CONTENTS

ARTICLES

No “Dilettante Affair”: Rethinking the Experimental Use Exception to Patent Infringement for Biomedical Research Tools
Janice M. Mueller 1

Rights, Rights of Action, and Remedies: An Integrated Approach
Donald H. Zeigler 67

NOTES & COMMENTS

Qualifying Immunity: Protecting State Employees’ Right to Protect Their Employment Rights After Alden v. Maine
Raymond J. Farrow 149

Roberts v. Dudley: An Unnecessary Broadening of the Public Policy Exception to the Employment-At-Will Doctrine in Washington
Brian Hersey 179

The Reasonable Girl: A New Reasonableness Standard To Determine Sexual Harassment in Schools
Carrie L. Hoon 213

Protecting the Tax-Exempt Status of Housing Developers Participating in Low-Income Housing Tax Credit Partnerships
Marni Hussong 243

The Equal Pay Act as Appropriate Legislation Under Section 5 of the Fourteenth Amendment: Can State Employers Be Sued?
Thane Somerville 279

Copyright © 2001, Washington Law Review Association. Except as otherwise provided, the author of each article in this issue has granted permission for copies of that article to be made for classroom use, provided that (1) copies are distributed at or below cost, (2) the author and journal are identified, (3) proper notice of copyright is affixed to each copy, and (4) the Washington Law Review Association is notified of the use.
University of Washington School of Law 2000–2001

OFFICERS OF ADMINISTRATION
Richard L. McCormick, B.A., Ph.D., President of the University
Roland L. Hjorth, A.B., LL.B., Dean, Professor of Law
Richard O. Kummert, B.S., M.B.A., LL.B., Associate Dean, Professor of Law
Dexter Bailey, B.S., M.B.A., Assistant Dean for Development
Paula Littlewood, B.A., M.A., J.D., Assistant Dean for Administration and Public Relations
Sandra E. Madrid, B.S., Ph.D., Assistant Dean for Student Affairs and Administration

FACULTY EMERITI
William T. Burke, B.S. J.D., J.S.D.
Harry M. Cross, B.A., J.D.
Robert L. Fletcher, A.B., LL.B.
Dan Fenno Henderson, B.A., B.A., LL.B., Ph.D.
John Huston, B.A., J.D., LL.M.
Cornelius J. Peck, B.S., LL.B.
John R. Price, A.B., LL.B.
Luvern V. Rieke, B.S., LL.B., LL.M., LL.D.
Marjorie D. Rombauer, B.A., J.D.
Charles Z. Smith, B.S., J.D.
Frank W. Smith, Jr., B.A., J.D., LL.M.

FACULTY OF LAW
Craig H. Allen, B.S., J.D., Associate Professor of Law, Director of Law and Marine Affairs Program
William R. Andersen, B.S.L., LL.B., LL.M., Judson Falknor Professor of Law
Karen Boxa, B.A., J.D., Assistant Professor of Law
Steve P. Calandraillo, B.A., J.D., Assistant Professor of Law
Donald C. Clarke, A.B., M.Sc., J.D., Professor of Law, Washington Law School Foundation Scholar
Sam A. Donaldson, B.A., J.D., LL.M., Assistant Professor of Law
G. Meade Emory, A.B., LL.B., LL.M., Professor of Law, Director of Graduate Program in Taxation
Joan M. Fitzpatrick, B.A., J.D., Jeffery and Susan Brotman Professor of Law
Julia Gold, B.A., J.D., Senior Lecturer
James H. Hardisty, A.B., LL.B., Professor of Law
Penny A. Hazleton, B.A., J.D., LL.M., Professor of Law, Law Librarian, Foundation Scholar
Marc J. Hershman, A.B., J.D., Adjunct Professor of Law
Gregory A. Hicks, B.A., J.D., Professor of Law
Linda S. Hume, B.A., J.D., Professor of Law
Stewart M. Jay, A.B., J.D., Professor of Law
Nancy S. Jecker, B.A., M.A., Ph.D., Adjunct Professor of Law
Michele E. Jones, B.A., J.D., Senior Lecturer
John M. Junker, B.A., J.D., Professor of Law
Alan Kirtley, B.A., J.D., Associate Professor of Law, Director of Clinics
Patricia C. Kolstol, B.A., M.D., J.D., Associate Professor of Law, Washington Law School Foundation Scholar
Elizabeth F. Lofthus, B.A., M.A., Ph.D., Adjunct Professor of Law
Deborah Maranville, B.A., J.D., Associate Professor of Law
Anna Marszian, B.S./B.A., J.D., M.P.H., Assistant Professor of Law
Michael W. McCann, B.A., A.M., Ph.D., Adjunct Professor of Law
Kathleen M. McGinnis, B.A., J.D., Lecturer
Jacqueline McMurtrie, B.G.S., J.D., Senior Lecturer
Peter Morris, B.A., M.A., J.D., LL.M., LL.D., Professor of Law, Adjunct Professor of Education
Kathleen O’Neill, B.A., J.D., Assistant Professor of Law, Director of Legal Research and Writing
Roy L. Prosterman, A.B., J.D., Professor of Law
Anita Ramasastry, B.A., M.A., J.D., Assistant Professor of Law
William H. Rodgers, Jr., B.A., LL.B., Professor of Law
Eric Schnapper, B.A., M.A., B.Phil., LL.B., Pendleton Miller Professor of Law
William B. Stoebeck, B.A., M.A., J.D., S.J.D., Professor of Law
Toshiko Takenaka, B.A., LL.M., Ph.D., Associate Professor of Law, Director, CASRIP
Veronica L. Taylor, LL.B., B.A., LL.M., Professor of Law, Asian Law Program Director
Michael E. Townsend, B.A., M.A., Ph.D., J.D., Associate Professor of Law
Philip A. Trautman, B.A., J.D., Professor of Law
Lea B. Vaughn, A.B., J.D., Professor of Law
Walter J. Walsh, B.C.L., LL.M., Assistant Professor of Law
Lisa W. Wiel, A.B., M.A., J.D., Associate Professor of Law
Louis W. Wolcher, B.A., J.D., Professor of Law
Richard O. Zerbe, Jr., A.B., Ph.D., Adjunct Professor of Law
CONTENTS

ARTICLES
Attorney-Client Confidentiality and the Assessment of Claimants Who Alleged Posttraumatic Stress Disorder
Robert H. Aronson, Lonnie Rosenwald, & Gerald M. Rosen 313

Conversion and Merger of Disparate Business Entities
Robert C. Art 349

DNA Typing: Emerging or Neglected Issues
Edward J. Imwinkelried & D.H. Kaye 413

NOTES & COMMENTS
One Step Forward, Two Steps Back: Vasquez v. Hawthorne Wrongly Denied Washington’s Meretricious Relationship Doctrine to Same-Sex Couples Amanda J. Beane 475

A Shield, Not a Sword: Involuntary Leave Under the Family and Medical Leave Act Megan E. Blomquist 509

Unopened Public Street Easements in Washington: Whose Right to Use That Land Is It, Anyway? Alfred E. Donohue 541

Grandma Got Run Over by the Supreme Court: Suggestions for a Constitutional Nonparental Visitation Statute After Troxel v. Granville Eric B. Martin 571

Two Wrongs Do Not Make a Defense: Eliminating the Equal-Opportunity-Harasser Defense Shylah Miles 603

The First Amendment Versus the World Trade Organization: Emergency Powers and the Battle in Seattle Aaron Perrine 635
Kim Ho Ma v. Reno: Cloaking Judicial Activism As Constitutional Avoidance

Matthew E. Hedberg 669

Copyright © 2001, Washington Law Review Association. Except as otherwise provided, the author of each article in this issue has granted permission for copies of that article to be made for classroom use, provided that (1) copies are distributed at or below cost, (2) the author and journal are identified, (3) proper notice of copyright is affixed to each copy, and (4) the Washington Law Review Association is notified of the use.
WASHINGTON LAW REVIEW

VOLUME 76
2001

2000–2001 EDITORIAL BOARD

Editor-in-Chief
CHARLES S. WRIGHT

Managing Editors
MARY E. CREGO
CHRISTOPHER B. DURBIN
BENTON J. GAFFNEY
H. MATTHEW MUNSON
KATIE HOSFORD TRAVERSE

Topics Development
CARA J. FREY
CARRIE L. HOON

Associate Editors-in-Chief
PATRICIA BAUGHER
AMY EDWARDS
CORY J. ALBRIGHT
LINELL A. LETENDRE

Executive Notes and Comments Editors
BRIAN HERSEY
MARNI M. HUSSONG
ERIC M. PETERSON
JEFFREY SCHICK
THANE SOMERVILLE
CARRIE TRACY
MARK A. WILNER

Executive Articles Editors
KEITH BUTLER
ANASTASIA J. CAVIRIS

Notes and Comments Editors
BRIAN HERSEY
MARNI M. HUSSONG
ERIC M. PETERSON
JEFFREY SCHICK
THANE SOMERVILLE
CARRIE TRACY
MARK A. WILNER

Articles Editors
LANCE ANDREE
TARA CAUSLAND
RAYMOND J. FARROW
MICHAEL HATCH
MARGARET TOUGH

EDITORIAL STAFF

ANTHONY AZURE
JILL M. BALLO
AMANDA J. BEANE
DEBORAH L. BENDER
D. GREG BLANKINSHIP
MEGAN E. BLOMQVIST
ANDREW R. BRYANT
PIER LUIGI COLOMBO

ALFRED DONOHUE
REBECCA R. GLASGOW
BRAD A. GOERGEN
RYAN P. HARKINS
MATTHEW E. HEDBERG
KASEY HUEBNER
MARKETA VOREL
ERIC B. MARTIN
JOSEPH P. MEARA

SHYLAH R. MILES
PAUL W. MOOMAW
AARON PERRINE
SID REYNOLDS
Evan SHAPIRO
MATTHEW J. SMITH
JENNIFER S. STEVENSON
CHRISTOPHER WARAKSA

BUSINESS STAFF

Faculty Business Manager
PENNY HAZELTON

Faculty Adviser
RICHARD O. KUMMERT

Business Office
JONATHAN FRANKLIN
FRAN JOHNSON
University of Washington School of Law 2000–2001

OFFICERS OF ADMINISTRATION
Richard L. McCormick, B.A., Ph.D., President of the University
Roland L. Hjorth, A.B., LL.B., Dean, Professor of Law
Richard O. Kummert, B.S., M.B.A., LL.B., Associate Dean, Professor of Law
Dexter Bailey, B.S., M.B.A., Assistant Dean for Development
Paula Littlewood, B.A., M.A., J.D., Assistant Dean for Administration and Public Relations
Sandra E. Madrid, B.S., Ph.D., Assistant Dean for Student Affairs and Administration

FACULTY EMERITI
William T. Burke, B.S. J.D., J.S.D.
Harry M. Cross, B.A., J.D.
Robert L. Fletcher, A.B., LL.B.
Dan Fenno Henderson, B.A., B.A., LL.B., Ph.D.
John Huston, B.A., J.D., LL.M.
Cornelius J. Peck, B.S., LL.B.
John R. Price, A.B., LL.B.
Luvern V. Rieke, B.S., LL.B., LL.M., LL.D.
Marjorie D. Rombauer, B.A., J.D.
Charles Z. Smith, B.S., J.D.
Frank W. Smith, Jr., B.A., J.D., LL.M.

FACULTY OF LAW
Craig H. Allen, B.S., J.D., Associate Professor of Law, Director of Law and Marine Affairs Program
William R. Andersen, B.S.L., LL.B., LL.M., Judson Falknor Professor of Law
Helen A. Anderson, B.A., J.D., Senior Lecturer
Robert Anderson, B.A., J.D., Lecturer, Director of the Native American Law Center
Thomas R. Andrews, B.A., M.A., J.D., Professor of Law
Robert H. Aronson, B.A., J.D., Professor of Law
Daniel M. Bodansky, A.B., M.Phil., J.D., Professor of Law, Washington Law School Foundation Scholar
Karen Boxa, B.A., J.D., Assistant Professor of Law
Steve P. Calandrillo, B.A., J.D., Assistant Professor of Law
Donald C. Clarke, A.B., M.Sc., J.D., Professor of Law, Washington Law School Foundation Scholar
Sam A. Donaldson, B.A., J.D., LL.M., Assistant Professor of Law
G. Meade Emory, A.B., LL.B., LL.M., Professor of Law, Director of Graduate Program in Taxation
JOAN M. FITZPATRICK, B.A., J.D., Jeffery and Susan Brotman Professor of Law
Julia Gold, B.A., J.D., Senior Lecturer
James H. Hardisty, A.B., LL.B., Professor of Law
Penny A. Hazleton, B.A., M.L., Professor of Law, Law Librarian, Foundation Scholar
Marc J. Hershman, A.B., J.D., Adjunct Professor of Law
Gregory A. Hicks, B.A., J.D., Professor of Law
Linda S. Hume, B.A., J.D., Professor of Law
Stewart M. Jay, A.B., J.D., Professor of Law
Nancy S. Secker, B.A., M.A., Ph.D., Adjunct Professor of Law
Michele E. Jones, B.A., J.D., Senior Lecturer
John M. Junker, B.A., J.D., Professor of Law
Alan Kirtley, B.A., J.D., Associate Professor of Law, Director of Clinics
Patricia C. Kozlizier, B.A., M.D., J.D., Associate Professor of Law, Washington Law School Foundation Scholar
Elizabeth F. Loftus, B.A., M.A., Ph.D., Adjunct Professor of Law
Deborah Maranville, B.A., J.D., Associate Professor of Law
Anna Mastroianni, B.S.B.A., J.D., M.P.H., Assistant Professor of Law
Michael W. McCann, B.A., M.A., Ph.D., Adjunct Professor of Law
Kathleen M. McGinnis, B.A., J.D., Lecturer
Jacqueline MCMURTRIE, B.G.S., J.D., Senior Lecturer
Peter Nicolas, B.A., M.A., J.D., LL.M., LL.D., Professor of Law, Adjunct Professor of Education
Kathleen O’Neill, B.A., J.D., Assistant Professor of Law, Director of Legal Research and Writing
Roy L. Prosternessman, A.B., LL.B., Professor of Law
Anita Ramasastry, B.A., M.A., J.D., Assistant Professor of Law
William H. Rodgers, Jr., B.A., LL.B., Professor of Law
Eric Schnapper, B.A., M.A., Ph.D., LL.B., Pendleton Miller Professor of Law
William B. Stoebeck, B.A., M.A., J.D., S.J.D., Professor of Law
Toshiko Takenaka, B.A., LL.M., Ph.D., Associate Professor of Law, Director, CASRIP
Veronica L. Taylor, B.L.L., B.A., LL.M., Professor of Law, Asian Law Program Director
Michael E. Townsend, B.A., M.A., J.D., Associate Professor of Law
Philip A. Trautman, B.A., J.D., Professor of Law
Lea B. Vaughn, A.B., J.D., Professor of Law
Walter J. Walsh, B.C.L., LL.M., Assistant Professor of Law
Lis W. Wiehl, A.B., M.A., J.D., Associate Professor of Law
Louis E. Wolcher, B.A., J.D., Professor of Law
Richard O. Zerbe, Jr., A.B., Ph.D., Adjunct Professor of Law
CONTENTS

ARTICLES

The Declaration of Independence: A 225th Anniversary Re-Interpretation
Carlton F.W. Larson 701

E-Proxies for Sale? Corporate Vote-Buying in the Internet Age
Douglas R. Cole 793

NOTES & COMMENTS

Echazabal v. Chevron: A Direct Threat to Employers in the Ninth Circuit
Deborah Leigh Bender 859

Can Students Sue When Schools Don’t Make the Grade?
The Washington Assessment of Student Learning and Educational Malpractice
Rebecca R. Glasgow 893

In the Litigation Business: Insurance Company Liability for Acts Occurring in the Course of Litigation Under the Washington Consumer Protection Act
Kasey D. Huebner 927

Closing a Discrimination Loophole: Using Title VII’s Anti-Retaliation Provision to Prevent Employers from Requiring Unlawful Arbitration Agreements as Conditions of Continued Employment
Sidney Charlotte Reynolds 957

Copyright © 2001, Washington Law Review Association. Except as otherwise provided, the author of each article in this issue has granted permission for copies of that article to be made for classroom use, provided that (1) copies are distributed at or below cost, (2) the author and journal are identified,
WA SHINGTON LA W REV IEW

VOLUME 76
2001

2000–2001 EDITORIAL BOARD

Editor-in-Chief
CHARLES S. WRIGHT

Managing Editors
MARY E. CREGO
CHRISTOPHER B. DURBIN
BENTON J. GAFFNEY
H. MATTHEW MUNSON
KATIE HOSFORD TRAVERSE

Topics Development
CARA J. FREY
CARRIE L. HOON

Associate Editors-in-Chief
PATRICIA BAUGHER
AMY EDWARDS

Executive Notes and Comments Editors
CORY J. ALBRIGHT
LINELL A. LETENDRE

Notes and Comments Editors
BRIAN HERSEY
MARNI M. HUSSONG
ERIC M. PETERSON
JEFFREY SCHICK
THANE SOMERVILLE
CARRIE TRACY
MARK A. WILNER

Executive Articles Editors
KEITH BUTLER
ANASTASIA J. CAVIRIS

Articles Editors
LANCE ANDREE
TARA CAUSLAND
RAYMOND J. FARROW
MICHAEL HATCH
MARGARET TOUGH

EDITORIAL STAFF

ANTHONY AZURE
JILL M. BALLO
AMANDA J. BEANE
DEBORAH L. BENDER
D. GREG BLANKINSHIP
MEGAN E. BLOMQUIST
ANDREW R. BRYANT
PIER LUIGI COLOMBO
ALFRED DONOHOE

MARK K. FUNKE
REBECCA R. GLASGOW
BRAD A. GOERGEN
RYAN P. HARKINS
MATTHEW E. HEDBERG
KASEY D. HUEBNER
O. YALE LEWIS III
ERIC B. MARTIN
JOSEPH P. MEARA

SHYLAH R. MILES
PAUL W. MOOMAW
AARON PERRINE
SID REYNOLDS
EVAN SHAPIRO
MATTHEW J. SMITH
JENNIFER S. STEVENSON
CHRISTOPHER WARAKSA

BUSINESS STAFF

Faculty Business Manager
PENNY HAZELTON

Faculty Adviser
RICHARD O. KUMMERT

Business Office
JONATHAN FRANKLIN
FRAN JOHNSON
CONTENTS

HARRY MAYBURY CROSS: IN MEMORIAM

William B. Stoebuck  991

ARTICLES

Punitive Damages, Explanatory Verdicts, and the Hard Look
Richard W. Murphy  995

The Asymmetry of State Sovereign Immunity
Richard H. Seamon  1067

NOTES & COMMENTS

Festo’s Effect on After-Arising Technology and the Doctrine of Equivalents
Anthony H. Azure  1153

Holding Tortfeasors Accountable: Apportionment of Enhanced Injuries Under Washington’s Comparative Fault Scheme
Ryan P. Harkins  1185

Fire Sale! The Admissibility of Evidence of Environmental Contamination To Determine Just Compensation in Washington Eminent Domain Proceedings
Paul W. Moomaw  1221

The Religious Land Use and Institutionalized Persons Act: An Analysis Under the Commerce Clause
Evan M. Shapiro  1255
2001–2002 EDITORIAL BOARD*

Editor-in-Chief
ERIC B. MARTIN

Managing Editors
ALFRED DONOHUE
KASEY HUEBNER
PAUL W. MOOMAW
AARON PERRINE
MATTHEW J. SMITH

Topics Development
JILL M. BALLO
DEBORAH L. BENDER

Associate Editors-in-Chief
BRAD A. GOERGEN
SHYLAH R. MILES

Executive Notes and Comments Editor
REBECCA R. GLASGOW

Associate Notes and Comments Editors
AMANDA J. BEANE
MEGAN E. BLOMQUIST

Notes and Comments Editors
P. LUIGI COLOMBO
RYAN P. HARKINS
MATTHEW E. HEDBERG
O. YALE LEWIS III
EVAN M. SHAPIRO
CHRISTOPHER WARKSA

EDITORIAL STAFF
KIMBERLY ALDERSON
AMY ALLISON
REBECCA ANDREWS
KRISTI DARNELL
KAUSTUV M. DAS
ANDRE EGGLE
SARAH E. FARLEY
ELIZABETH FARNAM
CECILY FUHR
ROGER JANSSON
RYAN JENSEN
TOBIAS KAMMER
AARON MATTHEW LAING
SHANNON M. McMINNIE
JANNA J. ANNEST
TIM MITROVICH

Executive Articles Editors
JOSEPH P. MEARA
SID REYNOLDS

Articles Editors
ANTHONY AZURE
D. GREG BLANKINSHIP
ANDREW R. BRYANT
MARK K. FUNKE
JENNIFER S. STEVENSON

Julie NAPIER
JAMES R. PUTNAM
KRISTIN RELYEA
JILL SIMMONS
JANIS SNOEY
AIMEE SUTTON
MEGAN WALSETH

BUSINESS STAFF
Faculty Business Manager
PENNY HAZELTON

Faculty Adviser
RICHARD O. KUMMERT

Business Manager
JONATHAN FRANKLIN
FRAN JOHNSON

*Washington Law Review issues are produced on a calendar-year schedule; the Editorial Board’s term of office, however, corresponds to the school year. Accordingly, editorial responsibilities for October issues are apportioned between consecutive Boards.
University of Washington School of Law 2001–2002

OFFICERS OF ADMINISTRATION
Richard L. McCormick, B.A., Ph.D., President of the University
W. H. Knight, Jr., B.A., J.D., Dean, Professor of Law
Richard O. Kummert, B.S., M.B.A., LL.B., Associate Dean for Academic Affairs, Professor of Law
Sandra E. Madrid, B.S., Ph.D., Assistant Dean for Student Affairs and Administration
Dexter Bailey, B.S., M.B.A., Assistant Dean for Development
Paula Littlewolf, B.A., M.A., J.D., Assistant Dean for Administration and Public Relations

FACULTY EMERITI
William T. Burke, B.S. J.D., J.S.D.
Robert L. Fletcher, A.B., LL.B.
John Huston, B.A., M.A., J.D., LL.M., LL.D.
Arval Morris, B.A., M.A., J.D., LL.M., LL.D.
Cornelius J. Feck, B.S., LL.B.
John R. Price, B.A., LL.B.
Luvern V. Rieke, B.S., LL.B., LL.M., LL.D.
Maryrose D. Rombauer, B.A., J.D.
Charles Z. Smith, B.S., J.D.
Frank W. Smith, Jr., B.A., J.D., LL.M.

FACULTY OF LAW
Craig H. Allen, B.S. J.D., Professor of Law, Director of Law and Marine Affairs Program
William A. Anderson, B.S.L., LL.B., LL.M., Judson Falknor Professor of Law
Helen A. Anderson, B.A., J.D., Senior Lecturer
Robert T. Anderson, B.A., J.D., Assistant Professor of Law, Native American Law Center
Thomas R. Andrews, B.A., M.A., J.D., Professor of Law
Robert H. Aronson, B.A., J.D., Professor of Law
Melissa M. Berry, B.A., J.D., Lecturer
Daniel M. Bodansky, A.B., M.Phil., J.D., Professor of Law, Wash. Law School Foundation Scholar
Karen Boxx, B.A., J.D., Assistant Professor of Law
Steve P. Calandrillo, B.A., J.D., Assistant Professor of Law
Donald C. Clarke, A.B., M.Sc., J.D., Professor of Law, Wash. Law School Foundation Scholar
Sam A. Donaldson, B.A., J.D., LL.M., Assistant Professor of Law
G. Meade Emsory, A.B., LL.B., LL.M., Professor of Law, Director of Graduate Program in Taxation
Joan M. Fitzpatrick, B.A., J.D., Jeffery and Susan Brotman Professor of Law
James H. Hardisty, A.B., LL.B., Professor of Law
Penny A. Hazelton, B.A., M.L.L., Professor of Law, Law Librarian, Law Foundation Scholar
Gregory A. Hicks, B.A., J.D., Professor of Law
Roland L. Hjorth, A.B., LL.B., Dean Emeritus and Garvey, Schubert, & Baer Professor of Law
Mary A. Hotchkiss, B.A., M.S.L.S., J.D., LL.M., Senior Lecturer
Linda S. Hume, B.A., J.D., Professor of Law
Stewart M. Jay, A.B., J.D., William L. Dwyer Professor of Law
John M. Junker, B.A., J.D., Professor of Law
Alan Kirtley, B.A., J.D., Associate Professor of Law, Director of Clinics
Patricia C. Kuszler, B.A., M.D., J.D., Professor of Law, Wash. Law School Foundation Scholar
Deborah Maranville, B.A., J.D., Associate Professor of Law
Anna Mastroianni, B.S./B.A., J.D., M.P.H., Assistant Professor of Law
Kathleen M. McGinnis, B.A., J.D., Senior Lecturer
Jacqueline McMurtrie, B.G.S., J.D., Senior Lecturer
Peter Nicolas, B.A., M.P.P., J.D., Assistant Professor of Law
Kathleen O’Neill, B.A., J.D., Associate Professor of Law, Director of Legal Research and Writing
Roy L. Prosterman, A.B., LL.B., Professor of Law
Anita Ramasastry, B.A., M.A., J.D., Assistant Professor of Law
William H. Rodgers, Jr., B.A., LL.B., Stimson Bullitt Professor of Environmental Law
Eric Schnapper, B.A., M.A., B.Phil., LL.B., Pendleton Miller Professor of Law
Scott A. Schumacher, B.A., J.D., Lecturer
William B. Sheepherd, B.A., M.A., J.D., S.J.D., Judson Falknor Professor of Law Emeritus
Toshiko Takeda, B.A., LL.M., Ph.D., Associate Professor of Law, Director, CASRIP
Veronica L. Taylor, LL.B., B.A., LL.M., Professor of Law, Asian Law Program Director
Michael E. Townsend, B.A., M.A., Ph.D., J.D., Associate Professor of Law
Philip A. Trautman, B.A., J.D., Professor of Law
Lea B. Vaughn, A.B., J.D., Professor of Law
Maria D. Victoria, B.A., J.D., Lecturer
Walter J. Walsh, B.C.L., LL.M., Assistant Professor of Law
Lis W. Wiehl, A.B., M.A., J.D., Associate Professor of Law
Louis E. Wolcher, B.A., J.D., Professor of Law

ADJUNCT AND AFFILIATE FACULTY
G. Andrew H. Benjamin, B.A., M.A., J.D., Ph.D., Affiliate Professor of Law
Daniel H. Foote, A.B., J.D., Affiliate Professor of Law
John R. Price, B.A., M.A., J.D., Affiliate Professor of Law
Marc I. Herrman, A.B., J.D., Adjunct Professor of Law
Nancy S. Jecker, B.A., M.A., Ph.D., Adjunct Professor of Law
Elizabeth F. Loftus, B.A., M.A., Ph.D., Adjunct Professor of Law
Michael W. McCann, B.A., M.A., Ph.D., Adjunct Professor of Law
Hugh D. Spitzer, B.A., J.D., LL.M., Affiliate Professor of Law
Richard O. Zerbe, Jr., A.B., Ph.D., Adjunct Professor of Law
NO “DILETTANTE AFFAIR”: RETHINKING THE
EXPERIMENTAL USE EXCEPTION TO PATENT
INFRINGEMENT FOR BIOMEDICAL RESEARCH TOOLS

Janice M. Mueller*

Abstract: Scientists who require multiple “research tools” (i.e., laboratory resources such as
transgenic animals and biological receptors) to develop new drugs and medical diagnostic
products are frequently finding that these tools are patented or subject to other proprietary
constraints. Stacking royalty obligations and heightened transaction costs resulting from the
proliferation of patents on research tools threaten to slow or stop the development of new drugs
and devices critical to public health. Because U.S. courts have very narrowly interpreted the
common law “experimental use” defense of patent law as limited to “dilettante” uses of
inventions for mere “amusement” or “philosophical” inquiry, scientists face the daunting choice
of either negotiating numerous licenses or risking the possibility that their research and
development will be enjoined. In response to this dilemma of mounting transaction costs and
increasingly restricted access to patented research tools, this Article argues for a broadened rule
of “development use” that would permit scientists to use certain patented research tools without
prior authorization, but require that the research tool patent owner be paid an ex post royalty
based on the ultimate commercial success of the new drugs or other products developed through
use of the tool. This “reach-through” royalty approach maintains incentives for the development
and patenting of new research tools, but alleviates the access restrictions and up-front costs
currently associated with their acquisition and use.

I.   INTRODUCTION ......................................................................2
II.   DEFINING AND ACCESSING “RESEARCH TOOLS” .... 10
III.  LIMITED RECOGNITION OF A “TRULY NARROW”
     EXPERIMENTAL USE DOCTRINE IN THE UNITED
     STATES ................................................................................... 17
     A. United States Common Law Evolution of a “Truly
        Narrow” Experimental Use Exemption .............................. 19
        1. Early Development: Commercial Intent
           Prohibited ........................................................................ 19
        2. Federal Circuit Development: Toward a
           Restrictive Application of the Experimental Use
           Exemption ........................................................................ 22
     B. The Hatch-Waxman Act: A Limited Exemption for
        Regulatory Data Gathering ................................................. 25

* Associate Professor, The John Marshall Law School, Chicago, Illinois. The author thanks Yvette
Barksdale, Charles Gholz, Jorge Goldstein, Michael Jakes, Walter Kendall, Robert Merges, and Samuel
Olken for reviewing and commenting on earlier drafts of this Article.
The thermostable enzyme *Thermus aquaticus* YT1 DNA polymerase (Taq)\(^1\) is a basic and widely used biotechnology tool. Purified and isolated from a bacterium discovered in the hot springs of Yellowstone National Park,\(^2\) a small amount of Taq is added to the test tube every time a forensic scientist analyzes blood from a crime scene or a laboratory technician tests an AIDS patient’s HIV levels.\(^3\) Taq is also widely used in DNA sequencing.\(^4\) In both its native and recombinant forms, Taq is patented.\(^5\)

---

\(^1\) The identification, purification, and introduction of Taq are described in PAUL RABINOW, MAKING PCR: A STORY OF BIOTECHNOLOGY 128–32 (1996).


\(^4\) *All Things Considered: Court Calls Biotechnology Patent Into Question* (National Public Radio broadcast, Aug. 14, 1996), at 1996 WL 12726284 (statement of ProMega official Randy Diamond that over half of world’s life science research laboratories use Taq “in either the PCR process or in DNA sequencing”).
1991, the Swiss pharmaceutical giant Hoffman-La Roche (Roche) obtained from the now-defunct Cetus Corporation the patent rights in Taq as well as polymerase chain reaction or “PCR,” the revolutionary DNA amplification process that utilizes Taq.6

By means of an unprecedented federal court filing in May 1995, Roche accused more than forty U.S. universities and research institutes (including Harvard, Stanford, Massachusetts Institute of Technology, the Salk Institute, the Scripps Research Institute, and the National Cancer Institute) and more than 200 individual scientists of infringing these patents.7 The scientists’ alleged wrongdoing was purchasing Taq for use in the PCR process from a bioscience supply company that, according to Roche, did not have a proper license to sell Taq for use in PCR.8 Roche officials professed no concern about the use of Taq for “pure research” purposes, but stated that they felt compelled to take action against those scientists engaged in what Roche termed “highly practical” research with profit-making potential.9 Although Roche officials denied any intent to formally join the scientists as parties to Roche’s ongoing patent litigation against the Taq supplier, Promega Corporation of Wisconsin, a Roche spokesperson obliquely warned that she “wouldn’t want to predict what action Roche would take relative to any patent . . . in the future.”10 At a San Francisco press conference hastily called by Promega in the days following Roche’s filing, Nobel laureate Dr. Arthur Kornberg of Stanford University decried Roche’s attempts to restrict the use of Taq and PCR technology as “violat[ing] practices and principles basic to the advancement of knowledge for the public welfare.”11


7. See generally LaRoche, supra note 5; see also Mario C. Aguilera, Local Institutes Named in Lawsuit, SAN DIEGO DAILY TRANSCRIPT, May 25, 1995, at 1; Marcia Barinaga, Scientists Named in PCR Suit, 268 SCIENCE 1273, 1273 (1995).


9. Rubenstein, supra note 2, at 28.


Roche’s naming of hundreds of prominent scientists as potential patent infringers dramatically underscores the current debate over proprietary rights in biomedical “research tools,” the many varied resources used by scientists to conduct research and development of new drugs, therapies, diagnostic methods, and other therapeutic products. At bottom, the dispute stems from the broad rights conferred by the patents covering these tools. A U.S. patent grants its owner the right, inter alia, to prevent others from using the patented invention, without qualification as to the nature or purpose of the use. Non-consensual uses of patented inventions that lead to the development of other products may result in patent infringement liability, even though sales of these products do not involve selling the patented invention. For example, a researcher may infringe if he or she uses without a license a patented research tool, such as Taq, a biological receptor, or a transgenic animal model in the research and development

12. “Research tools” are further defined in Part II infra.
13. 35 U.S.C. § 154(a)(1) (1994) (“Every patent shall contain . . . a grant to the patentee . . . of the right to exclude others from making, using, offering for sale, or selling the invention throughout the United States or importing the invention into the United States . . . .”) (emphasis added); id. § 271(a) (“Except as otherwise provided in this title, whoever without authority makes, uses, offers to sell, or sells any patented invention, within the United States or imports into the United States any patented invention during the term of the patent therefore, infringes the patent.”) (emphasis added).
14. The “use” contemplated by § 271(a) as an act of infringement is limited by the statute to use “without authority.” Id. By contrast, “use” of a patented device legally purchased is not infringement. A purchaser is deemed to have obtained an implied license to use the purchased patented device. Aro Mfg. Co. v. Convertible Top Replacement Co., 377 U.S. 476, 484 (1964) (“[I]t is fundamental that sale of a patented article by the patentee or under his authority carries with it an ‘implied license to use.’”) (citing Adams v. Burke, 84 U.S. (17 Wall.) 453, 456 (1873)). The limits of authorized “use” are informed by the “permissible repair” versus “infringing reconstruction” debate. See Hewlett-Packard Co. v. Repeat-O-Type Stencil Mfg. Corp., 123 F.3d 1445, 1451–52 (Fed. Cir. 1997).
15. See John H. Barton, Patents and Antitrust: A Rethinking In Light of Patent Breadth and Sequential Innovation, 65 ANTITRUST L.J. 449, 451 (1997) (suggesting scenario in which firm obtaining patent protection on biological receptor useful in schizophrenia research could preempt others from further research in schizophrenia, without itself having any truly “marketable product”); see also Eliot Marshall, Patent on HIV Receptor Provokes an Outcry, 287 SCIENCE 1375, 1375–77 (2000) (describing academic researchers’ criticism of patent issued to Human Genome Sciences, Inc. (HGS) on CCR5 cell-surface receptor that HIV uses as cell entry point, and reporting HGS’s position that it will enforce patent against “anyone [who] wants to use the receptor to create a drug”). A “receptor” is a portion of a cell’s surface that binds with specific molecules “like a lock accepting a key.” Human Pheromone Link May Have Been Found, N.Y. TIMES, Sept. 28, 2000, at A22.
16. For example, U.S. Patent No. 5,675,060 (issued Oct. 7, 1997) (“Transgenic arthritic mice expressing a T-cell receptor transgene”) discloses and claims transgenic arthritic mice that are useful as animal models for the evaluation of human arthritogenic and therapeutic anti-arthritic compositions. Genetically altered mice are preferred models for many human diseases because the mouse genome is similar to the human genome. David Malakoff, The Rise of the Mouse, Biomedicine’s Model Mammal, SCIENCE, Apr. 14, 2000, at 248.
Experimental Use Exception to Patent Infringement

of new drugs, therapies, or diagnostic products to be sold commercially. “Use” liability arises under the patent laws even though the researcher has not physically incorporated the patented tool into the new product that is ultimately marketed.17

Many industrialized countries recognize an exception to patent infringement liability for non-consensual uses of patented inventions for experimental or research purposes.18 An experimental use exception has met with little success in the United States, however.19 The U.S. Court of Appeals for the Federal Circuit has grudgingly recognized the existence of a common law experimental use defense, but characterizes it as “truly narrow” and applicable only to trifling “dilettante affairs.”20 Banished from the experimental use defense is any activity viewed as “commercialization” or otherwise grounded on profit motive.21 The current narrow interpretation of the doctrine virtually assures that it cannot be relied on by the rapidly growing number of university and industry collaborations whose research and development efforts are ultimately targeted at the commercialization of new biomedical products.22

The shortcomings of the experimental use doctrine as currently interpreted in the United States are receiving increasing attention in a climate of heightened concern over access to patented research tools. The explosion of biotechnological and biomedical research and development in the United States in the past twenty years,23 with a corresponding increase

17. Because the patented research tool is not incorporated into the commercial product, the researcher has not violated the “sells” prohibition of 35 U.S.C. § 271(a). See supra note 13.
18. Part V infra examines the implementation of the experimental use doctrine in foreign patent laws.
21. E.g., Roche, 733 F.2d at 863 (refusing to adopt broader view of experimental use doctrine that would “allow a violation of the patent laws in the guise of ‘scientific inquiry,’ when that inquiry has definite, cognizable, and not insubstantial commercial purposes”); Pitcairn v. United States, 547 F.2d 1106, 1125–26 (Ct. Cl. 1976) (rejecting government’s experimental use defense because government’s unauthorized use of infringing helicopters for testing, demonstrations, and experiments was “in keeping with the legitimate business of the using agency”); Deuterium Corp. v. United States, 19 Cl. Ct. 624, 633 (1990) (Rader, J.) (rejecting experimental use defense because government agency’s participation in demonstration project with for-profit partner corporation “was not strictly intellectual experimentation, but development of technology and processes for commercial applications”).
22. See Part IV infra.
23. The U.S. Supreme Court’s 1980 decision in Diamond v. Chakrabarty, 447 U.S. 303, which upheld the patentability of a living, genetically engineered bacterium, see id. at 318, is generally viewed as having given the green light to the U.S. biotechnology industry. The initial public offering of Genentech, Inc. in October, 1980, is also considered a watershed event in biotechnology commercialization. Lynne G. Zucker et al., Geographically Localized Knowledge: Spillovers or Markets?, 36 ECON. INQUIRY 65 (1998). By 1998, there were reportedly 1,283 biotechnology companies

24. Guide to Biotechnology, supra note 23 (providing graphical representation of total numbers of U.S. biotechnology patents granted per year, shows that biotechnology patenting has escalated from less than 2000 patents issuing in 1985 to more than 9000 patents issuing in 1998); Report of the National Institutes of Health (NIH) Working Group on Research Tools 3, at http://www.nih.gov/news/researchtools/index.htm (June 4, 1998) [hereinafter NIH Research Tools Report] (reporting “increasing use of the patent system to obtain proprietary rights in research tools and increasing use of license agreements and material transfer agreements (MTAs) delineating the terms and conditions under which research tools can be used”).

25. See generally Karen Hall, Genomic Warfare, AM. LAW., June 2000, at 68. “Genomics” refers to that subset of the biotechnology industry which is engaged in finding, sequencing, and frequently patenting purified and isolated sequences of genes. Many of the genomics firms have piggybacked on the work of the Human Genome Project, a federal government effort to identify all genes in the human body. As of September 1999, reportedly 1800 U.S. patents had issued on animal, plant, and human genes; 7000 other genes were the subject of pending U.S. patent applications. Should Congress Liberate Gene Data?, CHI. TRIB., Sept. 16, 1999, at A26 [hereinafter Gene Data]. In January 2000, more than seventy genomics companies were poised to issue public stock offerings. Another Boom in Biotechnology Stocks: Genetic Research Lures Internet Cash, N.Y. TIMES, Jan. 23, 2000, at B9.

26. See, e.g., Principles and Guidelines for Recipients of NIH Research Grants and Contracts on Obtaining and Disseminating Biomedical Research Resources: Final Notice, 64 Fed. Reg. 72,900, 72,909 (Dec. 23, 1999) [hereinafter NIH Principles and Guidelines] (reporting formation of advisory committee to Director of NIH, Dr. Harold Varmus, to “look into problems encountered in the dissemination and use of proprietary research tools, the competing interests of intellectual property owners and research users underlying these problems, and possible NIH responses”); Irving N. Feit, Biotechnology Research and the Experimental Use Exception to Patent Infringement, 71 J. PAT. & TRADEMARK OFF. SOC’Y 819, 821 (1989) (predicting that “[a]s commercial products of biotechnology become available, the infringement of patents covering the basic research methods that led to the product will become an increasingly difficult problem”); Lita Nelsen, The Rise of Intellectual Property Protection in the American University, 279 SCIENCE 1460, 1460–61 (1998) (listing “[r]estricted availability or delays in exchange of ‘research tools’ (such as vectors or transgenic mice) in biological research” as one of several unresolved problems currently facing university technology-transfer management); Gene Data, supra note 25 (“Holders of genetic patents have been charging steep licensing
Experimental Use Exception to Patent Infringement

ever-greater numbers of proprietary tools,\(^\text{27}\) giving rise to transaction costs associated with acquiring the right to use each such tool. In some cases, the patentee may refuse to license the research tool altogether. The sum total of the transaction costs involved with acquisition of all necessary research tools may be so severe as to impede, postpone, or stop the development of important new products.

Within the highly patent-centric environment of biotechnological and biomedical research and development, Michael Heller and Rebecca Eisenberg have argued that the scientific community is approaching a “tragedy of the anti-commons.”\(^\text{28}\) The anti-commons theory predicts that the proliferation of patents on “upstream” basic tools of biotechnological and biomedical research will stymie the development of sufficient numbers of downstream application products. Innovation is impeded by the “royalty stacking” problem imposed by the numerous upstream patents that must be practiced in order to make the new downstream product.\(^\text{29}\) The problem of too many restrictions on upstream research tools resulting in an impoverishment of downstream products is the reverse of the famous “tragedy of the commons” theorized by Garrett Hardin in 1968.\(^\text{30}\) In Hardin’s metaphor, the absence of restrictions on access to public lands resulted in a tragedy of over-grazing; here the result is under-development of potentially important commercial drugs and therapeutic products.

The anti-commons theory is far from a merely academic construct. The National Institutes of Health (NIH) Working Group on Research Tools found in 1998 that “all segments” of the biotechnology research and development community surveyed agreed that “the stacking of intellectual property obligations as successive tools are used in the course of an extended research project has the potential to impede or even preclude the development of new and better diagnostic and therapeutic products.”\(^\text{31}\)

---

27. Kyla Dunn, *A Look at . . . Patents & Biotech*, WASH. POST, Oct 1, 2000, at B3 (reporting that “[t]he cost of doing [biotechnology] research now includes the cost of accessing these [research tool] patents—at whatever price the market will bear”).


29. Id. at 699.


31. NIH Research Tools Report, supra note 24, at 22; see also Rebecca S. Eisenberg, *Technology Transfer and the Genome Project: Problems With Patenting Research Tools*, 5 RISK 163, 168 (1994) (asserting that patents on research tools “can create obstacles to subsequent [research and development] and add to a thicket of rights that firms must negotiate their way past before they can get their products
patenting of numerous “knockout” (gene-deleted) laboratory mice in recent years has incensed scientists concerned about the high prices and burdensome licensing conditions associated with these research tools. University licensing directors list “[r]estricted availability or delays in exchange of ‘research tools’ (such as vectors or transgenic mice) in biological research” as a key, unresolved challenge. Scientists using “DNA chip” technology to screen patients for genetic variations envision a nightmare scenario in which a license is required for each of the thousands of DNA sequences attached to a thumbnail-sized chip. In the plant biotechnology sector, technology-transfer officials report stifled development because of the veto power wielded by the owners of patents on research tools such as promoters and transformation systems that are necessary for the commercial development of transgenic plants. In an attempt to overturn what they view as overly restrictive licensing terms and excessive licensing fees, the families of children afflicted with Canavan disease, a rare genetic disorder of the brain, are suing the research scientists who isolated and patented the gene responsible for the illness.

Specific types of obstacles to access that the NIH Working Group investigated included “refusals to license, onerous royalty obligations, restrictions on the dissemination of materials and information, restrictions on the ability to collaborate with commercial firms, and advance commitments regarding intellectual property rights in future discoveries.” NIH Research Tools Report, supra note 24, at 4.

32. Eliot Marshall, A Deluge of Patents Creates Legal Hassles for Research, 288 SCIENCE 255, 255–57 (2000). Marshall reports that “[t]his tension between the creators and the controllers of knockout mice is indicative of a tension throughout the research world.” Id. at 255.

33. Nelsen, supra note 26, at 1460–61 (Ms. Nelsen is director of the Technology Licensing Office of the Massachusetts Institute of Technology); see also Richard Florida, The Role of the University: Leveraging Talent, Not Technology, ISSUES SCI. & TECH., Summer 1999, at 69–70 (noting concerns of smaller industrial firms that partner with universities over “protracted negotiations by university technology-transfer offices or attorneys over intellectual property rights,” which impede goal of getting new innovations to market as quickly as possible).


36. Peter Gorner, Parents Sue Over Patenting of Genetic Test, CHI. TRIB., Nov. 19, 2000, at 1. On October 30, 2000, the parents of children suffering from Canavan disease, along with the New York-based Canavan Foundation and other foundations, sued Dr. Reuben Matalon and his institution, Miami Children’s Hospital, in federal district court in Chicago seeking to block the hospital’s commercial use of the Canavan gene and recover money damages based on royalties that the hospital has collected from licensing a patented test that screens for the Canavan gene. Id. at 9. The Canavan Foundation alleges that it was forced to stop offering free genetic screening after learning that it would have to pay royalties and comply with licensing terms it considered onerous. Id.
Experimental Use Exception to Patent Infringement

This dilemma of gaining access to patented research tools needed in biotechnological and biomedical development suggests the importance of re-conceptualizing the experimental use doctrine as a partial, if not complete, solution. This Article argues that where significant transaction costs are associated with accessing the patented research tools necessary to develop downstream application products such as new drugs, therapies, and diagnostics, the non-consensual use of those tools, even though for ultimately commercial purposes, should no longer be automatically disqualified from the benefits of the experimental use doctrine. This Article further proposes a "liability rule" model that, while prohibiting the patent owner from enjoining the non-consensual use of the research tool, would appropriately compensate the patent owner in the form of an ex post

The lawsuit does not attack the validity of the Canavan gene patent, but instead seeks damages and equitable and injunctive relief to redress the defendants’ alleged breach of informed consent, breach of fiduciary duty, unjust enrichment, fraudulent concealment, conversion, and misappropriation of trade secrets. Complaint at 1, Daniel Greenberg v. Miami Children’s Hosp. Research Inst., Inc., (N.D. Ill. Oct. 30, 2000) (No. 00C 6779) (copy on file with author). The plaintiffs allege that, as owner of the Canavan gene patent, Miami Children’s Hospital "substantially restricted the number of laboratories authorized to conduct Canavan disease testing through exclusive licensing agreements" and "restricted public accessibility to testing through ‘volume caps’ that limited the number of tests to be performed by licensed laboratories and by requiring all such laboratories to pay royalty and licensing fees." Id. at 10. Among other remedies, the plaintiffs seek a permanent injunction restraining the hospital “from restricting access to prenatal and carrier testing for Canavan disease through exclusive licensing and/or collection of royalties, and from impeding research on finding a cure or therapies for Canavan disease through enforcement of Patent No. 5,679,635 or related international patents, or pursuit of pending international patents.” Id. at 13.

37. See NIH Research Tools Report, supra note 24, at 23 n.5 (proposing “clarification of the research exemption” as one of several issues for further consideration).

A number of academic commentators have suggested a larger role for the experimental use exemption, or at least the importance of a clearer understanding of the doctrine’s scope. See Barton, supra note 15, at 457 (proposing revision of experimental use exemption “so as clearly to permit use of patented technology for technology improvement purposes without needing to obtain an explicit license”); Jon Cohen, Chiron Stakes out Its Territory, 285 SCIENCE, 28 (1999) (quoting Robert Merges’s description of experimental use exemption’s reach as “one of the great unsolved mysteries of contemporary patent metaphysics”); Rebecca S. Eisenberg, Patents and the Progress of Science: Exclusive Rights and Experimental Use, 56 U. CHI. L. REV. 1017, 1020 (1989) (asserting that purpose and scope of experimental use defense “are not well defined” and that “this vaguely defined doctrine is becoming less satisfactory” as level of research utilizing patented materials continues to accelerate); Robert P. Merges & Richard R. Nelson, On the Complex Economics of Patent Scope, 90 COLUM. L. REV. 839, 866 n.118 (1990) (characterizing “precise contours” of experimental use defense as “unclear”); Arti Kaur Rai, Regulating Scientific Research: Intellectual Property Rights and the Norms of Science, 94 NW. U. L. REV. 77, 139 (1999) (suggesting broader interpretation of experimental use exception as one possible mechanism for reducing transaction and creativity costs associated with patenting basic scientific research).

38. “Liability rules” and “property rules” are defined infra note 278.
royalty based on the marketplace value of any new products developed through use of the tool; in other words, a “reach-through” royalty approach.

Part II provides a further definition of the nature of research tools and the increasingly problematic restrictions on access to the tools. Part III addresses the traditional narrow interpretation of the experimental use doctrine by the U.S. courts. Part IV discusses the implications of that interpretation for university-industry research and development collaborations. Part V examines the more liberal treatment of the experimental use doctrine in Europe and Japan. Part VI responds to arguments commonly cited against the experimental use exemption and suggests additional justifications for its broadening in the context of research tools. In Part VII, this Article proposes an expanded model of the experimental use doctrine that would permit the non-consensual “development use” of research tools, coupled with an ex post royalty payment based on the marketplace-determined value of the new products that result from that use. It concludes that the present “truly narrow” formulation of the experimental use doctrine in the United States is inapplicable to most research tool users, and the doctrine should be expanded in a manner that will maximize the development of important new therapeutic products while maintaining appropriate investment incentives for the ongoing creation of new research tools.

II. DEFINING AND ACCESSING “RESEARCH TOOLS”

The debate over patenting research tools begins with defining what research tools truly are. This Part surveys candidate meanings and provides examples of prominent research tools that have been subject to proprietary constraints. It contends that restrictions on access to such research tools are currently most problematic in the field of biotechnological research and development, and describes the types of escalating transaction costs faced by these enterprises.

“Research tools” is a phrase of many meanings depending on perspective. While researchers view the resources they rely on in the laboratory as “tools,” firms whose primary business is to manufacture and sell these resources may consider the same “tools” as “end products.”\(^\text{39}\) A clear definition of “research tools” is an essential prerequisite to any

---

Experimental Use Exception to Patent Infringement

proposal for modification of the patent owners’ right to prohibit non-consensual uses of the tools.40

The NIH Working Group on Research Tools (Working Group) was formed in 199741 to investigate the problems of NIH grantees in obtaining access to patented research resources.42 The Working Group report defines “research tool” in its broadest sense as “embrac[ing] the full range of resources that scientists use in the laboratory,”43 including “cell lines, monoclonal antibodies, reagents, animal models, growth factors, combinatorial chemistry libraries, drugs and drug targets, clones and cloning tools (such as PCR [polymerase chain reaction]), methods, laboratory equipment and machines, databases and computer software.”44

The problem of access to patented research tools is currently more acute and better documented in biotechnology than in any other scientific field.45 Biotechnology is research-intensive.46 A high percentage of the basic research tools and laboratory techniques of biotechnology are subject to proprietary restraints such as patents47 or material transfer agreements.48

40. Evelyn H. McConathy & Clifford K. Weber, Committee Report: University and Government IP Issues, AM. INTELL. PROP. L. A SS’N BULL., Mid-Winter 2000, at 178 (identifying "gray areas" in definition of research tools that impact determination of how their use "can, or should be, protected").
42. NIH Research Tools Report, supra note 24, at 4.
43. Id. at 3.
44. Id.; see also James G. Cullem, Panning for Biotechnology Gold: Reach-Through Royalty Damage Awards for Infringing Uses of Patented Molecular Sieves, 39 IDEA 553, 553 (1999) (describing “molecular sieves” as next-generation research tools used to “screen for, and identify, novel, specific compounds for known molecular targets”).
45. In contrast with the field of biotechnology, restrictions on proprietary research tools have not yet been reported as problematic in the development of Internet- and computer-implemented business methods and products. As patenting activity in these technologies continues to increase at an explosive pace, however, access to developmental tools may become an issue. For example, the developer of a new software program might need access to various patented algorithms in order to test the performance of a new program. Of course, developers who use “tools” that would qualify as “methods” under the new Prior Inventor Defense provisions of the American Inventors Protection Act of 1999 (codified at 35 U.S.C.A. § 273 (West Supp. 1999)) may be able to assert that defense without needing to rely on an experimental use doctrine as proposed in this Article.
46. The Biotechnology Industry Organization, a trade group for the biotechnology industry, reports that the U.S. biotechnology industry’s expenditure for research and development was $9.9 billion in 1998. Guide to Biotechnology, supra note 23.
47. See Feit, supra note 26, at 819 (contending that experimental use exception “is likely to have a particularly important impact on biotechnology, where basic laboratory methods and materials constitute the subject matter of patent claims”).
48. A material transfer agreement (MTA) is a negotiated contract between the owner of a tangible material, either patented or un patented, and a party who seeks the material for research use. See NIH Research Tools Report, supra note 24, at App. B. MTA’s tend to be less formal and shorter than patent
proprietary restraints such as patents or material transfer agreements. Proponents of patent protection contend that the high cost of developing these tools can be recouped only through temporary, sole-source pricing control. Moreover, because of the scientific complexity of their work, biotechnology researchers generally need access to a relatively greater number of proprietary research tools in order to conduct their research than do workers in other technologies.

Some of the most important research tools in biotechnology, all subject to proprietary restraints, include:

- The **cre-**\(\text{loxP}\) mouse. E.I. du Pont de Nemours owns the patented technology used to create “conditional mutants,” mice in whom a targeted gene is deleted when the **cre** gene encounters two **loxP** DNA segments bracketing the targeted gene.

- The Cohen-Boyer patents covering the basic method and plasmids for gene cloning, assigned to the University of California and Stanford University. These patents have been

license agreements, and generally do not require financial payments at the time of the material transfer. Id. The NIH and academic community have developed a standardized MTA for transfers of biological materials known as the “Uniform Biological Material Transfer Agreement” (UBMTA). Id.

50. For example, DNA-chip technology involves layering chains of nucleotides onto silicon. See Sandeep Junnarkar, ‘GeneChip’ Encodes DNA on Silicon, N.Y. TIMES (Mar. 15, 1997), at http://search1.nytimes.com/search/daily/bin/fastweb?getdoc+site+site+126360+2+wAAA+genechip. These thumbnail-size chips offer a myriad of potential applications in diagnosing genetic mutations and studying genes believed to be responsible for various types of cancer. Id. More than 40,000 gene sequences may be attached to a single 2.5-centimeter chip. Service, supra note 34, at 397. If each of those sequences is covered by a patent, a nightmarish licensing scenario may ensue. Id.


52. Eliot Marshall, The Mouse that Prompted a Roar, 277 SCIENCE 24 (1997). Research institutions including the NIH protested when DuPont required academic researchers to sign no-cost “research licenses.” Id. at 25. Some commercial institutions have taken licenses for more than $100,000. See id. The science press has characterized the restrictions on cre-**loxP** as “a lightning rod for scientists chafing at restrictions on the free flow of research materials;” id. at 24, and the National Academy of Sciences has cited the restrictions as a “commercial barrier[] to basic research;” id. at 25.

Experimental Use Exception to Patent Infringement

widely licensed\(^{54}\) and are often cited as a positive example of the benefits of patenting research tools.\(^{55}\)

- PCR (polymerase chain reaction) technology.\(^{56}\) Cetus Corporation originally owned the patents on PCR, a basic method of amplifying DNA sequences and the key reagent used in PCR, the enzyme Taq DNA polymerase.\(^{57}\)

- The Harvard “oncomouse” patent of Leder-Stewart, assigned to Harvard University and exclusively licensed to E.I. du Pont de Nemours.\(^{58}\) The tumor-prone “oncomouse” is useful as a model in cancer research.\(^{59}\)

- The expressed sequence tags (ESTs)\(^{60}\) that have been identified in the decoding of the human genome. Many of these small

---

54. The Cohen-Boyer patents have been “licensed to all comers.” Merges & Nelson, supra note 37, at 906. There are reportedly more than seventy licensees of the Cohen-Boyer patents. See Feit, supra note 26, at 820 (reporting that Cohen-Boyer patent license “provides for a minimum royalty of $10,000 per year and royalties based on net sales of final products at a rate of 0.5% to 10% based on the type of final product and country of sale”).


56. The PCR process selectively and exponentially amplifies (or multiplies) a specific region of DNA, producing quantities of DNA sufficient for experimentation and analysis. DRLICA, supra note 6, at 153–57, 314. The invention of PCR is detailed in RABINOW, supra note 1.

57. Eliot Marshall, Battling Over Basics, 277 SCIENCE 25 (1997). Cetus’s efforts to impose licenses on academic researchers using PCR were widely criticized. E.g., Barinaga, supra note 7, at 1273. Roche acquired the patent rights from Cetus in 1991 for $300 million. Service, supra note 5, at 2251. Since that time, it has imposed a multi-tiered licensing system. Marshall, supra note 57, at 25. Roche’s patent on the “native” form of the Taq enzyme (i.e., purified from the bacterium Thermus aquaticus) was recently held unenforceable by a federal district court as having been procured through inequitable conduct. Hoffman-La Roche, Inc. v. Promega Corp., No. C-93-1748 VRW, 1999 WL 1797330 (N.D. Cal. Dec. 7, 1999) (holding all claims of U.S. Patent No. 4,889,818 unenforceable); Service, supra note 5, at 2251. The ruling, now on appeal, was limited to n-Taq and did not directly effect Roche’s other patents on a recombinant form of Taq and on the PCR process. Id. at 2253.


59. In 1988, DuPont was selling the genetically modified mice to cancer researchers for approximately $50 each. See Eisenberg, supra note 37, at 1084.

60. An EST is a short fragment of complimentary DNA (cDNA), typically 150–400 base pairs in length, that is potentially useful as a probe to find the corresponding full-length gene. Rebecca S. Eisenberg & Robert P. Merges, Opinion Letter As To the Patentability of Certain Inventions Associated With the Identification of Partial cDNA Sequences, 23 AIPLA Q.J. 1, 2, 13–14 (1995). Complimentary DNA is DNA synthesized in test tubes from ribonucleic acid (RNA). DRLICA, supra note 6, at 305.
segments of complimentary DNA have no presently known utility, although they are believed to be useful as probes in searching for corresponding full-length genes.61

• Human embryonic stem cells, from which any type of human tissue can be grown.62 Stem cell research may someday permit doctors to grow transplant organs that identically match a patient’s tissue.63 The Wisconsin Alumni Research Foundation (WARF) owns a broad patent on these cells64 and has granted an exclusive license to Geron Corporation for commercial use.65

This Article limits the analysis of “research tools” to those patented tools used in development of new biotechnological or pharmaceutical products that do not themselves physically incorporate the tool. Thus defined, the sale of such products would not trigger the “sells” or “offers to sell”

61. Eisenberg & Merges, supra note 60, at 13–14. Whether ESTs are properly characterized as “research tools” requires that a distinction be made between ESTs themselves and patents on ESTs. An EST itself, i.e., the short segment of complimentary DNA, is a research tool in the sense that it is used as a tool to find a corresponding full-length gene. Public comments received by the U.S. Patent and Trademark Office (USPTO) in response to its first round of Interim Written Description Guidelines, published in June 1998, reflect this view. Commentators asserted that “ESTs are genomic research tools that should be available for unencumbered research to advance the public good.” Department of Commerce, Patent and Trademark Office, Revised Utility Examination Guidelines, 64 Fed. Reg. 71,440, 71,441 (Dec. 21, 1999) [hereinafter Examination Guidelines].

“Research tool” may not be a proper characterization for the many patents and pending patent applications directed to ESTs. Many in the biotechnology industry are concerned about EST patent claims of “open” scope (i.e., those reciting a purified and isolated “DNA comprising the sequence [EST nucleotide sequence data]”). Janice M. Mueller, Patent Law Examination Guidelines, Nat. L.J., Jan. 24, 2000, at B7; Rai, supra note 37, at 104 (stating that many pending EST patent applications claim “not only the EST but also the full gene of which it is a part and future uses of the gene”). Such claims appear to read on the full-length gene (as yet unknown), which could include within it the recited EST sequence plus a multitude of other sequences or regulatory elements. Thus, the EST sequence recited in the “comprising” claim would not be merely a tool, but actually a physical subset of the full-length gene once located. This is a very different paradigm from the use of the cre-loxP mouse to develop a new drug, where the mouse “tool” is used to make a new product but is not part of or incorporated into that product. See Robert Blackburn, Chief Patent Counsel, Chiron Corporation, remarks at the National Academies Board on Science, Technology, and Economic Policy’s Conference on “Intellectual Property Rights: How Far Should They Be Extended?” (Washington, D.C., Feb. 3, 2000) (contending that ESTs are not true “research tools” because they have no clear utility other than as part of final product).

62. Dunn, supra note 27, at B3.

63. Id.


65. Dunn, supra note 27, at B3 (reporting that although WARF is permitting scientists to access patented stem cells for “purely academic research,” researchers and patent experts are concerned that Geron’s right to exclude others from commercial use of cells “may determine whether or not new lifesaving therapies reach the public”).
liability provisions of the Patent Act.66 Rather, the liability for infringement of patents on these research tools, absent a license or exemption, would occur only under the “uses”67 (and in some cases, the “makes”)67 provisions of the statute.

Another important qualifier on the meaning of “research tools” as used herein is that the analysis is limited to those patented tools for which access is problematic. The patenting of a particular research tool does not necessarily create dissemination problems for that tool. Likewise, any resulting impediment to innovation of products requiring use of the tool is not uniformly high for all types of tools. Economically rational researchers would not risk infringement liability when they can simply buy widely available tools such as patented chemical reagents or genetically modified laboratory mice via supplier catalog or other anonymous market transactions.68 Transaction costs and access barriers would not be at issue here, and a broadened research exemption would not be necessary in such cases.69

In contrast, the possibility that research will be delayed or foregone, or that it will be conducted without authorization to use the patented research tool and lead to subsequent litigation, is much greater where the research tool must be acquired through direct license negotiations.70 Researchers may balk at the prospect of having to disclose the nature of their research in the course of obtaining the licenses they need.71 Alternatively, the suppliers of patented research tools may simply refuse to license them,72 or place

---

66. 35 U.S.C. § 271(a) (Supp. IV 1998) (“Except as otherwise provided in this title, whoever without authority makes, uses, offers to sell, or sells any patented invention, within the United States or imports into the United States any patented invention during the term of the patent therefore, infringes the patent.”); see also supra notes 13–14.

67. The “making” liability under 35 U.S.C. § 271(a) would involve the construction or synthesis of the research tool as a precursor to its use in developing a new product that does not incorporate the tool. For example, a cre-loxP mouse might be made (rather than purchased from an external source) by the research worker before it was used, but the ultimate sale of a new therapeutic product developed by using the patented mouse would not involve a sale of the mouse itself.

68. Eisenberg, supra note 31, at 171.

69. Cf. Barton, supra note 15, at 457 (advocating broadened experimental use exemption for follow-on innovators, “unless, of course, the technology is readily available, as through a research kit”).

70. Id.; see generally Mark A. Lemley, The Economics of Improvement in Intellectual Property Law, 75 Tex. L. Rev. 989, 1053–55 (1997) (describing various transaction costs involved in licensing intellectual property and concluding that these costs are “significant”).


72. Refusal to license a patent, even where the patent owner has market power in the antitrust law sense, has not been held to violate the antitrust laws. See Intergraph Corp. v. Int'l Corp., 195 F.3d 1346, 1362 (Fed. Cir. 1999); see also 35 U.S.C. § 271(d)(4) (1994). Section 271(d)(4) provides:
significant limitations on licenses, such as refusing to grant non-exclusive licenses to multiple researchers. 73

Even where research tool patent owners are amenable to licensing, the price demanded can represent a barrier to entry. When there is not yet any commercial product in existence, the research-tool patent owner and the tool user may have very different views about the proper economic valuation of the tool. 74 With increasing frequency, research tool patentees are demanding the payment of “reach-through royalties.” 75 Such royalties are computed as a share of the ultimate market value of some future commercial product to be developed with the tool, rather than the current market value of the research tool itself. 76 Although reach-through royalties are often attractive to biotech start-up firms because they minimize or eliminate up-front licensing costs, other potential licensees may object to them. 77 The NIH have recently issued a strongly-worded criticism of reach-through royalties, at least for tools licensed by or to NIH grantees. 78

No patent owner otherwise entitled to relief for infringement or contributory infringement of a patent shall be denied relief or deemed guilty of misuse or illegal extension of the patent right by reason of his having done one or more of the following: . . . (4) refused to license or use any rights to the patent.

Id.

73. Nelsen, supra note 26, at 1460–61 (criticizing exclusive licensing of receptor “targets” for high-throughput drug screening in situations where non-exclusive licensing “might better foster development”).

74. Flores, supra note 49, at 819 (contending that often “the promise of generating commercial value is remote and the up-front use fee that scientists are willing to pay without extended negotiation is correspondingly low”); cf. Josh Lerner & Robert P. Merges, The Control of Technology Alliances: An Empirical Analysis of the Biotechnology Industry, 46 J. INDUS. ECON. 125, 126 (1998) (describing early-stage biotechnology research and development efforts as “highly complex and uncertain” and characterized by great difficulty in “specify[ing] the features of the product to be developed”).


76. For example, assume that Firm X is the supplier of a transgenic mouse useful in pharmaceutical research. Firm Y wants to use the mouse in laboratory testing in order to develop a new drug. By demanding a reach-through royalty, Firm X seeks to obtain from Firm Y some portion of future revenues from marketplace sales of the end product, the new drug. See Jorge A. Goldstein, Research Tools and Reach Throughs, in PROCEEDINGS OF THE 17TH ANNUAL AMERICAN TYPE CULTURE COLLECTION BIOTECH PATENT & LICENSING FORUM (Sept. 23–26, 1999); see also Barton, supra note 75 (noting patenting of research tools by biotechnology entities and universities and efforts of these firms to obtain reach-through royalty provisions in licenses); Cohen, supra note 37, at 28 (reporting attempts of Chiron Corporation, owner of patents on hepatitis C viral protease enzyme, to obtain reach-through royalties based on future sales of protease inhibitors developed by firms that use Chiron’s enzyme to screen for these inhibitors).

77. Basing royalties on sales of a possible future end-product rather than the patented research tool itself creates the potential for greatly enhanced royalty income to the patentee, particularly if the licensee develops a product that is successful in the marketplace. Some scholars contend that this reach-through mechanism works to reduce the licensee’s incentives to innovate and develop

16
Experimental Use Exception to Patent Infringement

These difficulties in defining and accessing research tools suggest the need for an experimental use doctrine that would, if not completely exempt research and development uses of patented tools from infringement liability, at least permit the use of patented research tools without prior consent, so long as appropriate compensation were subsequently paid to the research tool patent owner. As illustrated in the next Part, however, the current formulation of the experimental use doctrine in the United States is far too narrow to permit such activity and must therefore be expanded.

III. LIMITED RECOGNITION OF A “TRULY NARROW” EXPERIMENTAL USE DOCTRINE IN THE UNITED STATES

For almost 200 years U.S. patent jurisprudence has paid homage to the concept of an exception (or exemption) from infringement liability for unlicensed or otherwise unauthorized uses of patented inventions carried out for a research or experimental purpose. In practice, however, the experimental use doctrine has rarely been applied in favor of an accused commercially successful products. E.g., Barton, supra note 15, at 461 (asserting that reach-through royalties “directly reduce the follow-on inventor’s incentive to develop new technologies”). On the other hand, reach-through royalties may facilitate increased research and development activity because they allow authorized access to a research tool at little or no up-front cost.

78. NIH Principles and Guidelines, supra note 26, at 72,091 (finding that reach-through royalties “contribute not only to specific restriction of access to subsequent tools arising out of the NIH-funded work, but also to the general proliferation of multiple ties and competing interests that is the source of the current access problems”); Barton, supra note 75, at 350 (noting opposition of pharmaceutical firms to reach-through royalties). Patent bar groups have also identified concerns with demands for reach-through royalties in consensual licensing transactions of research tools. McConathy & Weber, supra note 40, at 178 (noting that reach-through license “increases the royalty burden, requires full disclosure of the licensee’s intended use of the technology, and may so enlarge the stacking provisions that the entire project is killed”). But see Flores, supra note 49, at 819–20 (criticizing NIH Proposed Guidelines for Recipients of NIH Research Grants and Contracts on Obtaining and Disseminating Biomedical Research Resources; Request for Comments, 64 Fed. Reg. 28,205 (May 25, 1999), as overly broad in attempting to prohibit reach-through royalties in licenses to commercial entities as well as to academics). The propriety of reach-through royalties is further addressed infra notes 286–305 and accompanying text.

79. Part VII infra details this proposed liability rule.

80. Although the U.S. case law has generally referred to an experimental use “exception,” the term “exemption” is more precise and will be used herein.

81. The terms “research use” and “experimental use” appear interchangeably in U.S. case law and literature that recognizes an exemption from infringement liability.

infringer. Although Congress enacted legislation in 1984 that created a safe harbor for generic drug manufacturers to test patented drugs for purposes of preparing FDA bioequivalency data, that narrow and specific statutory safe harbor is not the subject of this Article. Rather this Part focuses on what remains of the common law doctrine of experimental use and suggests that the burgeoning nonprofit and for-profit collaborative environment of biotechnology research and development calls for a fundamental rethinking of this doctrine’s very limited contours.

In the great majority of cases, U.S. courts have recognized the experimental use exemption as doctrinally legitimate but found it inapplicable to the facts of the particular cases before them. The accused infringer’s use of a patented invention, even for a socially-beneficial purpose such as scientific research, typically has been labeled a commercial or profit-making endeavor and, therefore, ineligible for the doctrine’s protections. The courts have considered even a minimal flavor of commerciality sufficient to take the accused activity outside the realm of protected experimental or research use.

85. Although the U.S. Patent Act includes among the enumerated defenses to patent infringement the “absence of liability for infringement,” 35 U.S.C. § 282(1), no court has pointed to that statutory provision as encompassing the experimental use doctrine. E.g., Pfizer, Inc. v. Int’l Rectifier Corp., 217 U.S.P.Q. 157, 160 (C.D. Cal. 1982) (“There is nothing in the Patent Act of 1952 which even mentions any experimental use exception. The experimental use exception is . . . purely case law.”). The commentary by a co-author of the 1952 codification of the Act does not mention the experimental use doctrine. P.J. Federico, Commentary on the New Patent Act, 35 U.S.C.A. 1 (1954 ed.) reprinted in 75 J. PAT. & TRADEMARK Off. Soc’y 161, 215 (1993) (stating that item one of § 282 “would include the defenses such as that the patented invention has not been made, used or sold by the defendant; license; and equitable defenses such as laches, estoppel and unclean hands”). Thus, the general experimental use doctrine, as opposed to the explicit regulatory data-gathering safe harbor of 35 U.S.C. § 271(e), is one of common law rather than statute.
86. CHISUM, supra note 82, § 16.03[1] (characterizing defense as limited to those acts conducted “solely for an experimental or other nonprofit purpose”).
87. See, e.g., Pfizer, 217 U.S.P.Q. at 161 (“The underlying rule of permissible experimental use demands there must be no intended commercial use of the patented article, none whatsoever, if the exception is to be recognized at all.”).

That almost all of the reported cases invoking the experimental use defense involved some degree of commercialization is not surprising because litigation costs will most likely deter enforcement where a defendant’s use is merely for amusement or philosophical inquiry, not involving some significant sales diversion or other profit-taking from the patentee. The U.S. Supreme Court’s decision in Beedle v. Bennett, 122 U.S. 71 (1887), is a rare exception to this general rule. The plaintiff in Beedle held a patent on a driven well for drawing water from the ground. Id. at 72. The defendant, without
A. United States Common Law Evolution of a “Truly Narrow” Experimental Use Exemption

1. Early Development: Commercial Intent Prohibited

The experimental use exemption from patent infringement liability originated in the opinions of Justice Joseph Story, one of the country’s early leading intellectual property jurists. Justice Story’s first and most commonly cited case on the subject is Whittemore v. Cutter, involving an alleged infringement of a patent directed to a machine for making cards. In discussing the trial court’s instruction on infringement to the jury, Justice Story opined that:

It could never have been the intention of the legislature to punish a man, who constructed such a machine merely for philosophical experiments, or for the purpose of ascertaining the sufficiency of the machine to produce its described effects.

In Sawin v. Guild, Justice Story cited his earlier decision in Whittemore as establishing that patent infringement must concern:

the making [of the invention] with an intent to use for profit, and not for the mere purpose of philosophical experiment, or to ascertain the
verity and exactness of the specification . . . . In other words, that the making must be with an intent to infringe the patent-right, and deprive the owner of the lawful rewards of his discovery. 92

In Sawin, Justice Story thus makes the accused infringer’s commercial intent the hallmark of liability, and the absence of such motive as the prerequisite for exemption under the experimental use doctrine. 93 Justice Story’s dichotomy was adopted in subsequent decisions, 94 and it remains the rule today. 95

Several explanations have been posited for Justice Story’s profit-dispositive view. Suits for patent infringement were brought in Justice Story’s era as actions for trespass on the case, a tort theory requiring a showing of wrongful intent or negligence. 96 Justice Story simply may have recognized the absence of such intent. Alternatively, Justice Story may have relied on the doctrine of de minimis non curat lex ("the law takes no account of trifles"). 97 One commentator suggests that Justice Story was merely invoking the common law principle of injuria absque damno (wrong without damage), 98 which applies when a party’s rights have been violated but the party has not suffered any legally recognizable damage. This explanation seems unlikely, however, in view of Justice Story’s statement in Whittemore that “where the law gives an action for a particular act, the doing of that act imports itself a damage to the party. Every violation of a right imports some damage, and if none other be proved, the law allows a nominal damage.” 99

Professor William Robinson in his famous 1890 treatise seconded Justice Story’s “intent to deprive the patentee of his rewards” theory, titling

92. Id. at 555.

93. It is unclear whether Justice Story would view an infringement that is unknowing or “innocent” (i.e., where the accused infringer was not aware of the plaintiff’s patent) as equally disqualified from any experimental use exemption as a willful infringement, so long as in either case the accused infringer intended to make a profit.


97. Radio Corp. of Am. v. Andrea, 15 F. Supp. 685, 687 (D.C.N.Y. 1936) (characterizing as understandable view that “the law, not concerning itself with trifles, would ignore a mere casual appropriation [of a patented invention] for amusement or even scientific purpose”), modified on other grounds, 90 F.2d 612 (2d Cir. 1937).

98. Bee, supra note 96, at 365.

99. 29 F. Cas. 1120, 1121 (C.C.D. Mass 1813) (No. 17,600).
section 898 of the treatise with the phrase that “no act [is] an infringement unless it affects the pecuniary interests of the owner of the patented invention.” \(^{100}\) In Robinson’s view, these interests, or “emoluments,” are not affected when the accused infringer’s use of the patented invention “produce[s] no pecuniary result.” \(^{101}\) Robinson would have exempted an invention “made or used as an experiment, whether for the gratification of scientific tastes, or for curiosity, or for amusement.” \(^{102}\) In these cases, Robinson contended, “the interests of the patentee are not antagonized, the sole effect being of an intellectual character in the promotion of the employer’s knowledge or the relaxation afforded to his mind.” \(^{103}\)

In the years following Justice Story’s decision in Whittemore, a number of federal trial and appellate courts recognized the existence of an experimental use exemption from infringement liability, but following the Story-Robinson rule generally refused to apply the exemption because the cases involved commercial activity or some degree of profit motive. \(^{104}\) The defense of experimental use was limited to those rare instances where the accused infringer’s use of a patented invention was for “the sole purpose of gratifying a philosophical taste, or curiosity, or for mere amusement.” \(^{105}\)

---

100. 3 W. ROBINSON, THE LAW OF PATENTS FOR USEFUL INVENTIONS § 898 (1890).
101. Id.
102. Id.
103. Id.
104. CHISUM, supra note 82, § 16.03[1]. Although the opinion in Deuterium Corp. v. United States, 19 Cl. Ct. 624 (1990) (Rader, J.), seemed to recognize the reality of profit motive when it described the experimental use exemption as “protect[ing] an individual making unauthorized use of a patent (a potential infringer) during tests seeking advancement or commercialization of the patented teaching,” id. at 632 (emphasis added), the court declined to apply the exemption in that case because the accused steam cleaning, pilot plant-scale demonstration project represented “not strictly intellectual experimentation, but development of technology and processes for commercial applications,” id. at 633.
105. Poppenhusen v. Falke, 19 F. Cas. 1048, 1049 (C.C.N.Y. 1861) (No. 11,279). One of the few cases in which the experimental use doctrine was applied in favor of an accused infringer is Ruth v. Stearns-Roger Manufacturing, Co., 13 F. Supp. 697, 703, 713 (D. Colo. 1935) (exempting from infringement flotation machines and their parts used by Colorado School of Mines, which were “all used in the laboratory and . . . cut up and changed from day to day”), rev’d on other grounds, 87 F.2d 35 (10th Cir. 1936). See also Chesterfield v. United States, 159 F. Supp. 371, 375–76 (Cl. Cl. 1958) (stating in dicta that if patent in suit was not invalid, then defendant’s use of patented metallic alloy was for testing and experimentation and thus not infringing).
2. Federal Circuit Development: Toward a Restrictive Application of Experimental Use Exemption

The key case addressing the experimental use exemption in the modern era is *Roche Products v. Bolar Pharmaceutical Company*, which has attracted significant scholarly commentary. The *Roche* decision led Congress to enact 35 U.S.C. § 271(e), which legislatively overruled part, but not all, of *Roche* by creating a safe harbor for using patented inventions for regulatory data gathering. The “residue” of *Roche*, in terms of what remains of the common law exemption from patent infringement for experimental uses, is a focal point of this Article.

*Roche* sued *Bolar* for infringing *Roche’s* U.S. patents on the active ingredients of Dalmane, *Roche’s* commercially successful prescription sleeping aid. *Roche’s* patent was due to expire in early 1984. *Bolar*, planning to market a generic version of Dalmane as soon as the drug went off-patent, chose to begin testing and gathering the data that would be required for submission of a generic equivalent drug application to the FDA before *Roche’s* patent expired. *Bolar* obtained the drug from foreign sources and began using it in tests in 1983. The New York federal district court held that *Bolar’s* use of the patented compound for federally mandated testing was experimental and de minimis and thus not infringement.

---

108. See infra Part IV.B.
109. *Roche*, 733 F.2d at 860. *Roche* was assignee of U.S. Patent No. 3,299,053. One of the chemical compounds claimed in the ’053 patent is flurazepam hydrochloride, the active ingredient in *Roche’s* “Dalmane” brand sleeping pill. *Id.*
110. *Id.*
111. *Id.*
112. *Id.* Specifically, *Bolar* formed the drug into “dosage form capsules” with which it commenced testing to obtain stability data, dissolution rates, bioequivalency studies, and blood serum studies needed for an FDA New Drug Application. See *id.*
Experimental Use Exception to Patent Infringement

The Federal Circuit reversed, holding that Bolar was liable for infringement.114 The Federal Circuit explained that on its face, § 271(a) of the Patent Act prohibits any “making, using or selling.”115 A mere “use” that does not result in a sale is still actionable.116 Thus, “the patentee does not need to have any evidence of damage or lost sales to bring an infringement action.”117

Although the Federal Circuit held Bolar’s use to be infringing, it stopped short of adopting a definition of § 271(a) “use” in its “utmost possible

114. Roche, 733 F.2d at 867.
115. Id. at 861 (emphasis added).
116. Id.
117. Id. The Roche court’s stance would thus seem to have evolved away from the Story-Robinson view that actionable harm to the patentee’s pocketbook is required for patent infringement.

The historical origin of the inclusion of “use” of a patented invention as one of the U.S. patent owner’s exclusionary rights is uncertain. In post-Elizabethan England, patents were recognized as an explicit exception to the general rule against monopolies, but the right granted by the Crown was positively defined as the patent owner’s affirmative right to “work” or “make” the invention:

Provided also and be it declared and enacted, that any declaration [against monopolies] before mentioned shall not extend to any letters patent and grants of privilege for the term of fourteen years or under, hereafter to be made of the sole working or making of any manner of new manufactures within this Realm, to the true and first inventor and inventors of such manufactures, which others at the time of making such letters patent and grants shall not use, so as also they be not contrary to the law or mischievous to the State, by raising prices of commodities at home, or hurt of trade, or generally inconvenient . . . .

English Statute of Monopolies, 1623, § 6 (emphasis added), reprinted in EDITH TILTON PENROSE, THE ECONOMICS OF THE INTERNATIONAL PATENT SYSTEM 7 (1951). “Use” is of course an explicit part of the qualification that the patented invention be one “which others at the time of making such letters patents and grants shall not use;” see id. (emphasis added), but “use” in this manner merely denoted a novelty requirement and did not form part of the exclusive right granted by the patent. The recited exclusive right to “work” the patented invention may have been intended as synonymous with an exclusive right to “use,” but that interpretation is inconsistent with the separate presence of the terms “work” and “use” in the same statutory provision.

The early U.S. patent system was strongly influenced by that of England. In re Bergy, 596 F.2d 952, 958 n.2 (C.C.P.A. 1979) (Rich, J.). But in contrast with Section 6 of the Statute of Monopolies, the first federal patent statute in the United States explicitly included “use” within the definition of infringement. Act of 1790, § 1 (defining patent grant as “sole and exclusive right and liberty of making, constructing, using and vending to others to be used” the patented invention) (emphasis added), reprinted in 9 LIPSComB’S WALKER ON PATENTS app. 2, at 9 (3d ed. 1990). The broader definition of the patent right through the addition of a “use” right may represent the first foreshadowing of the broader treatment of patents generally in the United States, as compared to foreign systems.

Alternatively, the presence of “use” in the list of a patentee’s exclusive rights may have been intended to specifically refer to patents on processes or methods (then termed “arts”), which are used rather than made or sold. If this assumption is correct, it does not explain the extension of the “use” right to the other statutory categories of patentable subject matter, such as machines, manufactures, and compositions of matter.
The court cited Justice Story’s opinion in *Whittemore* as the origin of the common law experimental use doctrine and recognized that one of the Federal Circuit’s predecessor courts, the Court of Claims, had considered the doctrine in several decisions.

The Federal Circuit in *Roche* ultimately interpreted this earlier authority as having established a very restricted experimental use defense, but refused to apply the defense to Bolar’s clearly commercial activity:

> [W]e hold the experimental use exception to be truly narrow, and we will not expand it under the present circumstances. Bolar’s argument that the experimental use rule deserves a broad construction is not justified. . . . Bolar’s intended “experimental” use is solely for business reasons and not for amusement, to satisfy idle curiosity, or for strictly philosophical inquiry. . . .

> [Bolar’s] unlicensed experiments conducted with a view to the adaptation of the patented invention to the experimenter’s business is a violation of the rights of the patentee to exclude others from using his patented invention. It is obvious here that it is a misnomer to call the intended use de minimis. It is no trifle in its economic effect on the parties even if the quantity used is small. It is no dilettante affair such as Justice Story envisioned. We cannot construe the experimental use rule so broadly as to allow a violation of the patent laws in the guise of “scientific inquiry,” when that inquiry has definite, cognizable, and not insubstantial commercial purposes.

Thus, after *Roche*, scientists engaged in research and development having more than negligible commercial purpose could no longer rely on the experimental use doctrine to exempt their experiments from patent infringement liability.

---

118. *Roche*, 733 F.2d at 861.
119. *Id.* at 862.
120. The Federal Circuit in *Roche* characterized *Pitcairn v. United States*, 547 F.2d 1106 (Ct. Cl. 1976) as “the most persuasive” of the Court of Claims cases concerning the experimental use defense. *Roche*, 733 F.2d at 863; see also *Pitcairn*, 547 F.2d at 1124–26 (recognizing that experimental use may be defense to infringement and that government’s “tests, demonstrations, and experiments” were not eligible for defense because they were “in keeping with the legitimate business of the using agency”).
121. *Roche*, 733 F.2d at 863.
B. The Hatch-Waxman Act: A Limited Exemption for Regulatory Data Gathering

The generic-drug industry quickly and successfully moved to devise an escape route from *Roche.* The industry argued to Congress that if a generic-drug manufacturer had to wait to begin testing of an equivalent drug until after the relevant patent had expired, the patentee of the branded drug would receive a de facto extension of the patent term. Such an extension was contrary to the public’s interest in obtaining lower-cost drugs as soon as possible and, of course, contrary to the profit-maximizing goals of the generics.

The generic-drug companies’ lobbying efforts succeeded. Congress enacted the Drug Price Competition and Patent Term Restoration Act of 1984, popularly known as the “Hatch-Waxman Act.” The legislation added, inter alia, § 271(e) to the Patent Act, which provides in pertinent part that the use of a patented invention solely for purposes reasonably related to gathering data to support an FDA application for generic versions of previously approved drugs (i.e., an Abbreviated New Drug Application) is not patent infringement.

The U.S. Supreme Court subsequently interpreted § 271(e) as broad enough to encompass not only regulatory data gathering on pharmaceuticals, but also the comparable testing of medical devices. The Court in *Eli Lilly & Co. v. Medtronic, Inc.* limited its review of the Federal Circuit’s decision in *Roche* to the Circuit’s refusal to exempt Bolar’s particular testing for generic-drug equivalency, however. The Federal Circuit’s recognition in *Roche* of the less-specific common law exemption for experimental uses of patented inventions was not addressed in *Eli Lilly,* nor any subsequent U.S. Supreme Court decision.

---

122. See *Eli Lilly & Co. v. Medtronic, Inc.*, 872 F.2d 402, 404 (Fed. Cir. 1989).
123. *Id.* at 405.
124. *Id.* at 404–05.
126. A generic-drug manufacturer submits an Abbreviated New Drug Application to the FDA in order to seek expedited approval of the generic version of a “listed drug” (one previously approved by the FDA). *Bayer AG v. Elan Pharm. Research Corp.*, 212 F.3d 1241, 1244 (Fed. Cir. 2000). The generic drug must be the bioequivalent of the listed drug if an Abbreviated New Drug Application is used. *Id.*
129. *Id.* at 665.
There are valid reasons to conclude that the common law exemption for experimental use survived Congress’s enactment of 35 U.S.C. § 271(e)’s safe harbor, albeit in narrow form. In responding to the generic manufacturers’ concerns, it is reasonable to believe that Congress intended to overrule the Roche holding only insofar as it impacted the generic manufacturers—limited to unlicensed use of an invention for purposes of gathering data for the FDA—but not to wipe away Roche’s broader recognition of the common law experimental use exemption. The legislative history of the Hatch-Waxman Act makes no mention of the common law doctrine discussed in Roche. It addresses no “use” of patented inventions other than for the purpose of regulatory data gathering. Nor does the legislative history indicate that by enacting § 271(e), Congress intended to fill the field on this issue and preempt any and all common law doctrines related to experimental use.

Enactment of the Hatch-Waxman Act supports, rather than detracts from, a broadened interpretation of the common law experimental use exemption that does not turn solely on the commerciality of the accused infringer’s use. Bolar’s testing for purposes of gathering data needed to expedite the FDA approval of its generic equivalent of Dalmane was clearly profit-driven activity. Congress’s exemption from liability of this particular form of non-consensual, commercial use of a patented invention should be viewed as supportive of, rather than in opposition to, the broader proposition that certain unlicensed uses of patented inventions should be exempted from liability. The legislative history does not indicate that by enacting § 271(e) Congress meant to exempt one particular variety of research use from infringement liability while forever excluding all others. That Congress’s intent was not so limited is also evidenced by subsequent


The legislation was narrowly tailored to uses of patented products that are necessary for data-gathering for regulatory approval. H.R. REP. No. 98-857, pt. 1, at 14–15 (1984), reprinted in 1984 U.S.C.C.A.N. 2647, 2647–48 (stating that purposes of this legislation are “to make available more low cost generic drugs and to create a new incentive for increased expenditures for research and development of certain products which are subject to premarket approval”).

131. But see id. at 30, reprinted in 1984 U.S.C.C.A.N. at 2714 (quoting legislative-history conclusion that “[j]ust as we have recognized the doctrine of fair use in copyright, it is appropriate to create a similar mechanism in the patent law. That is all this bill does.”). While a positive justification for passage of the legislation, this statement does not purport to define or broaden its scope beyond the particular context of the exemption for regulatory data gathering purposes.
legislative proposals to statutorily implement a second type of experimental use provision in the Patent Act. Section 271(e)’s narrow purview merely reflects the successful lobbying of one particular industry: the generic-drug manufacturers. Absent lobbying pressure from other sectors of the patenting community, Congress would have had no reason or motivation to enact a broader safe harbor at the time of passage of the Hatch-Waxman Act. The fact that § 271(e) is limited to unlicensed use of patented inventions for purposes of gathering FDA data does not refute a broadened interpretation of the common law experimental use doctrine.

C. Whither the Residue of Roche?

Following its 1984 decision in *Roche*, the Federal Circuit did not revisit the common law experimental use doctrine for more than fifteen years. In *Embrex, Inc. v. Service Engineering Corp.*, decided in June 2000, the Federal Circuit re-confirmed the existence of the common law doctrine, but refused to apply it to the accused infringer’s “commercial” activity. *Embrex*’s patent in suit was directed to a method of inoculating chicks against diseases while still in ovo, i.e., before hatching. The claimed
method required administration of a vaccine within the “region defined by either the amnion or the yolk sac.” 136 Accused infringer Service Engineering retained scientific consultants to help it design around the Embrex patents, but was unsuccessful in avoiding infringement; Service Engineering’s scientists were unable to prevent injection into the claimed amnion or yolk region. 137 A jury found that Service Engineering had willfully infringed, and the district court denied Service Engineering’s motion for judgment as a matter of law. 138

On appeal, the Federal Circuit rejected Service Engineering’s argument that the accused tests were merely experimental and thus exempt from infringement liability under the common law experimental use doctrine. 139 Citing Roche, the Federal Circuit reiterated that it construes the experimental use exception “very narrowly.” 140 The court also acknowledged that “[b]inding precedent” recognizes a “narrow defense to infringement performed ‘for amusement, to satisfy idle curiosity, or for strictly philosophical inquiry.’” 141 Service Engineering’s tests did not qualify for the defense, the Federal Circuit concluded, in view of the district court’s findings that the tests were performed “expressly for commercial purposes,” 142 and were chiefly conducted by Service Engineering in order to sell its own in ovo injection machines to potential customers. 143

The Federal Circuit’s recognition of the common law experimental use doctrine in Embrex clarifies that Congress did not overrule that portion of Roche in which the Federal Circuit first recognized the common law doctrine. 144 Even for unlicensed uses of patented inventions that do not fall

136 Id.
137 Id. at 1346–47.
138 Id. at 1347.
139 Id. at 1349.
140 Id. (citing Roche Prods., Inc. v. Bolar Pharm. Co., 733 F.2d 858, 863 (Fed. Cir. 1984)).
141 Id. at 1349 (citing Roche Prods., Inc. v. Bolar Pharm. Co., 733 F.2d 858, 863 (Fed. Cir. 1984); Pitcairn v. United States, 547 F.2d 1106 (Ct. Cl. 1978)).
142 Id.
143 Id.
144 Notably, the Embrex court cites Roche as having been “superseded on other grounds by 35 U.S.C. § 271(e) (1994),” Embrex, 216 F.3d at 1349 (emphasis added). This characterization of Roche indicates that the Federal Circuit does not view Congress’s 1984 enactment of 35 U.S.C. § 271(e)’s safe harbor for regulatory data gathering use as having overruled Roche in its entirety; i.e., as having overruled those statements in Roche respecting the common law experimental use doctrine. In other words, Embrex establishes that the post-Roche 1984 passage of the Hatch-Waxman Act did not repeal sub silentio the common law experimental use doctrine.
within § 271(e)’s safe harbor for regulatory data-gathering, the common law experimental use defense remains available, at least theoretically. As illustrated by Service Engineering’s failure to prevail under the experimental use defense in *Embrex*, however, the Federal Circuit’s extremely narrow interpretation of the doctrine virtually precludes successful reliance by any commercial enterprise. Few users of patented methodologies will be able to mount a more compelling case for exemption from infringement liability than Service Engineering did; rather than proceeding to use Embrex’s patented method without permission, Service Engineering retained outside scientists for the commendable purpose of attempting to design around Embrex’s patent by developing an alternative method sufficiently different to avoid Embrex’s claims. This kind of activity should be facilitated, rather than penalized, by patent law.  

Federal Circuit Judge Rader concurred in *Embrex*, but wrote a separate opinion contending that the experimental use defense is no longer viable following the U.S. Supreme Court’s 1997 statement in *Warner-Jenkinson Co. v. Hilton Davis Chemical Co.* that “[a]pplication of the doctrine of equivalents . . . is akin to determining literal infringement, and neither requires proof of intent.” In Judge Rader’s view, “[t]he Supreme Court’s recent reiteration that infringement does not depend on the intent underlying the allegedly infringing conduct . . . precludes any further experimental use defense, even in the extraordinarily narrow form recognized in *Roche*.” According to Judge Rader, when “wholly non-commercial” infringement is proven, “the damage computation process provides full flexibility for courts to preclude large (or perhaps any) awards for minimal infringement.”

Contrary to Judge Rader’s concurrence, the U.S. Supreme Court in *Warner-Jenkinson* did not create new law nor change the law with respect to the common law experimental use doctrine. The accused infringer in *Warner-Jenkinson* did not rely on the experimental use doctrine, nor did the case involve the use of research tools; both parties were commercial

145. Rebecca S. Eisenberg, *Patents and the Progress of Science: Exclusive Rights and Experimental Use*, 56 U. Chi. L. Rev. 1017, 1078 (1989) (contending that patent holder should not be able to enjoin researcher who uses patented invention in course of attempting to develop alternative means of achieving same purpose).

146. 520 U.S. 17 (1997).


148. Id. (Rader, J., concurring).

149. Id. at 1352 (Rader, J., concurring).
manufacturers of purified dyes. The accused infringer’s dye purification method, operating at a pH of 5, infringed the patent owner’s claimed method for purifying dyes at a pH “from approximately 6.0 to 9.0.” The accused infringer asserted that it independently developed its own process with no knowledge of the plaintiff’s patent. On this basis, the accused infringer contended that the patent owner should be required to establish, as an equitable threshold factor, an “intent to copy” on the part of the accused infringer, before the patent owner could assert infringement under a doctrine of equivalents theory. The U.S. Supreme Court rejected this argument, explaining that because the “essential predicate” of the doctrine of equivalents is the identity between the claimed invention and its equivalent, there is no basis for treating an infringing equivalent any differently from a device that literally infringes. The Court concluded that “intent plays no role in the application of the doctrine of equivalents.” Thus, Warner-Jenkinson merely establishes that an accused infringer need not be aware of the plaintiff’s patent in order to be liable for infringing it. In no way does Warner-Jenkinson hold or suggest that an accused infringer’s experimental or research purpose is irrelevant to the question of infringement liability or remedy. Moreover, an accused infringer’s purpose or intent has been held relevant in other aspects of the infringement determination.

Judge Rader’s suggestion that accused infringers who assert the experimental use defense are adequately protected by the courts’ flexibility in adjusting the damages award for “minimal or non-commercial

---

151. Id. at 22–23.
152. Id. at 23 (noting that Warner-Jenkinson did not learn of Hilton Davis’s patent until “after [Warner-Jenkinson] had begun commercial use of its ultrafiltration process”).
153. Id. at 34–36.
154. Id. at 35.
155. Id. at 36.
156. For example, in determining whether an infringement was willful, which may lead to an award of enhanced damages under 35 U.S.C. § 284 (1994), the Federal Circuit looks to whether the accused infringer was aware of the plaintiff’s patent, and once aware, whether the accused infringer proceeded to act with due care and in good faith (e.g., obtained a competent non-infringement or invalidity opinion of counsel before proceeding with manufacture). E.g., Read Corp. v. Portec, Inc., 970 F.2d 816, 826–31 (Fed. Cir. 1992). More recently, the U.S. Supreme Court suggested that the nature and purpose of a state government’s patent infringement (i.e., whether it was merely “negligent” rather than “intentional or reckless”) governs whether the state’s infringement violated the Due Process Clause of the Fourteenth Amendment. Fla. Prepaid Postsecondary Educ. Expense Bd. v. Coll. Sav. Bank, 527 U.S. 627, 645 (1999).
Experimental Use Exception to Patent Infringement

infringement” is also problematic. Although compensatory damages might be reduced to a nominal level under Judge Rader’s proposed framework, a research tool user would nevertheless remain subject to an injunction against any further use of the patented tool. Moreover, compelling the research tool user to litigate the question of liability and damages in the hopes that damages might be reduced to a nominal level ignores the tremendous costs of patent litigation.

D. District Court Decisions: Continued Vitality of the Common Law Exemption

Several lower courts have had occasion to address the exemption following Roche. The emergent consensus is that Roche’s recognition of a general common law exemption continues to have vitality after the enactment of the Hatch-Waxman Act.

The clearest judicial support for a viable exemption is the Claims Court decision in Deuterium Corp. v. United States. Then-Claims Court Judge


158. The different judicial viewpoints expressed in Embrex with respect to the viability of the common law experimental use doctrine will likely be fleshed out in future Federal Circuit decisions as pending litigation over unlicensed use of research tools wends its way toward appeal. E.g., Cohen, supra note 37, at 28 (describing lawsuit filed in July 1998 by Chiron Corporation, owner of patents directed to hepatitis C viral protease enzyme, against Vertex, corporation involved in research and development efforts to find drugs that block enzyme, in which Vertex has invoked experimental use exemption as defense).

The experimental use defense is also at issue in Integra v. Merck, a patent infringement lawsuit currently pending in the Southern District of California. See Integra LifeSciences I Ltd. v. Merck KgA, 50 U.S.P.Q.2d 1846 (S.D. Cal. 1999) (granting summary judgment on single claim of one of five patents in suit). Merck, a German corporation, is a co-defendant in the action with the Scripps Research Institute, a nonprofit, public benefit corporation, and Dr. David A. Cheresh, a tenured professor in Scripps’s Immunology Department. Id. at 1847. Scripps and Merck entered into a research support agreement in 1988 to fund Dr. Cheresh’s study of “integrins,” proteins that serve as receptors on the surface of certain living cells. Id. As part of the agreement, Scripps granted Merck an option to license any inventions arising from Dr. Cheresh’s work. Id. When Dr. Cheresh reported his findings that the growth of new blood vessels can be inhibited by binding various molecules to receptors on the surface of certain cells, Integra sued Cheresh, Scripps, and Merck in order to enjoin Cheresh’s further work on his discovery and to “collect damages for the underlying research.” Id. The defendants asserted that the experimental use doctrine shields them from infringement liability, but Integra contended that the defense was unavailable to them because Merck is a for-profit entity. E-mail Interview with William C. Rooklidge, counsel for defendants (May 25, 2000). The trial court granted judgement as a matter of law that the defendant’s pre-1995 activities were exempt under the experimental use doctrine. Id. The trial court refused to instruct the jury on the experimental use defense with respect to the defendants’ post-1994 activities; however, a jury found that those activities infringed Integra’s patents. Id. As of January 2001, the case was still pending before the trial court on post-trial motions. Id.

Rader, now a member of the Federal Circuit, cited the Robinson treatise as support for the doctrine and explained that “[a]lthough [the Hatch-Waxman Act] changed that narrow application of the doctrine affecting reporting requirements for federal drug laws, Congress did not disturb the Federal Circuit’s enunciation [in Roche] of the parameters of the experimental use exception.”

The view that the experimental use exemption remains valid was seconded by the federal district court in Giese v. Pierce Chemical, Co.161 The end users of Giese’s patented methods of detecting cancer cells were largely academic researchers.162 Giese sued two chemical companies as contributory and inducing infringers based on their acts of supplying the academic research institutions with kits of chemical reagents for use in Giese’s patented method.163 Although the trial court denied the defendant chemical companies’ motion for summary judgment of non-infringement because of unresolved factual disputes,164 it gave strong support to the notion that the experimental use doctrine survives. The Roche decision “establish[es] a restrictive definition of the traditional common law doctrine, but in no way eliminat[es] it in those cases which involve experimentation for ‘idle curiosity or for strictly philosophical inquiry.’”165

160. Id. at 632 n.14.
162. Id. at 35.
163. Id. at 34–35.
164. Id. at 37.
165. Id. at 36 (quoting Roche Prods., Inc. v. Bolar Pharm. Co., Inc. 733 F.2d 858, 863 (Fed. Cir. 1984)). The court in Giese further suggested, albeit in dicta, that if end users of a patented method are exempt from infringement liability because they are beneficiaries of the experimental use exemption, then their suppliers by definition cannot be liable as contributory infringers. Id. at 36. Non-liability would hold even where the suppliers, as in Giese, were decidedly for-profit enterprises. Id. This result certainly follows from application of the black-letter patent rule that there can be no contributory infringement without direct infringement. See Aro Mfg. Co. v. Convertible Top Replacement Co., 377 U.S. 476, 483 (1964); see also Ruth v. Stearns-Roger Mfg. Co., 13 F. Supp. 697, 703 (D. Colo. 1935) (holding that sale of parts for use in accused flotation machine did not constitute contributory infringement where machine was used in laboratory of Colorado School of Mines for experimental purposes), rev’d on other grounds, 87 F.2d 35 (10th Cir. 1936). Moreover, the pro-innovation policies underlying the recognition of a viable research use exemption would be stymied if the research user could not obtain necessary supplies from for-profit sources for fear that these sources would be subject to indirect liability under § 271(b) or (c) or both of the Patent Act.

The Giese court’s decision allowing suppliers of the direct infringer to share in the benefits of the experimental use exemption, a judge-made doctrine of equitable roots, thus seems to assume without deciding that the experimental use exemption is not a personal defense, limited to the individuals actually conducting the experimentation. This treatment contrasts the newly enacted “prior inventor defense,” which is personal to the defendant prior user and can only be licensed or transferred as part of a good faith transfer of the prior user’s entire line of business. See American Inventors Protection Act of
IV. IMPLICATIONS FOR ACADEMIC-INDUSTRIAL COLLABORATIVE RESEARCH AND DEVELOPMENT

Much of the burgeoning biotechnology research and development activity in the U.S. involves public-private partnerships, collaborations, joint ventures, sponsored research, and the like between nonprofit universities or research institutions and for-profit corporations.166 Many

---

1999, Pub. L. No. 106-113, 113 Stat. 1536 (amending Title 35, U.S.C., by adding new § 273 titled “Defense to Infringement Based on Earlier Inventor”). The potential for exploiting the exemption, thus interpreted, to absolve large commercial, profit-making suppliers of liability, whenever their customers are deemed to be conducting research or experimentation, looms large.

The contributory liability of a research-tool supplier is also at issue in the long-running patent litigation battle between Hoffman-La Roche, the Swiss corporate holder of patents on polymerase chain reaction (PCR) and Taq polymerase, the key reagent used in PCR, and Promega Corporation, a Wisconsin supplier of Taq to researchers across the United States. See supra notes 1–11 and accompanying text. Roche sued Promega in October 1992 for selling Taq in violation of the terms of Promega’s license granted by the now-defunct Cetus Corporation, the original owner of the patents. See Barinaga, supra note 7, at 1273. Roche reads that license as limiting Promega to sales of Taq for non-PCR purposes; Promega disputes this interpretation. Abate, supra note 3, at B1; Barinaga, supra note 7, at 1273; Rubenstein, supra note 2, at 28. Among other contentions, Hoffman-La Roche asserts that Promega is liable for contributory or inducing infringement of the PCR method patent based on Promega’s sales of Taq to more than 200 researchers whom Roche identified as direct infringers (though not named as parties to the lawsuit). See Rubenstein, supra note 2, at 28. One of the several patents involved in the dispute was held unenforceable in December 1999 for inequitable conduct. Hoffman-La Roche, Inc. v. Promega Corp., No. C-93-1748 VRW, 1999 WL 1797330 (N.D. Cal. Dec. 7, 1999) (holding all claims of U.S. Patent No. 4,889,818 unenforceable).

166. NIH Research Tools Report, supra note 24 (explaining that since mid-1970s “[b]iomedical researchers increasingly chose to collaborate with entrepreneurial companies that understood and valued basic science, or to leave academia and join these firms as founders or employees. Many biotechnology companies emerged with strong ties to the academic world.”); see also Florida, supra note 33, at 69 (reporting dramatic growth of joint university-industry research centers as evidenced by 1990 Carnegie Mellon University study of 1,056 such U.S. centers that received total funding of more than $4.12 billion); David E. Korn, Patent and Trade Secret Protection in University-Industry Research Relationships in Biotechnology, 24 HARV. J. LEGIS. 191, 191 (1987) (noting “rapid and substantial growth in biotechnology research conducted jointly by industry and academia” in recent years); id. at 222 (characterizing industry-sponsored research as “big business” for many U.S. research universities); Rai, supra note 37, at 110 (contending that legal developments in 1980s and 1990s, including passage of Bayh-Dole Act and pro-patent decisions by the Federal Circuit, have generated “large variety of academic-industrial relationships,” some of which resemble commercial joint ventures, and citing examples); Jeff Gerth & Sheryl Gay Stolberg, Drug Makers Reap Profits on Tax-Backed Research, N.Y. TIMES, Apr. 23, 2000, at 20 (reporting that since enactment of Bayh-Dole Act, universities and their scientists have become “more commercially oriented; many are spinning off their own biotech companies to develop their ideas.”).

A widely publicized biotechnology nonprofit and industry collaboration was formed in May 1998 when The Institute for Genomic Research (TIGR), a private, nonprofit research enterprise headed by
private corporations are shifting their research and development resources toward “external innovation,”167 building on the results of research conducted in universities rather than by their own in-house research staff.168 Private industry is estimated to have funded approximately twelve percent of university research and development activity in the life sciences in 1994,169 and that percentage has likely increased in subsequent years.170 Seventy percent of the funding for clinical trials of new drugs comes from private industry rather than the federal government.171

The 1980 enactment of the Bayh-Dole Act172 fundamentally enhanced the incentives for such collaborations; an explicit objective of the Act was

Doctor J. Craig Venter, joined forces with the Perkin-Elmer Corporation, a leading manufacturer of DNA sequencing devices. The collaboration announced plans to compete against the federal government’s Human Genome Project in attempting to sequence the entire human genome. Genetic Warfare, ECONOMIST, May 16, 1998, at 87.

167. Gerth & Stolberg, supra note 166, at 1, 20 (describing Bayh-Dole Act as having allowed private corporations engaged in drug development to “shift resources away from in-house research and development and towards outside collaborations, a strategy known as ‘external innovation’”).

168. Korn, supra note 166, at 196 (contending that firms in biotechnology industry are much more dependent on university research than are firms in other technologies); Philip H. Abelson, Editorial: Global Technology Competition, 277 SCIENCE 1587 (1997) (stating that only about six percent of U.S. industry research and development expenditures are “devoted to longer range directed basic research”); Richard C. Atkinson, Universities: At the Center of U.S. Research, 276 SCIENCE 1479 (1997) (describing changing pattern of U.S. corporate research and development in last decade, evolving from conduct of “significant basic research” in-house to current model in which corporations “build[] on the results of long-term university research” to solve “specific short-term problems”).

169. David Blumenthal et al., Relationships Between Academic Institutions and Industry in the Life Sciences—An Industry Study, 334 NEW ENG. J. MED. 368, 369 (1996). More than ninety percent of U.S. firms in the life sciences reported some relationship with academia in 1994; almost sixty percent of such firms supported research conducted by academic institutions. Id.

170. Cf. Abelson, supra note 168, at 1587 (reporting that total corporate support for U.S. research and development has increased by about twenty-seven percent since 1994); Nelsen, supra note 26, at 1460–61 (noting that following corporate downsizing of late 1980s and early 1990s, industry has displayed increased interest in establishing research partnerships with universities); Timothy Caulfield, The Commercialization of Human Genetics: Profits and Problems, MOLECULAR MED. TODAY, Apr. 1998, at 148 (reporting “substantial growth in the number of academic industry collaborations”); Charles F. Larson, The Boom in Industry Research, ISSUES SCI. & TECH., Summer 2000, at 27 (stating that industry support of all university research has grown from $1.45 billion in 1994 to $2.16 billion in 1999).

171. Thomas Bodenheimer, Health Policy Report: Uneasy Alliance—Clinical Investigators and the Pharmaceutical Industry, 342 NEW ENG. J. MED. 1539, 1539 (2000). “Clinical” research involves the design, conduct, and interpretation of drug testing on human subjects, and thus represents a later stage of the research process than “basic” research. See Marcia Angell, Is Academic Medicine for Sale?, 342 NEW ENG. J. MED. 1516, 1517 (2000) (arguing that in clinical research most technology development has already been completed and is simply being tested).

“to promote collaboration between commercial concerns and nonprofit organizations, including universities.”173 As a result of the Bayh-Dole Act, the interest level in university-industry partnerships is significantly greater than before the legislation’s enactment.174 The Bayh-Dole Act encourages universities to patent “subject inventions” made with federal government funds,175 and contemplates the grant of exclusive licenses under those patents to the universities’ private-sector partners.176 Predictably, university patenting activity177 and university revenues from patent licensing fees have increased dramatically since the enactment of the Bayh-Dole Act.178 So too have problems of access to patented research tools.179

173. 35 U.S.C. § 200 (1994) (“Policy and Objective”). The enactment of the Bayh-Dole Act has been viewed as part of a broader push by the Reagan administration in the early 1980s to maintain the United States’s stance as an international leader in technological development in the face of increasing imports of high-quality electronic goods from Japan. Gerth & Stolberg, supra note 166, at 1, 20.

174. See Angell, supra note 171, at 1517 (describing Bayh-Dole Act as “frequently invoked to justify the ubiquitous ties between academia and industry”); Nelsen, supra note 26, at 1460–61 (describing increasing interest in research partnerships between industry and universities, on the part of both groups, as having developed in parallel with formation of Bayh-Dole Act university licensing framework); see also Blumenthal et al., supra note 169, at 368 (describing greater acceptance in 1990s among life-sciences academics of relationships with industry than in 1980s).


176. Nelsen, supra note 26, at 1460–61 (stating university position that right to grant exclusive licenses is “key aspect” of Bayh-Dole Act because substantial risk taken by licensees to develop “early-stage technologies” justifies exclusivity); Rai, supra note 37, at 97 n.113; Editorial: The Patent Craze and Academia, 342 LANCET 1435, 1435 (1993) [hereinafter Editorial] (characterizing Bayh-Dole Act as having made it easier for universities and nonprofit research institutions to obtain and exclusively license patents); see also 35 U.S.C. § 203 (referring to “exclusive licensee” as entity that is subject to march-in rights). The federal government always retains a nonexclusive, paid-up license to practice a subject invention. 35 U.S.C. § 202(c)(4).

177. The number of patents annually assigned to U.S. academic institutions, including colleges, universities, and associations thereof, has increased dramatically from 1984, when 551 utility patents were issued to these institutions, to 1997, when the number of patents issued to U.S. academic institutions had increased to 2436. Technology Assessment and Forecast Report, U.S. Colleges and Universities–Utility Patent Grants 1969–1997 5 (Sept. 1998), at http://www.uspto.gov/web/offices/ac/ido/oeip/taf/univ_97.pdf. Most of this patenting activity was in the life sciences. See id. at 6 (listing five biotechnological and chemical patent classes (Classes 800, 435, 530, 536, and 424) in six classes of highest university patenting activity). The 158 universities surveyed by the Association of University Technology Managers (AUTM) reported patent application filings of more than 6000 in the year 1997 alone. See Florida, supra note 33, at 68.

178. See Florida, supra note 33, at 68 (describing universities’ increasing focus on technology licensing to generate income and reporting that approximately 3000 licenses granted by U.S. universities to industry in 1998 generated approximately $500 million in royalties); see also Gerth & Stolberg, supra note 166, at 1, 20 (reporting that Bayh-Dole Act represents “a windfall” for universities).

179. See Flores, supra note 49, at 819 (reporting “[r]ising frustration among scientists about difficulties in accessing critical research tools”); Nelsen, supra note 26, at 1460–61 (listing “[r]estricted
Under the traditional “narrow” view of the experimental use exemption in the United States as set forth in *Roche*, this type of collaborative research and development generally does not qualify for exemption. It is difficult to posit a collaboration involving a for-profit firm as having anything other than “definite, cognizable, and not insubstantial commercial purposes.” Indeed, the Bayh-Dole Act expressly promotes the commercialization of inventions made by universities.

This Article contends that the *Roche* treatment of “pure” experimentation for “philosophical” purposes versus “commercialization” as two polar extremes is no longer supportable. A better way to view such use is as overlapping regions on a continuum of experimental use. The sharp distinctions drawn by the *Roche* court have blurred dramatically in the last twenty years. The fundamental changes to the U.S. research and technology sector, brought about by the Bayh-Dole Act and other legal, technical and marketplace factors, are breaking down traditional barriers between academic research and for-profit commercialization. Research tools are at the heart of the research and commercialization overlap. The experimental use doctrine requires appropriate flexibility to reflect this changing research and development landscape.

The public policies promoted by choosing to exempt “philosophical” research from liability while denying the benefits of the exemption to innovation having the slightest “commercial” flavor are suspect. Society benefits from new therapeutic and diagnostic products, whether or not they availability or delays in exchange of ‘research tools’ (such as vectors or transgenic mice) in biological research” as unresolved problem facing university technology-transfer management).

180. *Roche Prods., Inc. v. Bolar Pharm. Co.*, 733 F.2d 858, 863 (Fed. Cir. 1984). This would appear to hold even when industry funds “basic” rather than “applied” research. Cf. Eisenberg, *supra* note 37, at 1023 n.26 (contending that “[a]cademic researchers whose work is funded by industry are likely to be motivated by ‘philosophical’ and ‘commercial’ interests at the same time”).


182. Dunn, *supra* note 27, at B3 (“Defining when research becomes commercial is tricky, given that an experiment done in a university lab today may be the founding idea for a biotech company tomorrow.”).

183. See Nelsen, *supra* note 26, at 1460–61 (describing drive to balance federal budget and decline of communism (resulting in fewer military research expenditures) as two key events leading to reduced federal government funding of research and development, and contending that this reduction in government funding has led U.S. universities to seek increased research and development funding from private industry).

184. See McConathy & Weber, *supra* note 40, at 177–78 (describing research tools as existing at overlap “where universities seek patents and royalties, where industry finances and supports academic research, and where collaborations between the two occur”).
arose from a profit motive. Arguably, society may benefit more when profit motive drives innovation. This is because industry funding of university research tends to focus on short-term projects leading to marketable products rather than longer-term basic research. Thus, the "anti-commercialism" element of the experimental use doctrine as currently interpreted actually works against the prompt introduction of new drugs and therapies into the market place.

Profit motive should no longer be held antithetical to the experimental use doctrine. A re-conceptualization of the experimental use doctrine must consider the commercial realities of the twenty-first century research and development process. The involvement of a for-profit firm in the use of patented research tools to develop new products should not be treated as per se outside the scope of the experimental use doctrine. Foreign patent systems have adopted legal rules that come much closer to reflecting these notions than has the United States, as demonstrated in the next Part.

V. INTERNATIONAL ACCEPTANCE OF AN EXPERIMENTAL USE EXEMPTION

National patent systems other than that of the United States have generally accepted the concept of an exemption from patent infringement for experimental or research use of a patented invention. This acceptance is consistent with the international patent community’s greater tolerance of...
incursions on patent exclusivity such as working requirements and compulsory licensing.

For example, Germany’s patent laws provide that “‘[t]he effects of the patent shall not extend to acts performed for experimental purposes relating to the subject-matter of the patented invention.’” The Federal Supreme Court of Germany recently interpreted this provision to absolve from liability certain clinical trials of a patented pharmaceutical, although the trials were conducted for the purpose of finding new applications for the pharmaceutical. The court indicated that the exemption would be available even if the unlicensed use resulted in the accused infringer filing a patent application on the results of its research.

An experimental use exemption is included in the European Commission’s proposed Council Regulation on the Community Patent. Article 9 of the proposed regulation excludes from the effects of a Community patent those acts “done privately and for non-commercial purposes,” and those “acts done for experimental purposes relating to the subject-matter of the patented invention.” The term “relating to” is not

189. Von Meibom & Pitz, supra note 188, at 29 (describing German Supreme Court, GRUR 1996, 109-clinical trials (November 7, 1995)). According to these commentators, application of the exemption from liability does not turn on “whether, over and above the character of pure research, commercial interests are also in the background.” Id.; see also Christian Hertz-Eichenrode, Germany, in 1 WORLD INTELLECTUAL PROPERTY RIGHTS AND REMEDIES (1999), at GER-14 (characterizing German Supreme Court decisions as giving “wide interpretation” to Germany’s experimental use provision. German Patent Act of 16 December 1980, § 11.2).
190. Von Meibom & Pitz, supra note 188, at 29.
191. Commission of the European Communities, Proposal for a Council Regulation on the Community Patent, 69 (Aug. 1, 2000), available at http://europa.eu.int/comm/internal_market/en/intprop/indprop/412en.pdf [hereinafter Community Patent]. The proposed regulation provides for the grant of a unitary “Community patent” of equal effect throughout the European Community. Id. § 2.4.1, at 9 (“Explanatory Memorandum”). A newly-created, centralized “Community Intellectual Property Court” with Community-wide jurisdiction will determine enforcement and validity questions. Id. § 2.4.5.1, at 13. The proposed regulation on the Community Patent is independent from the European Patent Convention (EPC), which was signed in 1973. Id. § 1.1, at 4. The EPC provides a single procedure for the examination of patent applications in the European Patent Office. Id. Once a European patent has been granted, however, it becomes a national patent in each member country designated by the applicant, and is subject to the patent laws of each such designated country. Id. The EPC does not provide any Community-wide enforcement forum; rather, any infringement of a European patent “shall be dealt with by national law.” European Patent Convention, Art. 64(3), available at http://www.european-patent-office.org/legal/epc/ea64.html (last modified Nov. 8, 2000).
193. Id. Art. 9(b).
Experimental Use Exception to Patent Infringement

defined, but suggests a scope broad enough to envision the use of patented research tools. The proposed regulation follows a thirty-year effort to create a unitary Community patent.194 The European Council has recommended that the Community patent be implemented by the end of 2001.195

Japanese patent laws provide that “[t]he effects of the patent right shall not extend to the working of the patent right for the purposes of experiment or research.”196 Japan has implemented an experimental use exemption in reverse fashion from the United States; Japan provides a broad exemption for experimental use, but generic-drug manufacturers have had to rely on the judiciary for an interpretation placing them within the exemption. The Supreme Court of Japan recently held that the statutory exemption does apply to the testing of patented drugs for purposes of obtaining the data required for application to manufacture a generic equivalent under Japan’s Drugs, Cosmetics and Medical Instruments Act.197

The NIH Working Group contends that these foreign patent systems properly distinguish between “experimenting on a patented invention—i.e., using a patented invention to study the underlying technology or perhaps to invent around the patent,” and “experimenting with a patented invention to study something else.”198 The Working Group suggests that treating the former as eligible for the research exemption and the latter as ineligible is a “sensible distinction,”199 because:

It is difficult to imagine how a broader research exemption could be formulated without effectively eviscerating the value of patents on research tools. Researchers are ordinary consumers of patented research tools, and if these consumers were exempt from infringement liability, the patent holder would have nowhere else to turn to collect patent royalties. An excessively broad research exemption could eliminate incentives for private firms to develop and

194. Id. § 1.1, at 4 (“Explanatory Memorandum”).
195. Id. § 4, at 69 (“Impact Assessment Form”).
199. Id.
disseminate new research tools, which could on balance do more harm than good to the research enterprise.200

The Working Group’s position that a broadened experimental use rule should not be available to those “working with” a patented invention (e.g., those using the patented invention as a research tool) is reasonable only if such workers are truly “ordinary consumers” of the tool. In other words, these research workers can freely acquire the tools they need in the marketplace at reasonable cost via anonymous purchasing without the need for licensing transactions. The growing incidence of high transaction costs associated with accessing multiple patented research tools201 contravenes the ordinary consumer assumption, however. When research tool transaction costs are severe enough to impede or stop the development of new biomedical products, line-drawing between “experimenting on” and “experimenting with” is no longer justified. In such cases, access to the experimental use doctrine should not turn on the relatively fine distinction between experimenting on or experimenting with the patented invention.202

The Working Group’s concern that a broadened experimental use doctrine would leave holders of research tool patents uncompensated and without sufficient incentives to develop new research tools is a valid one if all non-consensual tool users were given a complete exemption from liability. A more viable alternative, however, is the adoption of a liability rule under which the patent holder cannot enjoin the researcher’s use, but will obtain an ex post royalty based on the marketplace valuation of products developed through use of the tool. The research user’s access problem is alleviated because a license need not be negotiated prior to the use and an appropriate level of royalty to the patent holder will ensure that incentives to innovate are not significantly decreased.203

200. Id.
201. See supra notes 31–36.
202. The general notion of discerning whether a patented invention has been “experimented on” rather than “experimented with” may be an exercise in semantics. Consider, for example, the case of a widget manufacturer who seeks to avoid infringement liability by designing around the widget patent of a competitor. The manufacturer’s goal is to develop an acceptable but non-infringing alternative. Has the manufacturer experimented on the competitor’s widget in designing around it, or experimented with it?
203. Part VII infra discusses this proposal.
VI. RESPONDING TO TRADITIONAL ARGUMENTS AGAINST A BROADENED EXPERIMENTAL USE DOCTRINE

This Article proposes to broaden the “truly narrow” experimental use defense as it is currently interpreted in the United States. Where transaction costs of accessing patented research tools are severe enough to impede or halt the development of new products important to public health, the non-consensual “development use” of such tools should be permitted without injunction, in exchange for an ex post royalty payment based on the marketplace value of the newly created products. This Part responds to a number of the traditional arguments against expanding the experimental use doctrine and provides additional justifications for re-thinking the doctrine to address the research tools access dilemma.

A. Incentive Function of Exclusivity

The most commonly stated objection to broadening the experimental use doctrine is the possibility that it would significantly reduce the incentives for invention of new research tools. Any reduction in the value of patents as drivers for new innovation might result in either a decreased level of innovation or a shift away from disclosure; i.e., protecting inventions as trade secrets rather than by patenting. If a reduction in innovation were to result from a broadened experimental use doctrine, then modifying the experimental use doctrine merely trades one problem for another.

The extent of any potential reduction in innovation, and the optimal balance between a desired level of innovation in research tools versus innovation in new commercial products developed through the unrestricted use of those tools, are probably impossible to determine. Although an expanded experimental use doctrine might feasibly result in some reduction in the development of new research tools, it is just as likely that lessening the probability of the royalty stacking problem will promote increased development of new products that require the use of multiple tools. The

204. Id.
205. Karp, supra note 107, at 2181 (contending that “[a]n expansive experimental use exception, which threatens the patentee’s potential for economic returns, would reduce inventive activity, particularly in those industries that rely heavily on patent protection”).
206. Eisenberg, supra note 37, at 1030–31 (stating that level of reduction in incentives to invent that would result from reducing strength of patents via research exemption is difficult to determine, as is determining whether recognition of research exemption would have different impact from other reductions of patent strength such as reducing patent term). Ultimately the question reduces to an empirical one. Id. at 1074.
diminution of incentive occasioned by divesting a patent owner’s injunctive remedy may be compensated for by creating the opportunity for a sufficiently generous after-the-fact monetary award. The adoption of a reach-through royalty approach that correlates the level of ex post royalty payments to the tool patent owner with the commercial value of the new products developed by use of the patented tool would work to minimize the disincentive effect.\textsuperscript{207}

B. Transformative Versus Commercial Purpose

An expanded experimental use doctrine is in keeping with the need for “safety valves” in all areas of intellectual property law, and especially so in view of the current climate of increasing “propertization” of intellectual property.\textsuperscript{208} Existing safety valves that safeguard against intellectual property rights of excessive scope include the reverse doctrine of equivalents in patent law\textsuperscript{209} and the fair use doctrines of the copyright\textsuperscript{210} and trademark\textsuperscript{211} laws.

The policies underlying the fair use doctrine of copyright law can readily support an expanded experimental use doctrine in patent law. Statutorily

\textsuperscript{207} The proposed adoption of reach-through royalties as the means for compensating the research tool patentee is detailed infra notes 286–305 and accompanying text.

\textsuperscript{208} Lawrence Lessig, The Problem With Patents, INDUS. STANDARD (Apr. 23, 1999), available at http://www.thestandard.com/article/display/0,1151,4296,00.html (decrying “feeding frenzy” in intellectual property law and asserting that Clinton administration was “obsessed” with intellectual property rights).

\textsuperscript{209} The reverse doctrine of equivalents absolves an accused infringer from infringement liability where the accused device, although literally falling within the scope of the asserted patent claim, is so far changed in function, way, or result that a finding of liability cannot be justified as a policy matter. Graver Tank & Mfg. Co. v. Linde Air Prods. Co., 339 U.S. 605, 608–09 (1950); Westinghouse v. Boyden Power Brake Co., 170 U.S. 537, 568 (1898). Although the topic of favorable academic attention, see Merges & Nelson, supra note 37, at 862–68 (advocating reverse doctrine of equivalents as mechanism to limit patent scope in face of significant technological improvement by accused infringer), the reverse doctrine of equivalents has rarely been applied by the courts to excuse liability. Lemley, supra note 70, at 1011.

\textsuperscript{210} 17 U.S.C. § 107 (1994) (providing that fair use of copyrighted work is not infringement of copyright). Unlike copyright law, patent law has not contemplated any fair use provision. The traditional, narrow experimental use doctrine of patent law has been characterized as the closest patent doctrine to copyright’s fair use. Lemley, supra note 70, at 1038 n.236.

\textsuperscript{211} See 15 U.S.C. § 1115(b)(4) (Supp. IV 1998) (codifying “fair use” defense to trademark infringement where defendant uses term “which is descriptive of and used fairly and in good faith only to describe” the defendant’s goods or services or their geographic origin); J. THOMAS MCCARTHY, 2 MCCARTHY ON TRADEMARKS AND UNFAIR COMPETITION § 11.17 (4th ed. 1996) (explaining that under trademark fair use doctrine, junior user of descriptive term always remains free to use such term in its “primary, descriptive sense;” rather than in its source-indicating sense).
enacted in the 1976 Copyright Act, the fair use doctrine provides that certain socially beneficial uses of copyrighted works, such as for research, criticism, and news reporting, are not copyright infringement. Recognition of the copyright fair use doctrine has been seen as necessary to fully promote the constitutional goal of stimulating the production of new copyrightable works. As a means of lessening or alleviating the restrictions on research and development that have been occasioned by the patenting of research tools, an expanded experimental use doctrine would likewise promote the constitutional goal of progress in the useful (technological) arts.

The commercial nature or profit motive of the accused infringer’s use is not fatal to enjoyment of the fair use defense in copyright law. In contrast with the traditional narrow understanding of patent law’s experimental use doctrine, recent judicial interpretations have expanded the copyright fair use doctrine to encompass unlicensed uses that are decidedly commercial in nature. The U.S. Supreme Court’s 1994 decision in *Campbell v. Acuff-Rose Music, Inc.* made clear that commercial uses are not per se unfair and thus infringing.

Rather than focusing on whether an accused infringer’s unauthorized use is commercial in its purpose, fair use in copyright law turns primarily on the degree to which that use is “transformative” in nature. In contrast with a use that would merely supercede or supplant the original work, a transformative use is one that generally furthers the goals of copyright


215. Id. at 572 (reversing court of appeals holding that defense of fair use for rap parody of copyrighted song was barred by parody’s commercial character and excessive borrowing from original). *Compare Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 451 (1984) (stating that “every commercial use of copyrighted material is presumptively . . . unfair”), with *Campbell*, 510 U.S. at 584–85 (rejecting narrow interpretation of *Sony* as creating presumption of unfair use arising from commerciality and recognizing that such presumption would “swallow nearly all of the illustrative uses listed in the preamble paragraph of [17 U.S.C.] § 107, including news reporting, comment, criticism, teaching, scholarship, and research, since these activities ‘are generally conducted for profit in this country’”) (quoting Harper & Row Pubs., Inc. v. Nation Enters., 471 U.S. 539, 592 (1985) (Brennan, J., dissenting)).

When the unauthorized use of another’s copyrighted work results in something that copyright law seeks to promote—i.e., “adds something new, with a further purpose or different character, altering the [copyright owner’s work] with new expression, meaning, or message”—the use is more likely to be held a fair one and therefore outside the realm of copyright infringement.

Judicial elevation of transformative character over commerciality in the copyright regime can serve as a model for a similar treatment in patent law, and particularly so with respect to the patent experimental use doctrine, given its proximity to copyright fair use principles. Another patent law doctrine, the doctrine of equivalents, has already been the subject of a transformative-use type of analysis, although not given that label. Compelling arguments have been made that patent infringement determinations under the doctrine of equivalents should take into consideration the extent to which an accused infringement represents a technological improvement over the patented invention. In other words, the technological contribution made by the accused infringer should be part of the calculus of patent infringement liability.

Analysis of the accused infringer’s contribution is particularly applicable when it is not operating within the literal boundaries of the patentee’s claims, but rather is liable, if at all, under the more imprecise rubric of the doctrine of equivalents. In such cases, the greater the extent of the accused infringer’s technological advance over the patented invention, the

---

217. *Id.* at 579.

218. *Id.* (“[T]he goal of copyright, to promote science and the arts, is generally furthered by the creation of transformative works.”).

219. *Id.*

220. The common law experimental use doctrine of patent law has been characterized as the closest patent doctrine to copyright’s fair use. *See* Lemley, *supra* note 70, at 1038 n.236.

221. Merges & Nelson, *supra* note 37, at 857–59 (citing with approval court’s analysis in *Texas Instruments, Inc. v. United States International Trade Commission*, 805 F.2d 1558 (Fed. Cir. 1986), that focused on “the merits of the accused device”); *id.* at 909–11 (advocating that after courts have determined significance of patented invention they should also consider “importance of the advance represented in the accused device”).

222. The infringer’s contribution could be relevant in determinations of equivalency under the doctrine of equivalents, as well as in assessing whether the reverse doctrine of equivalents should be applied to excuse a literal infringement that is so far changed in principle from the patented invention that it is unfair to find liability. *See* Merges & Nelson, *supra* note 37, at 862–68 (advocating reverse doctrine of equivalents as mechanism to limit patent scope in face of significant technological improvement by accused infringer); *id.* at 867 & n.120 (noting that same rationale applies to doctrine of equivalents analysis and is perhaps even more useful there).

223. *Id.* at 867 n.120.
Experimental Use Exception to Patent Infringement

less likely that the differences between the claimed and accused devices are merely “insubstantial”224 and therefore invoking application of the doctrine of equivalents.225 Indeed, the Federal Circuit has recognized that where an accused infringer’s product is itself patented, this fact may evidence that the differences between the claimed and accused devices are beyond those permissible for application of the doctrine of equivalents.226

Professor Mark Lemley has proposed a patent law model in which the class of “truly radical improvements” would be exempted from liability for infringement227 even if the improvement falls within the literal scope of the patent claim at issue.228 Such an exemption would create an incentive for the creation of improvements and avoid the licensing transaction costs that might otherwise prevent those improvements from reaching the marketplace.229 The same analysis can be extended to uses of research tools that produce a commercial product not physically incorporating the patented tool itself. Such products can be viewed as representing the ultimate category of improvement within Professor Lemley’s hierarchy, as well as the epitome of transformative use in the copyright fair use sense.

C. Research Tool Patentability and Claim Scope

Broadened acceptance of the non-consensual use of patented research tools might also be criticized as a makeshift solution to the more fundamental problem of the issuance of biotechnological patent claims of

224. Sage Prods., Inc. v. Devon Indus., Inc., 126 F.3d 1420, 1423 (Fed. Cir. 1997) (stating that doctrine of equivalents applies at claim-limitation level if only “insubstantial differences” distinguish claimed limitation from corresponding component of accused device); Hilton Davis Chem. Co. v. Warner-Jenkinson Co., 62 F.3d 1512, 1516–18 (Fed. Cir. 1995) (expressing test for application of doctrine of equivalents in terms of whether differences between claimed and accused devices are “insubstantial”).

225. Merges & Nelson, supra note 37, at 867.

226. Zygo Corp. v. Wyko Corp., 79 F.3d 1563, 1570 (Fed. Cir. 1996) (explaining in context of doctrine of equivalents infringement analysis that “[t]he nonobviousness of the accused device, evidenced by the grant of a United States patent, is relevant to the issue of whether the change therein is substantial”); Nat’l Presto Indus., Inc. v. W. Bend Co., 76 F.3d 1185, 1192 (Fed. Cir. 1996) (characterizing fact of patentability of accused device as “relevant” and “entitled to due weight” in infringement analysis, though not determinative); Atlas Powder Co. v. E.I. du Pont De Nemours & Co., 750 F.2d 1569, 1580 n.3 (Fed. Cir. 1984) (suggesting that patentability of accused device based on “unexpected results” over patentee’s device might evidence non-equivalency based on substantially different “result”).

227. Lemley, supra note 70, at 1010–13, 1070.

228. When the “radical improvement” does fall within the literal boundaries of the claim, this invokes the reverse doctrine of equivalents. Id. at 1010–11.

229. Id. at 1070.
unjustified scope. But administrative and judicial actors are beginning to respond to the scope problem. The U.S. Patent and Trademark Office (USPTO) recently issued guidelines requiring a more rigorous showing of utility as a prerequisite to patentability, and the Federal Circuit seems more than willing to wield the enablement and written description requirements of the first paragraph of 35 U.S.C. § 112 as claim scope-shrinking implements.

Others contend outright that research tools should be completely excluded from patenting. They argue that, rather than exempting

230. Cf. Barton, supra note 15, at 449 (criticizing broad patents as potentially “usable to prevent entry by others . . .” or to “affect the incentives for further research by others”); Merges & Nelson, supra note 37, at 916 (concluding that “[w]hen a broad patent is granted or expanded via the doctrine of equivalents, its scope diminishes incentives for others to stay in the invention game, compared again with a patent whose claims are trimmed more closely to the inventor’s actual results”); Rai, supra note 37, at 120–29 (criticizing “development-oriented perspective” of Edmund Kitch and others as advocating broad patents on basic inventions as “prospects” to encourage further development of that technology, coordinated by single rights holder).

231. See Examination Guidelines, supra note 61, at 71,441; see also Q. Todd Dickinson, Reconciling Research and the Patent System, ISSUES SCI. & TECH., Summer 2000, at 69 (asserting that USPTO is cognizant of concerns that patents on gene fragments “might retard basic research and that these claims will form an intricate licensing web that will impede their use in developing cures for diseases,” and that USPTO “continue[s] to take steps to ensure that patent applications in these areas are meticulously scrutinized for an adequate written description, sufficiency of the disclosure, and enabled utilities”).

232. 35 U.S.C. § 112 (1994). Other recent examples of the Federal Circuit’s application of the first paragraph of § 112 in a restrictive fashion include Enzo Biochem, Inc. v. Calgene, Inc., 188 F.3d 1362, 1377 (Fed. Cir. 1999) (affirming judgment of invalidity based on failure of patent directed to “highly unpredictable” antisense technology to satisfy enablement requirement of § 112’s first paragraph, where claims at issue were “quite broad” and extent of experimentation required to practice invention as claimed in all prokaryotic or eukaryotic cells would be undue in light of examples disclosing successful performance of invention with only single prokaryotic cell, e. coli), and Genentech, Inc. v. Novo Nordisk, A/S, 108 F.3d 1361, 1366 (Fed. Cir. 1997) (reversing preliminary injunction and holding invalid for non-enablement patent directed to method of producing human growth hormone through cleavable fusion expression that, in Federal Circuit’s view, provided only “a direction for further research”). E.g., Janice M. Mueller, The Evolving Application of the Written Description Requirement to Biotechnological Inventions, 13 BERKELEY TECH. L.J. 615 (1998) (characterizing Federal Circuit’s decision in Regents of the University of California v. Eli Lilly & Co., 119 F.3d 1559 (Fed. Cir. 1997), as having effectively elevated written description requirement of § 112’s first paragraph to “super enablement” standard, and having improperly applied written description requirement to originally filed claims).

233. Public comments received by the USPTO in response to its first round of Interim Written Description Guidelines, published in June 1998, asserted that “ESTs [Expressed Sequence Tags] are genomic research tools that should be available for unencumbered research to advance the public good.” Examination Guidelines, supra note 61, at 71,441; see also McConathy & Weber, supra note 40, at 177 (expressing alarm of many in biotechnological community that exclusive rights in research tools will preclude opportunities for development of products in long term, and arguing that research tool inventions should not be patentable); cf. Rebecca S. Eisenberg & Robert P. Merges, Opinion
unauthorized research uses of patented research tools from infringement liability, research tools would be better dealt with by excluding them from patentability in the first instance.

But prohibiting the patenting of any particular technology or type of invention is a draconian step that has generally been avoided in the United States. To take away any possibility of any remedy, injunctive or monetary, for unauthorized uses of all research tools could deleteriously constrict incentives for the creation of new research tools. A complete exclusion would also run afoul of the developmental history of U.S. patent jurisprudence, which traditionally has taken an expansive stance toward patentable subject matter. Exclusion of particular technologies from patentability also contravenes U.S. treaty obligations under the non-discrimination provisions of the General Agreement on Tariffs and Trade (GATT) Uruguay Round Agreements, Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS).
Reliance interests of patent owners and innovators who rely on the possibility of patent protection would also be threatened by excluding research tools from patentability. Numerous patents have already been granted on research tools. In view of the current industry practice of patenting to the greatest extent possible, any proposal for the outright prohibition of patents on research tools appears destined for failure.

Lastly, the issuance of patents on research tools results in social benefits as well as costs. Publication of enabling descriptions of how to make and use new research tools increases the store of public knowledge, despite the “royalty stacking” problem that may arise when development of commercial products demands multiple authorizations to practice several different research tool patents. As the USPTO’s Director has contended, “the need for a possible research tool exemption . . . should not drive a narrowing of subject matter in order to create . . . a de facto patenting exception.” The patenting of research tools should not be barred.

D. Constitutional Implications

A broadened experimental use exemption arguably might run counter to a textualist interpretation of the Intellectual Property Clause of the U.S. Constitution. The core of U.S. patent law, absent in many foreign regimes, is Congress’s constitutional power to promote the progress of the useful arts in a manner that the Framers in 1787 spelled out with an unusual degree of particularity by securing to inventors, for limited

238. See supra notes 51–61 and accompanying text for examples. An on-line search of the LEXIS U.S. Utility Patents database (which includes patents issued from 1971 through the present) conducted on June 6, 2000 for patents having the phrase “research tool” in the “Summary of the Invention” section retrieved 666 hits.

239. Dickinson, supra note 230, at 70.

240. See U.S. CONST., art. I, § 8, cl. 8 (“Congress shall have Power . . . To promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries.”). The quoted constitutional text serves a dual purpose by establishing the basis of both copyright and patent law in the United States. The Framers’ use of the word “Science” is believed to refer to copyrightable subject matter, while “useful Arts” is thought to mean “technological arts” or patentable subject matter. In re Bergy, 596 F.2d 952, 958–59 (C.C.P.A. 1979) (Rich, J.).

241. In many foreign countries, the existence of a patent system is viewed as a limited exception to a general prohibition against monopolies. E.g., PENROSE, supra note 117, at 7 (describing Statute of Monopolies of 1623 as basis of present British patent law); Edwin S. Flores Troy, The Development of Modern Frameworks for Patent Protection: Mexico, a Model for Reform, 6 TEX. INT’L PROP. L.J. 133, 137 (1998) (describing Article 28 of the Mexican Constitution of 1917 as prohibiting all monopolies except those that “serve the nation,” including privileges granted to inventors and authors).

242. No other enumerated power is spelled out in such detail. See U.S. CONST., art. I.
times, the “exclusive right” to their discoveries. As expressed in the Patent Act, the “exclusive right” is now understood as the patentee’s “right to exclude” all others from making, using, or selling (and more recently, offering to sell and importing) the patented invention.

The Framers’ use of the modifier “exclusive” to describe the right that Congress was authorized to secure to inventors probably does not create a Constitutional barrier to modification of that right, however. To conclude otherwise would require ignoring the instances in which Congress or the courts have already recognized exceptions to patent exclusivity. Moreover, a strict textualist interpretation of the Intellectual Property Clause is of minimal use given the Clause’s expansive terms and lack of “built-in limits.” The Intellectual Property Clause is probably best understood as giving Congress the power to grant exclusive rights, but not requiring that it do so. Thus, Congress has the power to grant less-than-exclusive patent rights. A broadened experimental use defense would not run afoul of the Constitution.

E. Conventional U.S. Norms of Patent Exclusivity

Re-conceptualizing the experimental use exemption will require a decided change in the culture of the American patent system. The tradition of sharing scientific information is the hallmark of the communalistic

244. 35 U.S.C. § 154 (a)(1) (1994) (“Every patent shall contain . . . a grant to the patentee . . . of the right to exclude others from making, using, offering for sale, or selling the invention throughout the United States or importing the invention into the United States . . . .”) (emphasis added); Bloomer v. McQuewan, 55 U.S. (14 How.) 539, 548–49 (1852) (“The franchise which the patent grants, consists altogether in the right to exclude every one from making, using, or vending the thing patented, without the permission of the patentee.”).
245. See infra notes 253, 256–67 and accompanying text.
247. Cf. NIMMER ON COPYRIGHT § 1.07 (1999). Professor Nimmer asserts that with respect to U.S. copyright law:

Inasmuch as Congress manifestly has the power either to grant complete exclusivity or no protection at all, it would seem that it may properly invoke protection somewhere between these two polar positions. Nonexclusivity under a compulsory license appears to constitute such a reasonable middle ground. It may then be concluded that the phrase ‘the exclusive right’ imports words of authority, but not of limitation.

Id.
nature of science. 248 This tradition would be greatly facilitated with less restricted access to patented research tools. On the other hand, expanding the doctrine clearly contravenes historical norms of expansive patent rights and traditionally fierce antipathy by the U.S. patent-holding community toward incursions on exclusivity.

The experimental use exemption is a decidedly unappealing notion to advocates of the view that the exclusionary power of a U.S. patent bestows its owner with broad-ranging control over future technology. This view is grounded in the well-established rule that the owner of a U.S. patent on a “basic” or “pioneer” invention may enjoin a follow-on developer from using the improvement invention, even if the improvement invention is independently patentable over the basic invention. 249 The basic patent “dominates” the improvement patent, so long as the claims of the basic patent are interpreted broadly enough to read on the improvement. The existence of “dominant” and “subservient” patents flows from the fact that a U.S. patent grants its owner a negative exclusionary right to prohibit others from using the patented invention, rather than bestowing on the patentee an affirmative right to use the invention. 250

The doctrine of equivalents also fosters the U.S. norm of patents as broad property rights relatively immune from incursions. New technology developed after the issuance of a particular patent, in an effort to improve on or even design around the patented invention, may still infringe that patent under the doctrine of equivalents, even if it does not fall within the literal boundaries of the patentee’s exclusionary right as defined by the patent’s claims. 251 The doctrine of equivalents in essence creates a penumbra of exclusionary power around the explicit boundaries specified in the patent instrument, and its extent is determinable only through litigation to enforce the patent. Notably, the fact that the infringed patent

248. Cf. Rai, supra note 37, at 89–90 (defining “communalism norm” of U.S. scientific community as one that views scientific information as “shared resource” and promotes “public domain of freely available scientific information”).


250. See 35 U.S.C. § 154(a)(1) (1994) (defining patent grant as “the right to exclude others”). Thus, the holder of the improvement patent does not obtain from the patent any affirmative right to practice the improvement invention, even though it is patented.

Experimental Use Exception to Patent Infringement

did not (and could not) disclose nor factually enable others to make and use the after-arising accused equivalent is not fatal to the patentee’s claim.\textsuperscript{252}

When viewed from this doctrinal and historical perspective, the U.S. patent community’s reticence to embrace a meaningful experimental use doctrine is entirely understandable. At least until recently, the far-reaching exclusionary power of the patent property right was sacrosanct. Absent very limited exceptions,\textsuperscript{253} the patentee’s right to exclude was not burdened by governmental intrusion. Courts reacted negatively to any limitations on enforcement rights.\textsuperscript{254} At the behest of patent owners and patent bar groups, U.S. legislators have traditionally rejected derogations of the patentee’s

\textsuperscript{252} The basic patent need only comply with the enablement requirement in the first paragraph of 35 U.S.C. § 112 insofar as the claimed invention was understood as of its patent application filing date; it need not enable the later-developed infringing device. United States Steel Corp. v. Phillips Petroleum Corp., 865 F.2d 1247, 1250–52 (Fed. Cir. 1989); In re Hogan, 559 F.2d 595, 605–07 (C.C.P.A. 1977).

\textsuperscript{253} Historically, only a small number of exceptions were recognized to the general rule against derogations of a patent’s exclusivity. For example, the federal government retains the power to use any patented invention but must pay just compensation for the taking if infringement is proved. 28 U.S.C. § 1498 (1994).

\textsuperscript{254} E.g., Smith Int’l, Inc. v. Hughes Tool Co., 718 F.2d 1573, 1578 (Fed. Cir. 1983) (“Without the right to obtain an injunction, the right to exclude granted to the patentee would have only a fraction of the value it was intended to have, and would no longer be as great an incentive to engage in the toils of scientific and technological research.”).
right to exclude that are widely recognized in foreign systems, such as compulsory licensing, working requirements, and prior-user rights. 255

More recently, however, fundamental changes in U.S. patent law have begun to foster an increasingly hospitable environment for acceptance of a broader experimental use doctrine. Congressional, executive, and judicial actors are implementing a number of significant incursions into patent exclusivity. As described previously, Congress passed the Hatch-Waxman Act in 1984, adding to the Patent Act § 271(e)’s safe harbor for testing of patented drugs and medical devices for purposes reasonably related to regulatory data gathering. 256 In 1996, Congress enacted the remedies limitation found in § 287(c) of the Patent Act, which precludes the owner of a patent directed to a “medical procedure” from enjoining or obtaining damages from an infringer of that patent. 257 By passage of the American Inventors Protection Act of 1999, 258 the United States has for the first time a limited form of prior-user rights, 259 which have long been recognized in Europe and Japan. 260 Over the protests of American pharmaceutical manufacturers, President Clinton in May 2000 issued an Executive Order

255. E.g., Dawson Chem. Co. v. Rohm & Haas Co., 448 U.S. 176, 215 (1980) (describing compulsory licensing as “rarity” in U.S. patent system); see also PENROSE, supra note 117, at 172 (explaining that compulsory licensing has been “violently opposed” in United States because it “can be such a serious derogation of the monopoly ‘rights’ of the patentee”); F.M. SCHERER, INDUSTRIAL MARKET STRUCTURE AND ECONOMIC PERFORMANCE 456 (1970) (“[E]very attempt to alter the U.S. law in this direction [of introducing general compulsory licensing provisions] has been beaten down as a result of determined opposition from industrial groups and the patent bar.”); Feit, supra note 26, at 840 n.102 (characterizing U.S. policy as “not favoring compulsory licenses.”); McConathy & Weber, supra note 40, at 178 (describing compulsory licensing as “abhorrent to both academia and industry”); Merges & Nelson, supra note 37, at 911 (describing compulsory licensing as “anathema” to U.S. patent law).

Congress considered but ultimately dropped the idea of compulsory licensing as part of the 1952 Patent Act. Dawson, 448 U.S. at 215 n.21.

256. See supra notes 125–27 and accompanying text.


259. In general terms, a prior-user right permits one who, prior to the filing of the patent in suit, had independently invented the same subject matter, to continue making, using, or selling it at pre-suit levels. A prior-user right is in essence a license to continue practicing the invention subsequently patented by another.

Experimental Use Exception to Patent Infringement

providing that the U.S. government will not challenge those sub-Saharan African governments that grant compulsory licenses under pharmaceutical patents on HIV/AIDS drugs or permit parallel imports of these drugs from other countries where they are available at lower prices. And in November 2000, the Federal Circuit sitting en banc announced in Festo Corp. v. Shoketsu Kinzoku Kogyo Kabushiki Co. stringent new rules of prosecution history estoppel that significantly contract the ability of patent owners to rely on the doctrine of equivalents.

Perhaps the most significant recent incursion into patent exclusivity occurred in 1999 when the U.S. Supreme Court held in Florida Prepaid Postsecondary Education Expense Board v. College Savings Bank that Congress’s 1992 abrogation of Eleventh Amendment immunity from patent infringement liability for states, instrumentalities of states, and state employees acting in their official capacity was unconstitutional. As a result of College Savings Bank, state universities are immune from patent infringement liability under the federal patent laws.

The potential ramifications of the College Savings Bank decision are quite troubling in at least two respects. First, the state courts may begin to


262. 234 F.3d 558 (Fed. Cir. 2000) (en banc).

263. Id. at 569 (holding that when amendment of claim creates prosecution history, no range of equivalents is available for amended claim limitation).


265. As a general rule, state governments are immune from lawsuits under the Eleventh Amendment of the U.S. Constitution, which provides that the “Judicial Power of the United States shall not be construed to extend to any suit in law or equity, commenced or prosecuted against one of the United States by Citizens of another State, or by Citizens or subjects of any foreign state.” U.S. CONST. amend. XI. In 1992, Congress stripped the states of their immunity for patent infringement by enacting the Patent Remedy Act. Patent and Plant Variety Protection Remedy Clarification Act, Pub.L. No. 102-560, 106 Stat. 4230 (1992) (codified at 35 U.S.C. §§ 271(h), 296(a) (1994)).

In Seminole Tribe v. Florida, 517 U.S. 44 (1996), the U.S. Supreme Court held that Congress cannot abrogate the states’ Eleventh Amendment sovereign immunity under Congress’s Article I powers. Seminole Tribe thus foreclosed reliance on congressional authority under either the Commerce Clause or the Intellectual Property Clause to sustain the Patent Remedy Act. In College Savings Bank, the argument for constitutionality of the Patent Remedy Act was accordingly premised on the Due Process Clause of the Fourteenth Amendment. 527 U.S. at 636. The petitioner contended that patents are private property and to allow states to use them without authorization and without remedy was a deprivation of private property without due process. Id. The U.S. Supreme Court rejected the Due Process theory on the grounds that there had not been a sufficient showing that state governments routinely infringe, that such infringement is willful rather than merely negligent or “innocent,” or that such infringement results in a property deprivation without due process. Id. at 639–46. In support of its ruling, the Court pointed to the possibility of alternative remedies in state court such as unfair competition causes of action. Id. at 643–44 & nn.8–9.
hear patent infringement cases brought under the guise of state unfair competition actions. State court activity in this arena is contrary to federal preemption principles of patent law and the federal nature of patent rights as recognized in U.S. law since 1790. Placing quasi-patent jurisdiction in the state courts is also antithetical to the “national uniformity” rationale underlying the Federal Circuit’s formation.266

Second, College Savings Bank may have an unforeseen impact on the formation, arrangement, and management of research and development collaborations between state universities and private industry. Private-sector partners will presumably seek to maximize their opportunities to be shielded from liability under the state universities’ umbrella of immunity. However, the minimum level of state funding or control or both that would render a state university-industry collaboration eligible for the immunity as an instrumentality of the state is yet unknown. Moreover, state universities may balk at the potential negative impact on their academic independence with good reason. Research agendas ought not to be controlled by funding manipulations for the purpose of helping industry collaborators gain the protections of Eleventh Amendment immunity.

The College Savings Bank holding is especially significant to the debate over access to patented research tools insofar as state universities and their industry collaborators may be frequent users of these tools. Because College Savings Bank gives states immunity from patent infringement liability under Title 35, state university research would appear to have no immediate need for a broadened experimental use doctrine as proposed here.267 Yet as frequent users of patented research tools, the state university is a critical participant in the debate over a broadened exemption.

VII. A PROPOSED “DEVELOPMENT USE” MODEL

This Part reviews a previous proposal for expanding the experimental use doctrine and concludes that further modifications are needed for the case of patented research tools. It proposes a “liability rule” model that would permit the non-consensual “development use” of patented research tools that are not readily available for licensing or purchase, while providing an ex post royalty payment to the owner of the patented research


267. Exceptions would include state and industry collaborations that are not sufficiently “instrumentalities of the state” to qualify for Eleventh Amendment immunity and state officials acting in the scope of their official capacity who may still be subject to prospective injunctive relief under Ex Parte Young, 209 U.S. 123, 167–68 (1908).
Experimental Use Exception to Patent Infringement

tool of sufficient amount to maintain adequate incentives for innovation in new tools. The proposed “reach-through” royalty approach, as it is referred to in consensual biotech licensing transactions, would compute the royalty payment to the tool patent owner based on the marketplace value of the new products or diagnostics developed through use of the patented research tool.

In her seminal 1989 article, Professor Rebecca Eisenberg proposed a three-pronged model for an expanded experimental use doctrine. First, Professor Eisenberg contended that use of a patented invention to verify that the patent’s written description and drawings adequately enabled the claimed invention should be altogether exempt from infringement liability. This aspect of Professor Eisenberg’s proposal is probably closest to Justice Story’s view that the construction of a patented machine “for the purpose of ascertaining the sufficiency of the machine to produce its described effects” is not infringement. Such use is also most likely to fall within the current post-Roche “truly narrow” view of the experimental use defense as limited to “non-commercial” uses.

Second, Professor Eisenberg argued that use of an invention having a “primary or significant market” among research users who are “ordinary consumers” of the invention should not be exempt from liability; in Professor Eisenberg’s view, these “consumers” must obtain a license prior to their use. For example, in a scenario where users of patented transgenic mice as tools for conducting cancer research represent the target market of the mouse patent holder, it is reasonable to assume that the patentee will want to make the mouse widely available at reasonable licensing terms. These “ordinary consumers,” as Professor Eisenberg’s use of that term suggests, do not face access problems if they can readily license or buy the research tool on the open market.

Third, Professor Eisenberg proposed that those who use a patented invention in a manner that leads to improvements in the technological field of that patent, or for the purpose of “designing around” the patent’s claims to avoid infringement, should not have to negotiate for a license prior to

---


269. Id. at 1078 (“Research use of a patented invention to check the adequacy of the specification and the validity of the patent holder’s claims about the invention should be exempt from infringement liability.”).


271. See supra note 121 and accompanying text.

272. Eisenberg, supra note 268, at 1078.

273. Id. at 1085 (suggesting that patentee “will want to sell to these users—they fall squarely within the market for the patented invention”).
Such users would still be considered infringers in Professor Eisenberg’s model but would enjoy a remedies limitation. The patent owner would not be able to enjoin their uses, but “in some cases” would receive an after-the-fact “reasonable royalty” in recognition of the patent owner’s initial investment in developing the patented invention. Professor Eisenberg did not identify criteria for choosing among those patent owners for whom the royalty remedy should be available, nor a method for quantifying its amount.

The third “improver” prong of Professor Eisenberg’s proposed framework can be viewed as a variation on compulsory licensing, without the necessity of first petitioning the patentee or the government for a license. The limitation of the patentee’s remedy to a royalty rather than

274. Id. (citing example of researchers attempting to design improved transgenic mouse or non-infringing transgenic mouse).
275. Id. at 1078.
276. See Rai, supra note 37, at 139 (noting that system under which users pay reasonable royalties to patent owner after unlicensed use is “not very different from a compulsory license scheme”). Professor Scherer has defined “compulsory licensing,” which he also refers to as “mandatory licensing,” as the act of “waiving a patent holder’s normally exclusive right to his invention under specified conditions such as nonutilization of the invention, monopolistic abuses, or other circumstances engendering a public interest in wider availability.” F.M. Scherer, The Economic Effects of Compulsory Patent Licensing 5 in NEW YORK UNIVERSITY’S MONOGRAPH SERIES IN FINANCE AND ECONOMICS (1977).
277. Edith Tilton Penrose has described the most extreme form of such a system as “unconditional compulsory licensing,” under which a license would be available as of right to any requestor without the need for petitioning a governmental agency or establishing the patentee’s failure to supply the domestic market with the patented invention or license on reasonable terms. See PENROSE, supra note 117, at 184 (contending that if unconditional compulsory licensing were adopted “the worst of the social costs of the patent system would be abolished at one stroke,” but also noting traditional fears that such system would abolish social gains from patenting because royalty remedy alone might be insufficient to motivate desired level of innovation).

Yet another variation on unconditional compulsory licensing, available in the United Kingdom and Germany, is a scheme under which a patentee may voluntarily request that the government make licenses available to all comers upon payment of a reasonable royalty. In exchange for the patentee’s agreement to license without restriction, it obtains reduced maintenance fees over the life of the patent. See id. at 177–78 & 177 n.29; German Patent Act of 16 December 1980, § 23(1) (reducing maintenance fees by one half prescribed amounts after patentee files written declaration of willingness to allow anyone to practice patented invention in return for reasonable compensation), reprinted in 2D SINNOTT ET AL., supra note 187, at WEST GERMANY-78.24.

A system of compulsory licensing without satisfaction of any threshold conditions would likely not comply with current international patent agreements. See Agreement on Trade-Related Aspects of Intellectual Property Rights, Including Trade in Goods, Dec. 15, 1993, art. 31, 33 I.L.M. 81, 95 (GATT Uruguay Round Agreements) (permitting use without authorization of rights holder only if “prior to such use, the proposed user has made efforts to obtain authorization from the rights holder on reasonable commercial terms and conditions and that such efforts have not been successful within a reasonable period of time”); Paris Convention for the Protection of Industrial Property, Mar. 20, 1883, art. 5.A.4, 828 U.N.T.S. 107, 123 (Stockholm 1967 rev.) (permitting compulsory licenses to be applied for in cases of patentee’s “failure to work” or “insufficient working” of patent, after later of four years
Experimental Use Exception to Patent Infringement

an injunction also illustrates the choice of a “liability rule” over a “property rule.”  

The royalty stacking problem in biotechnology, occasioned by increasing need for patented tools that are not freely available for purchase by ordinary consumers in the marketplace, has escalated in severity since the 1989 publication of Professor Eisenberg’s article. In the current environment, the assumption inherent in Professor Eisenberg’s second prong that research tools are readily available to “ordinary users” with minimal transactions costs is increasingly less certain. Moreover, innovation in one technology increasingly involves the use of patented research tools from other technologies. For example, a genetically modified mouse may be used to develop and screen new pharmaceuticals for the treatment of cancer in humans, or a DNA chip (formed by layering chains of nucleotides onto silicon) may be used to determine the link between specific single nucleotide polymorphisms (genetic variations) and particular diseases, leading to new screening techniques for these diseases. To the extent that they are not improving the technology of the research tool patent itself (i.e., resulting in improved research tools of the same type), these trans-technologic uses of research tools would appear to fall outside the third, “improver” prong of Professor Eisenberg’s model.

Yet such uses of research tools result in valuable new products that are just from filing or three years from grant, but mandating refusal of request for compulsory license if “patentee justifies his inaction by legitimate reasons”).

278. Ian Ayres & Eric Talley, Solomonic Bargaining: Dividing a Legal Entitlement To Facilitate Coasean Trade, 104 YALE L.J. 1027, 1036–37 (1995) (contrasting “property rules” that protect legal entitlements against all non-consensual takings with “liability rules” that permit non-consensual takings but compensate entitlement holders). A framework for the setting and protection of entitlements under “property rules” and “liability rules” was first proposed in Guido Calabresi & A. Douglas Melamed, Property Rules, Liability Rules and Inalienability: One View of the Cathedral, 85 HARV. L. REV. 1089 (1972). In the patent context the right to an injunction against infringement represents a property rule, while compulsory licensing of patented inventions represents a liability rule. Ayres & Talley, supra, at 1036–37. In the context of a patented improvement invention that is blocked by a dominant patent, Ayres and Talley contend that liability rules are preferable to property rules. Id. Liability rules would facilitate revelation of information which, when it is otherwise not disclosed in cross-licensing negotiations, would lead to significant delay if not complete breakdown of those negotiations. Id. at 1092–93. Ayres and Talley’s implementation of a liability rule in the patent setting accordingly contemplates a compulsory licensing scheme that would “giv[e] the improver an option to infringe the pioneer’s patent in exchange for a fee determined by a licensing tribunal.” Id. at 1093.

279. See supra notes 28–35 and accompanying text; see also Marshall, supra note 57, at 25 (describing debate over restrictions on research use of cre-loxP mouse as “just the latest skirmish in a decade-long battle over commercial controls on basic tools in biomedical research”).


281. See Eisenberg, supra note 268, at 1078 (limiting proposed reasonable royalty limitation to “use of a patented invention in subsequent research in the field of the invention”).
as important, if not more so, from a societal benefit standpoint.  

This Article proposes to extend and adapt Professor Eisenberg’s model to the current research tool milieu so as to permit non-consensual use of research tools not readily available for licensing on reasonable terms or via anonymous marketplace purchase. This “development use” rule would not discriminate against research tool users who seek to develop commercial products. It would be limited to “uses” of research tools as tools, however, and would not encompass the sale of products by the tool user that physically incorporate the patented invention. This latter scenario involves liability for “selling” the patented invention under 35 U.S.C. § 271(a), which should not be exempted from injunctive or damages remedies.

This Article further proposes the adoption of a reach-through royalty structure that would link the royalty payment with the ultimate commercial value of the products developed from use of the patented research tool. The new products would serve as the royalty base. In this manner the royalty payment to the research tool patentee would approximate the true value of the research tool to the tool user and product developer. Such an approach would also work to minimize the disincentive effect on innovation in research tools. The research tool patent owner would acquire the right to a potentially commercially significant future royalty stream, while the tool user would avoid the burdens of pre-use license negotiations, up-front payments, and blocked access to the proprietary research tools.

To ensure adequate notice to the research tool patentee in the proposed model, the putative user would be required to notify the patentee in writing of the user’s intent to use the patented research tool, prior to the use. In contrast with typical consensual licensing negotiations for use of research

282. See Barton, supra note 15, at 457.

283. The limitation of the proposed development use rule to situations where the research tool is not readily accessible through licensing or purchase in the marketplace is in keeping with the “failure of private bargaining” restriction on compulsory licensing under GATT TRIPS. Agreement on Trade-Related Aspects of Intellectual Property Rights, Including Trade in Goods, Dec. 15, 1993, art. 31, 33 I.L.M. 81, 95 (providing that compulsory licensing shall be available only after “the proposed user has made efforts to obtain authorization from the right holder on reasonable commercial terms and conditions and that such efforts have not been successful within a reasonable period of time”).

284. The phrase “development use” is employed in this Article to explicitly encompass those uses of research tools that lead to commercialization of products, and to make clear that the experimental use doctrine should no longer be limited to the purely “philosophical” (in the Roche sense) research component of “research and development.”

285. “Except as otherwise provided in this title, whoever without authority makes, uses, offers to sell, or sells any patented invention, within the United States or imports into the United States any patented invention during the term of the patent therefore, infringes the patent.” 35 U.S.C. § 271(a) (Supp. IV 1998) (emphasis added).

286. Reach-through royalties are discussed in supra note 76 and accompanying text.
tools, however, no disclosure of the nature or details of the use would be required. This minimal declaration of an “intent to use” will place the patentee sufficiently on notice so that it can police any subsequent introduction of new products into the marketplace by the tool user. The patentee’s detection ability can be supplemented by requiring that the tool user additionally provide written notice of the new products at or shortly before sales are commenced. Tool users who choose to opt out and not give proper notices could be made subject to treble damages if infringement is ultimately established.

Whether the utilization of a reach-through royalty approach triggers patent misuse or antitrust concerns has been the subject of some debate in the consensual (i.e., non-compulsory) licensing of biotechnological research tools. Advocates of reach-through royalties contend that patent misuse does not occur if the reach-through royalty results from a bargained, arms-length licensing transaction. The reach-through approach is seen as an expedient method of measuring the value of the use of the research tool rather than an unlawful leverage of the patent right. This treatment of reach-through royalties is consistent with the Federal Circuit’s approval of a license agreement in which the royalty base encompassed not just the

---

287. See McConathy & Weber, supra note 40, at 178 (characterizing conventional reach-through licenses for research tools as requiring “full disclosure” of licensee’s intended use).

288. Germany’s patent code provides a somewhat analogous notice framework, but requires disclosure of the nature of the licensee’s use. Owners of German patents have the option of filing a written declaration with the German Patent Office stating their willingness to allow anyone to practice the patented invention in return for reasonable compensation; the patent owner’s maintenance fees are thereafter reduced by half. German Patent Act of 16 December 1980, § 23(1), reprinted in 2D SINNOTT ET AL., supra note 187, at WEST GERMANY-78.24. Following the recording of the patentee’s declaration, any individual wishing to exploit the invention may do so by notifying the patentee of his or her intent. Id. at § 23(3). The notification must include a statement “of how the invention is to be exploited.” Id. The notifying party is then entitled to practice the invention in the manner stated. Id. The notifying party must also provide the patentee with quarterly reports of the “particulars of the use” and pay compensation for the use, which is assessed by the Patent Division. Id.

289. NIH Principles and Guidelines, supra note 26, at 72,091 (criticizing reach-through royalties); David S. Block & Daniel J. Curran, Patenting Genomic Technologies, 282 SCIENCE 1419 (1998) (denying, in letter responding to Heller & Eisenberg, supra note 28, at 698, that DuPont’s reach-through royalty agreement on patented cre-loxP mice allows DuPont to “leverage its proprietary position in upstream research tools into a broad veto right over downstream research and development products”).

290. Barton, supra note 15, at 461 (asserting that “if [reach-through royalties] are reasonable, they should be permitted and . . . insistence on such terms should not be read as an antitrust violation”); Goldstein, supra note 76; Robert Blackburn, Chief Patent Counsel, Chiron Corporation, remarks at the National Academies Board on Science, Technology, and Economic Policy’s Conference on “Intellectual Property Rights: How Far Should They Be Extended?” (Internet broadcast, Feb. 3, 2000).

291. DRATLER, supra note 253, § 4.03 (asserting that “a royalty base that extends beyond the scope of patent protection is lawful if accepted by both parties for their mutual convenience, for example, in calculating royalties and avoiding disputes”).
patented device but also unpatented components used in practice of the patented invention. The court characterized the inclusion of unpatented subject matter in the royalty base as merely a “convenient means for measuring the value of the license.”

Recent liberalization of the Federal Circuit’s damages jurisprudence provides additional support for the legitimacy of a reach-through royalty approach. The court has expanded traditional notions of recoverable damages in patent infringement cases to encompass virtually any type of economic harm that was “reasonably foreseeable” from the infringement. For example, damages may now be based on sales of infringing devices that do not directly compete in the marketplace with the patentee’s patented product, and lost profits may be awarded even when the patentee does not manufacture the patented device at all. Where the patented device is part of a larger system that also includes unpatented components, the “entire market value” rule recognizes that the damages award can be properly based on the value of the overall system. The Federal Circuit has specifically rejected arguments that its expansion of recoverable damages jurisprudence renders the reach-through royalty approach illegitimate.

---


293. Id. at 1408 (agreeing with magistrate judge that “royalties may be based on unpatented components if that provides a convenient means for measuring the value of the license”); see also Embrex, Inc. v. Serv. Eng’g Corp., 216 F.3d 1343, 1350 (Fed. Cir. 2000) (explaining that although compensatory damages based on “reasonable royalty” under 35 U.S.C. § 284 (1994) are “ordinarily computed based upon the sales of a patented product or process . . . parties may choose other methods to compute the amount that a licensee may pay for the right to use a patented product or process, such as flat fees or milestone payments in the case of pre-commercialization licenses”).

294. King Instruments Corp. v. Perego, 72 F.3d 855, 857 (Fed. Cir. 1995) (Nies, J., dissenting from denial of panel rehearing) (unpublished disposition) (characterizing 1995 Federal Circuit majority decision in Rite-Hite as having “expanded legal injury for patent infringement” and worked “fundamental change in patent rights”); see also Rite-Hite Corp. v. Kelly Co., 56 F.3d 1538, 1546 (Fed. Cir. 1995) (en banc) (holding that “[i]f a particular injury was or should have been reasonably foreseeable by an infringing competitor in a relevant market, broadly defined, that injury is generally compensable absent a persuasive reason to the contrary”).

295. Rite-Hite, 56 F.3d at 1549 (affirming award of lost profits damages based on lost sales of patentee’s ADL-100 vehicle-restraint device, which was not covered by patent in suit but directly competed with defendant’s infringing “Truk Stop” device).

296. King Instruments Corp. v. Perego, 65 F.3d 941, 947 (Fed Cir. 1995).

297. Id. at 950–51 n.4 (characterizing entire market value rule as recognizing that “the economic value of a patent may be greater than the value of the sales of the patented part alone,” and therefore permitting recovery based on value of unpatented as well as patented components of patentee’s product).
Experimental Use Exception to Patent Infringement

infringement damages facilitates patent owners’ restriction of competition in unpatented products; rather, the court’s stated objective is that “the patentee be made whole.”

The propriety of reach-through royalties is most problematic from a patent-misuse standpoint when the royalty payments extend beyond the enforceable life of the patent on the underlying research tool. This will likely occur with relative frequency in biotechnology, where long-term research and development projects (and regulatory approval of their results) often result in the marketplace introduction of a new product lagging several years behind the actual use of the patented research tools that led to the product’s development.

Rather than improper leveraging of the patent right, a better way to approach the use of reach-through royalties is simply as a time-shifting mechanism. Reach-through royalty payments continuing beyond the expiration date of an underlying research tool patent more accurately recognize the value of the patented research tool, which cannot be definitively established during the enforceable life of the patent. The premise underlying reach-through royalties is that the true value of the patented research tool will be determined by the ultimate marketplace

298. Rite-Hite, 56 F.3d at 1547 (explaining that court’s decision “simply asks, once infringement of a valid patent is found, what compensable injuries result from that infringement, i.e., how may the patentee be made whole”).

299. See Brulotte v. Thys Co., 379 U.S. 29, 31–32 (1964) (holding that license agreement requiring farmers to make royalty payments after expiration date of patent on hop-picking machine was unlawful per se where payments were for farmers’ use of machine during post-expiration period, not “deferred payments for use during the pre-expiration period”).

300. Cf. Harold See & Frank M. Caprio, The Trouble with Brulotte: The Patent Royalty Term and Patent Monopoly Extension, 1990 Utah L. Rev. 813 (1990). See and Caprio contend that Brulotte was incorrectly decided. Id. at 814. Licensees can be assumed to understand that a patentee’s right to exclude ends at the date of patent expiration. Id. Therefore, when licensees agree to make royalty payments beyond that date, “the parties base those payments on the licensees’ assessment of the value of the license during the patent period. These payments, therefore, do not represent an extension in time of the patent monopoly.” Id.; see also Mark A. Lemley, The Economic Irrationality of the Patent Misuse Doctrine, 78 Cal. L. Rev. 1599, 1630 (1990) (disagreeing with proposition that licensing practices which extend royalty payments beyond enforceable life of patent are per se patent misuse, and asserting that “[a] licensee will pay a fixed amount for a license, and the courts should not care whether the licensee pays that amount up front, in ten years, or in a hundred years”).

See and Caprio also dispute the Brulotte Court’s negative view of post-expiration royalty payments to the extent that such payments are perceived as unfairly imposing on the licensee a financial burden not shared by the licensor. See & Caprio, supra at 848 (noting concern that “[a]rguably, upon expiration of the patent, the royalty obligation impairs the licensee’s ability to compete with the licensor because only the licensee must pay a royalty”). The fallacy of this concern, See and Caprio point out, is that it assumes the licensor and licensee are competitors. Id. at 848. This was not the case in Brulotte, where the licensees were farmers using patented hop-picking machines, Brulotte, 379 U.S. at 31–32, nor would it generally be the case when researchers are using another’s research tool to produce a new drug or therapeutic product.
success of the new product developed through use of the tool. A reach-through license agreement merely time-shifts the royalty payments to the period when they are most accurately indicating the research tool’s true value to the user. That sales of the new product may extend beyond the life of the underlying tool patent will be neither surprising nor unknown to the patent owner and the tool user. Basing royalty payments on those sales should not be construed as patent misuse or as an antitrust violation.

Reach-through royalty payments are prima facie reasonable so long as the total time period over which they are paid is no longer than the term of the underlying tool patent, i.e., a period of twenty years less the patent application’s pendency.\textsuperscript{301} This limitation assures the patentee of obtaining full value for the researcher’s use and satisfies the constitutional requirement that exclusive rights in intellectual property are to be granted only “for limited times.”\textsuperscript{302}

The reach-through royalty approach is less straightforward when the non-consensual use of the patented research tool does not ultimately result in a commercial product; in these cases there is no end result to be “reached” as the royalty base. Nevertheless, the research tool user has obtained a benefit. Unproductive dead-ends have been identified and can be avoided in future research. The informational value of this understanding should not be underestimated; indeed, it may represent the largest cost component of other products that are ultimately developed and successfully marketed.\textsuperscript{303} To the extent that a specific research tool can be identified as instrumental in the development of a particular product, the royalty should be based on that product. In situations where this is not possible, however, a standardized schedule of royalty fees could be legislatively enacted. The mechanical license for musical works protected by copyright law suggests a possible model.\textsuperscript{304} Alternatively, a specialized administrative body similar

\textsuperscript{301}. For example, if a research tool patent is applied for in the United States in 2000 and issues in 2002, it will expire in 2020 (absent other extensions or term adjustments) and have an enforceable life of eighteen years. 35 U.S.C. § 154(a) (Supp. IV 1998). This Article proposes that reach-through royalties be paid beginning with the date of the first sale of the product developed through use of the patented tool, whenever that sale occurs, and ending eighteen years thereafter. The recovery period of eighteen years has been time-shifted to correspond with the period of sales of the new product.

\textsuperscript{302}. U.S. CONST., art. I, § 8, cl. 8.

\textsuperscript{303}. Gerth & Stolberg, supra note 166, at 1, 20 (stating that “only a small percentage” of $500 million average cost of developing new drug is attributable to actual development costs and that balance represents “cost . . . attributed to lost opportunities: years spent going down scientific ‘dry holes’ and research money that could have generated interest had it been invested instead”).

\textsuperscript{304}. 17 U.S.C. § 115 (Supp. IV 1998) (“Scope of exclusive rights in nondramatic musical works: Compulsory license for making and distributing phonorecords”). Once the owner of a copyright in a non-dramatic musical work has sold the work to the public, any other person has a compulsory license to record and sell phonorecords of the same work. See generally MARSHALL LEAFFER, UNDERSTANDING COPYRIGHT LAW § 8.7 (2d ed. 1995) (explaining how copyright mechanical license
to the Copyright Arbitration Royalty Panels could be created to adjudicate royalty disputes.\textsuperscript{305}

Adopting the notion of a royalty that will reach-through to use the downstream product as the royalty base, as proposed here, still leaves undetermined the royalty rate to be applied to that base.\textsuperscript{306} The difficulty of royalty quantification has been a leading argument against adoption of compulsory licensing in the United States.\textsuperscript{307} The determination of appropriate rates can be a very complex and expensive process.\textsuperscript{308} Absent
evidence of an established royalty rate, the traditional method of determining reasonable royalty in the patent litigation context involves application of the multi-factor “hypothetical license negotiation” framework of *Georgia-Pacific Corp. v. United States Plywood*. The Patent Act provides that a “reasonable royalty” represents the minimum amount of “compensation adequate to compensate for the infringement,” and invites the use of competing expert witnesses to assist the court in making this determination. Any such system, requiring elaborate individual litigations and competing expert testimony, would involve transaction costs far too high to be a solution of the research tools accessibility dilemma.

Recent scholarship suggests alternative methods of royalty rate determination that could short-circuit or at least simplify the elaborate multi-factor hypothetical negotiation method. These alternative methods could be adapted to a reach-through royalty approach where the royalty base is the commercial product rather than the research tool itself. One such method is a heuristic approach that involves payment of a royalty computed as twenty-five percent of the licensee’s pre-tax profit rate on its

309. It is unlikely that any established royalty rates would exist for patented research tools that have not been widely licensed or sold in the marketplace.

310. 318 F. Supp. 1116, 1120 (S.D.N.Y. 1970). Former Federal Circuit Chief Judge Howard Markey has termed the challenge of applying the *Georgia-Pacific* factors to hypothesize a license negotiation between willing parties an exercise in “fantasy and flexibility”:

The methodology encompasses . . . fantasy because it requires a court to imagine what warring parties would have agreed to as willing negotiators; flexibility because it speaks of negotiations as of the time infringement began, yet permits and often requires a court to look to events and facts that occurred thereafter and that could not have been known to or predicted by the hypothesized negotiators.

311. 35 U.S.C. § 284 (1994); see also *Rite-Hite Corp. v. Kelley Co.*, 56 F.3d 1538, 1544 (Fed. Cir. 1995) (en banc) explaining that purpose of reasonable-royalty provision of § 284 is to “set a floor below which damage awards may not fall”).

312. 35 U.S.C. § 284 (providing that “[a] court may receive expert testimony as an aid to the determination of damages or of what royalty would be reasonable under the circumstances”).

This twenty-five percent baseline figure may be “fine-tuned” as necessary to the circumstances of each individual case and apportioned between patent owners in the case of products developed through use of multiple patented research tools.

Another method is known as the “analytical approach,” which calculates the reasonable royalty to be paid to the patentee as the difference between the sales prices of the accused product and the sum of: (1) the infringer’s “direct or variable costs in the producing” the product; (2) the infringer’s “fixed costs, including allocated overhead to produce the article”; and (3) “normal profits to the infringer on similar products.” Thus, the analytical approach represents the “residual between the infringer’s anticipated net profit from practicing the infringed invention and the infringer’s normal net profit.” Reasonable royalty rates have also been estimated as a percentage of the infringer’s “net margin,” i.e., its operating income before taxes.

Despite the difficulties of assessing an appropriate royalty rate, the challenge of royalty quantification may be more of a problem in theory than practice. As foreign countries with compulsory licensing systems have recognized, the mere enactment of laws that contemplate judicial or administrative determination of licensing fees acts as an incentive to

---

314. The “twenty-five percent rule” was proposed by Robert Goldscheider based on his empirical observations of several successful commercial license negotiations in the 1950s. Robert Goldscheider, Measuring Damages in U.S. Patent Litigation, 5 J. PROPRIETARY RTS. 2, 6–7 (May 1993) (explaining that each licensee “earned about 20 percent pre-tax profit on sales and each paid a 5 percent royalty on sales—or 25 percent of the pre-tax profitability rate”). But see Toikka, supra note 313, at 292–93 (criticizing rule as overly simplistic for applying twenty-five percent split to total gross profit without distinguishing between monopoly profit and normal profit).


Goldscheider warns that because it computes the royalty based on the value of the entire product sold by the infringer, the analytical method is only appropriately used in circumstances where the “entire market value” rule can be applied; i.e., where the patented component is part of a larger product and it is the patented component that creates purchaser demand of the product. Goldscheider, supra note 314, at 8–9. Thus, the “analytical approach” would require modification if it were to be applied to compute a reasonable royalty for use of a patented research tool not incorporated into the defendant’s product.

317. Toikka, supra note 313, at 280 n.5.

voluntary licensing at more reasonable terms. This phenomenon may explain why few applications for compulsory licenses have actually been made in Europe. By analogy, the mere fact of legislative or judicial recognition in the United States of a broadened experimental use doctrine as proposed herein may encourage consensual agreements to reasonable licensing terms between research tool patent holders and research tool users.

VIII. CONCLUSION

The current narrow formulation of the experimental use doctrine in U.S. patent jurisprudence, which limits exemption from infringement liability to purely non-commercial, “philosophical” uses of patented inventions, means that the defense is not available to the vast majority of users of patented research tools. Many of these users are involved in collaborative efforts between universities and industry that require proprietary research tools to develop new therapeutic and diagnostic products, activity that by definition will involve some degree of commercialization or profit expectation. The proliferation of patents on research tools in the biotechnological and biomedical sector has resulted in stacking royalty obligations and heightened transaction costs that threaten to slow or stop the development of new drugs and devices critical to public health.

A potential solution is a “liability rule” model that permits the non-consensual “development use” of research tools not readily available for licensing or purchase, while providing an ex post royalty payment to the patent owner that would be correlated to the commercial value of the new product developed from the non-consensual use. This “reach-through” royalty approach provides the best approximation of the true worth of the research tool to its user. It ensures a royalty award of sufficient amount to maintain incentives for the development and patenting of new research tools, yet alleviates the access restrictions and up-front costs currently associated with acquisition and use of many proprietary research tools.

319. DRATLER, supra note 253, § 3.03[1][a] (contending that mere threat of royalty rate determination by judicial or administrative officials “may have the effect of spurring private voluntary transactions”); PENROSE, supra note 117, at 174–75 (explaining modest number of official adjudications of reasonable royalty in compulsory licensing context).

320. PENROSE, supra note 117, at 175 n.25 (recognizing that compulsory licensing schemes in Europe and Canada provide “every inducement for the patentee and the foreign concern desiring to use the invention to get together and settle their differences”).

321. Ayres & Talley, supra note 278, at 1094 (suggesting, in context of patented improvement inventions, that if liability rule such as compulsory licensing were applied, traditional perceptions of extensive litigation and courts’ inability to tailor royalty amount would actually facilitate parties’ bargaining “on their own terms, not those dictated by the underlying liability rule”).
RIGHTS, RIGHTS OF ACTION, AND REMEDIES: AN INTEGRATED APPROACH

Donald H. Zeigler*

Abstract: Traditionally, courts equated rights and remedies. Consequently, courts sought to provide remedies for the violation of statutory rights even if a statute did not contain detailed enforcement provisions. In the 1970s, however, the U.S. Supreme Court transformed what had been a unified inquiry into whether a statutory provision should be judicially enforceable into three distinct questions and developed separate criteria for deciding whether a statute should be read to create a right, imply a right of action, or provide a remedy. Rights, rights of action, and remedies are inextricably related. The Court’s attempt to separate these inseparable concepts has led to considerable confusion because decisions focusing on only one part of the equation fail to acknowledge the impact on other parts. Sometimes the Court disguises or actually misstates what it is doing. The Court has been wrong to divide rights, rights of action, and remedies and to develop separate tests to assess each of them. While a return to the traditional standards is impractical, the Court should integrate the separate tests and adopt a single test to answer what is actually a single inquiry: Does the applicable statutory provision entitle the plaintiff to the remedy he or she seeks? In answering that question, the Court should carefully examine statutory language, the overall statutory context, and possible reasons for caution in granting a remedy.

INTRODUCTION ...................................................................................68
I. THE TRADITIONAL STANDARDS AND THEIR DEMISE.....71
   A. The Traditional Standards .........................................................71
   B. The Modern Standards ..........................................................83
      1. Rights of Action .................................................................87
      2. Rights .................................................................................91
      3. Remedies ............................................................................95
II. TOWARD A NEW TEST FOR DECIDING WHETHER A STATUTORY PROVISION SHOULD BE JUDICIALLY ENFORCEABLE ...................................................105
   A. The Case for a Single, Integrated Standard .......................105
      1. The Interrelation of Rights, Rights of Action, and Remedies ..........................................................105
      2. Problems Caused by Trying To Separate the Inseparable .................................................................109

*  Professor of Law, New York Law School. A.B., Amherst College; J.D., Columbia University.
The author wishes to thank Brannon Heath, Randolf Jonakait, Richard Matasar, and Edward A. Purcell for their helpful comments on a draft of this Article.
INTRODUCTION

Traditionally, courts equated legal rights and remedies. A right without a remedy was said to be “a monstrous absurdity.” Consequently, whenever a right existed, courts strove to provide an appropriate remedy for its violation. Courts frequently applied this standard in actions to enforce a statute. Because legislatures often enacted statutes creating rights and duties without detailed enforcement provisions, courts filled the void. Moreover, they did so without conducting an independent inquiry into whether the statute provided a right of action. Violation of the right alone required a remedy.

2. See infra Part I.A.
3. This Article defines a legal right in Hohfeldian terms. A legal right is one that imposes a correlative duty on another to act or refrain from acting for the benefit of the person holding the right. Wesley N. Hohfeld, Some Fundamental Legal Conceptions as Applied in Judicial Reasoning, in FUNDAMENTAL LEGAL CONCEPTIONS AS APPLIED IN JUDICIAL REASONING AND OTHER LEGAL ESSAYS 35–38 (Walter W. Cook ed., 1919). In this context, a right of action, or cause of action, is the right “to seek judicial relief from injuries caused by another’s violation of a legal requirement.” Cannon v. Univ. of Chi., 441 U.S. 677, 730 n.1 (1979) (Powell, J., dissenting). A remedy is the relief a court grants, such as damages or an injunction. These definitions are discussed in detail infra Part II.
Although these principles did not guarantee effective redress in all cases, they remained the governing principles in the federal courts until the 1970s. Then, quite abruptly, the U.S. Supreme Court divided what previously had been a single, unified inquiry into whether a statutory provision should be judicially enforceable into three separate questions. The Court announced that rights, rights of action, and remedies are “analytically distinct” and developed separate criteria to decide whether a statute should be read to create a right, imply a right of action, or provide a remedy.

The Court’s trifurcated approach has caused several problems. Decisions that focus only on a right, a right of action, or a remedy are necessarily incomplete. Rights, rights of action, and remedies are inextricably related; one cannot make a decision about one of them without necessarily affecting the other two. The Court’s attempt to separate these inseparable concepts causes confusion. It also has led the Court to disguise and even misstate what it is doing.

This Article contends that the Court is wrong to divide rights, rights of action, and remedies and to use different tests to assess each of them. It further argues that one integrated test should be adopted to answer what is actually a single inquiry: Does the applicable statutory provision entitle a plaintiff to the remedy he or she seeks? Deciding what the test should be and how it should be applied is not a simple matter. How one

4. See infra notes 15–16 and accompanying text.

answers these questions depends on one’s views about underlying issues of federalism and separation of powers, and on several pragmatic concerns. To help lay the groundwork for a possible consensus, this Article presents two models governing the judicial enforcement of federal statutes denoted the “cooperative model” and the “adversarial model.” In formulating a new standard, it attempts to incorporate the generous and constructive aspects of the cooperative model and the cautionary concerns of the adversarial model.

This Article has three parts. Part I traces the development and application of the traditional standards equating rights and remedies from early English cases through U.S. Supreme Court decisions in the late 1960s. Part I then discusses the Court’s separation of rights, rights of action, and remedies in the 1970s. Finally, it reviews the Court’s development of separate criteria for deciding whether a statute should be read to confer a right, imply a cause of action, or provide a remedy.

Part II begins by making the case for the development of a single, integrated test to decide whether the statutory provision on which plaintiff relies entitles the plaintiff to the remedy he or she seeks. It explains why rights, rights of action, and remedies are inextricably related. Part II next explores the problems caused by the Court’s attempt to separate the inseparable. Part II then describes the two models governing judicial enforcement of federal statutes.

Part III proposes a new standard that seeks to find a middle ground between the two models. The test, which is an analytic process rather than a set of factors or criteria, begins with a careful look at the statutory language on which a plaintiff relies, examines the provision in its overall statutory context, and considers possible reasons for caution in granting a plaintiff the remedy sought.

The new test would lead to more honest court opinions that acknowledge the full interrelation of rights, rights of action, and remedies. Use of the test would change the results of some U.S. Supreme Court decisions. The biggest change would occur in implied right of action cases because the proposed test replaces the current single-factor test used in those cases with a broader, many-faceted standard. In the

7. The terms “cooperative” and “adversarial” refer to the relationship between the federal courts and Congress. The basic premise of the cooperative model is that federal courts should play a constructive, supportive role in interpreting and implementing federal legislation. The basic premise of the adversarial model is that federal courts should play a limited role in these endeavors, shifting most of the responsibility to Congress. See infra Part II.B.

8. The Court purports to consider “solely . . . whether Congress intended to create the private right of action asserted” by the plaintiff. Touche Ross & Co. v. Redington, 442 U.S. 560, 568 (1979) (emphasis added).
I. THE TRADITIONAL STANDARDS AND THEIR DEMISE

A. The Traditional Standards

The principle that rights must have remedies is ancient and venerable, and played an important role in English and American legal history. The principle underlies the rise of equity, the merger of law and equity, and the corresponding development of new codes of procedure. The principle also provided the impetus for actions to redress violations of statutory rights.

Early English and American cases enforcing statutes equated rights and remedies. As Professor H. Miles Foy states: “The essential notion . . . was that persons suffering legal wrongs were entitled to judicial remedies. What is more, they were entitled to adequate

---

9. Chancery courts developed because common law courts often did not provide complete and effective redress of legal wrongs. See George T. Bispham, The Principles of Equity: A Treatise on the System of Justice Administered in Courts of Chancery 9 (9th ed. 1915); William S. Holdsworth, A History of English Law 398 (7th ed. 1956). Equity courts enabled the English legal system to provide more complete relief. See John N. Pomeroy, A Treatise on Equity Jurisprudence 20–21 (5th ed. 1941) (“[T]he common law furnished a very meager system of remedies, utterly insufficient for the needs of a civilization advancing beyond the domination of feudal ideas.”). It became a maxim of equity jurisprudence that “equity will not suffer a right to be without a remedy.” Bispham, supra, at 56; see generally Zeigler, supra note 6, at 667–69.

10. Law and equity were merged so that legal rights could be better enforced. Zeigler, supra note 6, at 669. In both England and America, the existence of two separate judicial systems caused confusion and inconvenience. See Roscoe Pound, The Decadence of Equity, 5 Colum. L. Rev. 20, 23 (1905); Edward R. Taylor, The Fusion of Law and Equity, 66 U. Pa. L. Rev. 17, 23 (1917). With the merger of law and equity, litigants could obtain both legal and equitable relief in one lawsuit. See William F. Walsh, Merger of Law and Equity Under Codes and Other Statutes, 6 N.Y.U. L. Rev. 157, 169 (1929).

11. State legislatures promulgated new codes of civil procedure to better ensure the vindication of rights. See Charles M. Hepburn, The Historical Development of Code Pleading in America and England 21 (1897) (“[The code movement’s] one great purpose was to bring procedure into a simple and natural relation with substantive law . . . [and] to give a natural and vigorous vitality to a maxim which the law had long placed before itself as the ideal—wherever a right, there a remedy.”).

12. Actions on statutes probably have their origin in the early English Statute of Westminster II, which authorized an action on the case for those injured by breach of a statutory duty. The statute read: “Moreover, concerning the Statutes provided where the Law faileth, and for Remedies, lest Suitors coming to the King’s Court should depart from thence without Remedy, they shall have Writs provided in their Cases . . . .” Statute of Westminster II, 1285, 13 Edw., ch. 50, § 2 (Eng.).

13. See Foy, supra note 6, at 526–32.
Of course, this ideal was not always realized in practice. As Professors Richard H. Fallon and Daniel J. Meltzer point out, “the structure of substantive, jurisdictional, and remedial doctrines that existed [in the eighteenth century] and that evolved through the nineteenth century by no means guaranteed effective redress for all invasions of legally protected rights and interests.” Nonetheless, people in those times believed that rights required remedies. In Justice Harlan’s words, “contemporary modes of jurisprudential thought . . . appeared to link ‘rights’ and ‘remedies’ in a 1:1 correlation.”

In addition, when courts enforced rights in statutes that did not contain express remedies, they did not say that they were creating or implying a “cause of action” or a “private right of action” from the statute. The absence of such statements suggests that in this context courts did not view a cause of action as a separate procedural entity, independent of a right and remedy, that had to be present for an action to go forward. Thus, modern references to early English and American cases as “implied right of action” cases may mischaracterize what the courts were doing. Some early opinions suggested that an action on the case could be used to enforce a statutory right, and perhaps the form of action provided something roughly analogous to what we would today call a cause of action. But early cases plainly did not inquire whether the statute at issue provided a separate cause of action or make a separate

14. Id. at 529.
15. As Professor Paul Gewirtz notes, the law of remedies is by nature a “jurisprudence of deficiency, of what is lost between declaring a right and implementing a remedy.” Paul Gewirtz, Remedies and Resistance, 92 YALE L.J. 585, 587 (1983).
16. Richard H. Fallon, Jr. & Daniel J. Meltzer, New Law, Non-Retroactivity, and Constitutional Remedies, 104 HARV. L. REV. 1731, 1780 (1991). The authors note that common law privileges stood as barriers to full relief and that remedies were very limited in suits against the government and government officials. Id. at 1780–81.
19. See, e.g., California v. Sierra Club, 451 U.S. 287, 299–300 & n.3 (1981) (Stevens, J., concurring) (stating that “implication of private causes of action was a well-known practice at common law and in American courts” and citing early English treatises and cases); Stewart & Sunstein, supra note 6, at 1206 & n.39 (stating that “[a]t common law, courts created private rights of action . . . by creating an action in damages for statutory wrongs” and citing early English treatises and cases); Patrick B. Fazzone, Comment, Implied Rights of Action in Federal Legislation: Harmonization Within the Statutory Scheme, 1980 DUKE L.J. 928, 929 n.2 (“The doctrine of implied rights of action has been traced to an English case, Couch v. Steel, 118 Eng. Rep. 1193 (K.B. 1854).”).
decision whether to create or deny a private right of action. Violation of the right alone required a remedy.

_Ashby v. White_ 21 provides an example of courts equating rights and remedies without mention of a cause of action. The plaintiff claimed an official had improperly denied him the right to vote in an election and sought damages. 22 Chief Justice Holt cited an ancient statute conferring that right and concluded the plaintiff should have a remedy even though the statute did not provide one:

If the plaintiff has a right, he must of necessity have a means to vindicate and maintain it, and a remedy if he is injured in the exercise or enjoyment of it; and indeed it is a vain thing to imagine a right without a remedy; for . . . want of right and want of remedy are reciprocal. . . . Where a man has but one remedy to come at his right, if he loses that he loses his right. 23

Early American cases also equated rights and remedies and did not identify a cause of action as a separate requirement for relief. 24 _Marbury_
v. Madison\textsuperscript{25} provides a famous example. Marbury brought a mandamus action to compel the Secretary of State to deliver a commission signed by the President appointing him to the bench.\textsuperscript{26} Chief Justice Marshall quoted Blackstone concerning the relation between right and remedy:

```
"[I]t is a general and indisputable rule, that where there is a legal right, there is also a legal remedy by suit or action at law, whenever that right is invaded. . . . [F]or it is a settled and invariable principle in the laws of England, that every right, when withheld, must have a remedy, and every injury its proper redress."\textsuperscript{27}
```

Kendall v. United States\textsuperscript{28} provides another such example. Several individuals sought payment from the Postmaster General for performance of a contract to deliver mail.\textsuperscript{29} Congress passed a statute directing the Solicitor of the Treasury to investigate and to determine the equities of the matter.\textsuperscript{30} When the Postmaster refused to pay the full amount the Solicitor recommended, the individuals sued.\textsuperscript{31} In granting relief, the Court stated:

```
It cannot be denied but that congress had the power to command that act to be done; and the power to enforce the performance of the act must rest somewhere, or it will present a case which has often been said to involve a monstrous absurdity in a well organized government, that there should be no remedy, although a clear and undeniable right should be shown to exist.\textsuperscript{32}
```

\textsuperscript{25}. 5 U.S. (1 Cranch) 137 (1803).
\textsuperscript{26}. Id. at 153–54.
\textsuperscript{27}. Id. at 163 (quoting 3 William Blackstone, Commentaries *23). Chief Justice Marshall continued: "The government of the United States has been emphatically termed a government of laws, and not of men. It will certainly cease to deserve this high appellation, if the laws furnish no remedy for the violation of a vested legal right." Id.
\textsuperscript{28}. 37 U.S. (12 Pet.) 524 (1838).
\textsuperscript{29}. Id. at 608.
\textsuperscript{30}. Id. at 608–09.
\textsuperscript{31}. Id.
\textsuperscript{32}. Id. at 624. Franklin v. Gwinnett County Public Schools, 503 U.S. 60 (1992), mischaracterizes the statute involved in Kendall. Speaking for the majority, Justice White stated that the Act of Congress "accorded a right of action in mail carriers to sue for adjustment and settlement of certain claims for extra services but . . . did not specify the precise remedy available to the carriers." Id. at 67. As noted above, the statute was in the nature of a private bill that merely directed the Solicitor of the Treasury to look into the mail carriers' claim and recommend an equitable solution. Kendall, 37 U.S. at 528–29. The statute did not create a "right of action in mail carriers to sue." Kendall did not appear to consider a private right of action an independent or necessary part of the right-remedy equation. The modern Court is so accustomed to viewing a cause of action as a separate and necessary component of a lawsuit that it mistakenly read the statute in Kendall to create one.
Again, the Court closely related right and remedy without mentioning a cause of action.

During the late 1800s and early 1900s, American state courts routinely allowed private remedies for violations of statutes containing other sanctions, although sometimes on a slightly different legal theory. Violation of a statute was said to constitute "evidence of negligence" or "negligence per se." Plaintiffs could recover if they were members of the class of persons for whose benefit the statute was intended and the harm suffered was of a kind that the statute generally was intended to prevent. Arguably these cases recognize a cause of action for


34. See, e.g., Vandewater v. N.Y. & New Eng. R.R., 32 N.E. 636, 636–37 (N.Y. 1892) (stating that violation of statute could be evidence of negligence, but reversing and remanding for new trial because applicable statute had been repealed at time of accident); McRickard v. Flint, 21 N.E. 153, 153 (N.Y. 1889) (stating that failure to perform statutory duty "is evidence upon the question of negligence").

The federal courts occasionally relied on this theory. For example, in *Hayes v. Michigan Central Railroad*, 111 U.S. 228 (1884), the plaintiff was a young boy whose left arm was severed in a collision with defendant’s train. *Id.* at 231–32. The plaintiff claimed that the railroad had violated a Chicago ordinance granting it a right of way on the condition that it erect fences along the rail line to protect persons and property from danger. *Id.* at 229–30. The Court held that the failure to fence would support an action for personal injury, and that “this breach of duty will be evidence of negligence.” *Id.* at 240. The railroad argued that it could not be liable to a member of the public injured by violation of the ordinance, *id.* at 233, but the Court disagreed. Because the City had enacted the ordinance to protect the public, “considered as composed of individual persons,” it followed that “each person specially injured by the breach of the obligation is entitled to his individual compensation, and to an action for its recovery.” *Id.* at 240.


[It is an] axiomatic truth that every person while violating an express statute is a wrong-doer, and, as such, is ex necessitate negligent in the eye of the law, and every innocent party whose person is injured by the act which constitutes the violation of the statute is entitled to a civil remedy for such injury, notwithstanding any redress the public may also have.

*Id. But see* Brown v. Buffalo & State Line R.R. Co., 22 N.Y. 191, 195 (1860) (rejecting argument of plaintiff that violation of city ordinance prohibiting locomotives from running at more than six miles per hour “was alone evidence of carelessness, sufficient to charge the defendant with the consequences of the collision”). For support of the negligence per se rule, see Ezra R. Thayer, *Public Wrong and Private Action*, 27 HArv. L. REV. 317 (1914). The development of the theory that violation of a statutory duty constitutes negligence per se is discussed in Foy, supra note 6, at 540–48.

36. See KEETON ET AL., supra note 33, at 224–25. These standards were incorporated in the RESTATEMENT OF TORTS § 286 (1934): Violations Creating Civil Liability.
negligence as a separate element of the right-remedy equation. Negligence, however, was a general, pre-existing common law cause of action. The courts did not purport to imply a new cause of action from the specific statutes in question. Nor did they apply any special criteria for deciding whether to use the negligence theory, other than the general requirements noted above. The opinions also stressed the close relation between rights and remedies.

Courts did not automatically grant relief every time a plaintiff came to court seeking to enforce a provision of a statute. Courts sometimes concluded that the statutory language did not actually create the rights and duties that the plaintiffs claimed or that granting relief would frustrate rather than further legislative intent. Courts also denied relief

---

The violation of a legislative enactment by doing a prohibited act, or by failing to do a required act, makes the actor liable for an invasion of an interest of another if:

(a) the intent of the enactment is exclusively or in part to protect an interest of the other as an individual;

(b) and the interest invaded is one which the enactment is intended to protect;

(c) and, where the enactment is intended to protect an interest from a particular hazard, the invasion of the interests results from that hazard;

(d) and, the violation is a legal cause of the invasion, and the other has not so conducted himself as to disable himself from maintaining an action.

Id.

37. See WILLIAM A. PROSSER, THE LAW OF TORTS 139–40 (1971) (“About the year 1825, negligence began to be recognized as a separate and independent basis of tort liability.”).

38. The requirements that the statute be intended at least in part to protect the class of persons of which the plaintiff is a member and that the harm suffered was of a kind the statute generally was intended to prevent are plainly satisfied by, or are at least consistent with, the early English and American cases. The statute in Ashby v. White, 92 Eng. Rep. 126 (K.B. 1703), protecting the right to vote, plainly was intended to protect voters like Mr. Ashby and to ensure that voters were not improperly denied the right to vote. Id. at 135–36. Similarly, the statute in Couch v. Steel, 118 Eng. Rep. 1193 (K.B. 1854), requiring sufficient medicines on board during sea voyages clearly was intended to protect members of the crew like Couch and to prevent unnecessary suffering and sickness due to lack of proper medicine. Id. at 1196. The statute in Kendall v. United States, 37 U.S. (12 Pet.) 524 (1838), directing the Solicitor of the Treasury to look into the plaintiffs’ claim against the Postmaster was intended specifically to aid the plaintiffs in obtaining fair compensation under their contract to deliver mail. Id. at 608–09.

39. See, e.g., Parker v. Barnard, 135 Mass. 116, 120 (1883) (“The fact that there was a penalty imposed by the statute for neglect of duty in regard to the railing and protection of the elevator well does not exonerate those responsible therefor from such liability.”); Stout v. Keyes, 2 Doug. 184, 186 (Mich. 1845) (“It is a general principle of the common law, that whenever the law gives a right, or prohibits an injury, it also gives a remedy by action; and, where no specific remedy is given for an injury complained of, a remedy may be had by special action on the case.”); Martin v. Herzog, 126 N.E. 814, 816 (N.Y. 1920) (Cardozo, J.) (“A statute designed for the protection of human life is not to be brushed aside as a form of words, its commands reduced to the level of cautions, and the duty to obey attenuated into an option to conform.”).

when the legislature specifically forbade a particular remedy. But even when courts denied relief, they conducted an integrated inquiry that treated the claimed right and proposed remedy as a unit.

In 1916, the U.S. Supreme Court decided the often-cited case Texas & Pacific Railway v. Rigsby. Rigsby strongly reaffirmed the equation of rights and remedies, but it also appeared to treat a cause of action as a separate part of the equation, thus foreshadowing the radical developments in the later part of the twentieth century. The plaintiff was a switchman for the railway company. He was injured when a defective handhold gave way as he descended from the top of a box car. He sued for damages under a federal statute requiring trains in interstate commerce to have secure handholds, and the Court held that such relief was available. The Court reiterated the close, reciprocal relation between rights and remedies:

A disregard of the command of the statute is a wrongful act, and where it results in damage to one of the class for whose especial benefit the statute was enacted, the right to recover the damages

41. See supra note 23.
42. Pollard provides a good example of an integrated right-remedy inquiry in a case denying relief without mention of a cause of action as a separate requirement. In 1854 the Alabama legislature chartered a new bank. Pollard, 87 U.S. at 521. The legislation contained provisions in case the bank became insolvent. Id. at 521–22. Individual stockholders were to be personally liable for all outstanding debts of the bank in proportion to their share of the stock. Id. at 521. If the bank could not pay a debt, the creditor could file a bill in the chancery court asking for a temporary injunction freezing the bank’s assets and for appointment of a receiver. Id. After distributing the bank’s assets, remaining debts could be assessed against each of the stockholders in proportion to his shares of stock. Id. at 521–22. In due course the bank became insolvent. Id. at 522. Bailey, a creditor, brought an action at law against one of the stockholders, Pollard, seeking $17,000, the full amount that Bailey was owed. Id.

The Court held that the Alabama statute did not grant Bailey a right to the relief that he claimed. Id. at 524–25. Pollard had a duty to pay only his pro rata share of the debts. Id. at 525. This proportion could be determined only after a pro rata distribution of the indebtedness among all the stockholders in an equity action. Id. Granting relief to Bailey would frustrate the clear legislative intent “only to charge the stockholders upon a proper account and in the manner” provided by the statute. Id. Moreover, the legislature also intended to distribute funds for the common benefit of all of the creditors. Id. at 527. The Court thought that “[e]very provision [of the statute] is entirely inconsistent with the idea that one creditor could, by an individual suit, appropriate to himself the entire benefit of the security, and exclude all others.” Id. The court concluded with language that showed the close relation between the duty (and the right) and the remedy: “There was no liability except for the deficiency. That was to be apportioned and collected for the common benefit . . . . The liability and the remedy were created by the same statute. This being so the remedy provided is exclusive of all others.” Id.
43. 241 U.S. 33 (1916).
44. Id. at 36.
45. Id. at 37.
46. Id. at 40.
from the party in default is implied, according to a doctrine of the common law expressed . . . in these words: “So, in every case, where a statute enacts, or prohibits a thing for the benefit of a person, he shall have a remedy upon the same statute for the thing enacted for his advantage, or for the recompense of a wrong done to him contrary to the said law.” (Per Holt, C.J., Anon., 6 Mod. 26, 27.) This is but an application of the maxim, Ubi jus ibi remedium. See 3 Black. Com. 51, 123; Couch v. Steel, 3 El. & Bl. 402, 411; 23 L. J. Q. B. 121, 125. This language is often quoted by the U.S. Supreme Court and commentators and is generally seen as a ringing endorsement of traditional conceptions of rights and remedies. However, Rigsby also made several specific references to “a right of action” in the sentences immediately before and after the language quoted above, thus suggesting that the Court viewed a cause of action as a separate component of the rights-remedies equation. Moreover, although the language is ambiguous, the Court may have been suggesting that it was inferring the cause of action from the statute itself. Thus, Rigsby is a perplexing case in that it

---

47. Id. at 39–40. The term ubi jus ibi remedium is translated “[w]here there is a right, there is a remedy.” BLACK’S LAW DICTIONARY 1695 (7th ed. 1999).


49. See, e.g., Merrill Lynch, 456 U.S. at 374–75 (“Under this approach, federal courts, following a common-law tradition, regarded the denial of a remedy as the exception rather than the rule.”); Sierra Club, 451 U.S. at 299–300 (“[In 1890,] Members of Congress merely assumed that the federal courts would follow the ancient maxim, ‘ubi jus, ibi remedium.’”); Creswell, supra note 48, at 975 (“Justice Pitney emphasized that he was applying a well-recognized common-law doctrine.”); Foy, supra note 6, at 554 (“Justice Pitney wrote for the Court in Rigsby . . . . Justices Story or Marshall could have written the opinion. Indeed, Coke or Chief Justice Holt could have written it. Rigsby looked to the past, not to the future.”).

50. Referring to different versions of the Federal Safety Appliance Act in the sentence immediately before the quoted language, the Court stated:

None of the Acts, indeed, contains express language conferring a right of action for the death or injury of an employee; but the safety of employees and travelers is their principal object, and the right of private action by an injured employee, even without the Employers’ Liability Act, has never been doubted.

Rigsby, 241 U.S. at 39. In the sentence immediately following the quoted language, the Court stated:

“The inference of a private right of action in the present instance is rendered irresistible” by a provision of the Act stating that an employee injured by any car “in use contrary to the act shall not be deemed to have assumed the risk.” Id. at 40.

51. The block quote in supra note 50 suggests this possibility because it focuses so specifically on provisions of the Act. But see Cannon v. Univ. of Chi., 441 U.S. 677, 732 (1979) (Powell, J.,
Rights, Rights of Action, and Remedies

endorses the traditional close relation between rights and remedies while possibly introducing a new element into the equation.

While the seeds of current law treating rights, rights of action, and remedies as separate elements can be seen in *Rigsby*, in the years that followed the Court chose to reaffirm the close interrelation of rights and remedies, usually without mentioning a cause of action. In several cases seeking enforcement of the Railway Labor Act of 1926, the Court granted remedies that were not explicitly authorized by the statute. In *Bell v. Hood*, the Court reasoned that the remedies must be available or else the rights conferred by the statute would not exist, and it made no reference to a cause of action as a separate requirement. In *American Well Works Co. v. Layne & Bowler Co.*, 241 U.S. 257 (1916), the Court later talked about a cause of action in connection with subject matter jurisdiction. See generally *Steele v. Louisville & Nashville R.R. Co.*, 323 U.S. 192 (1944); *Tex. & New Orleans R.R. Co. v. Bhd. of Ry. & S.S. Clerks*, 281 U.S. 548 (1930).

Interestingly, the Court may have suggested that a cause of action is separate from a right and remedy, but the suggestion is only implicit. The plaintiff in *American Well Works* alleged that defendants maliciously and falsely libeled and slandered plaintiff’s title to a pump by asserting that the pump infringed defendants’ patent. Id. at 258. The Court held that the case did not arise under federal law because it was at base a tort case for defamation or for unlawful damage to plaintiff’s business. Id. at 259–60. Justice Holmes announced the famous test that “[a] suit arises under the law that creates the cause of action,” id. at 260, and since the plaintiff’s cause of action was state-created, the case arose under state law. It is unclear from Holmes’ opinion whether he was identifying a cause of action as a separate procedural entity or simply using the phrase as a short-hand to refer to the merits of the plaintiff’s claims. The Court appears to have denied federal question jurisdiction because plaintiff asserted state-created rights and duties that did not turn on any questions of federal law.

As the Court admitted in 1979, *Texas & New Orleans Railroad Co.* “is now understood as having implied a ‘cause of action’ although the opinion itself did not use the phrase.” Davis v. Passman, 442 U.S. 228, 239 n.17 (1979). In *Steele*, another case decided under the Railway Labor Act, the Court explicitly recognized that rights and duties do not actually exist without a remedy. *Steele* held that the Act imposed a duty on the exclusive bargaining representative of railroad firemen to represent all of the firemen without discrimination because of their race. 323 U.S. at 199–202. The Court stated:

[T]he right here asserted, to a remedy for breach of the statutory duty . . . is of judicial cognizance. That right would be sacrificed or obliterated if it were without the remedy which courts can give for breach of such a duty or obligation and which it is their duty to give in cases in which they have jurisdiction.
plaintiffs sought damages for violation of constitutional rather than statutory rights. The Court once again affirmed the broad availability of remedies to enforce all legal rights and did not identify a private right of action as a separate component of the rights-remedies equation:

[W]here federally protected rights have been invaded, it has been the rule from the beginning that courts will be alert to adjust their remedies so as to grant the necessary relief. And it is also well settled that where legal rights have been invaded, and a federal statute provides for a general right to sue for such invasion, federal courts may use any available remedy to make good the wrong done.55

The reign of the traditional standards continued through the 1960s. In J.I. Case Co. v. Borak,56 the Court allowed a shareholder of J.I. Case to

---

54. 327 U.S. 678 (1946).
55. Id. at 684 (citations omitted). Looking only at the quoted language, one might read the phrase “and a federal statute provides for a general right to sue for such invasion” to refer to a cause of action. Indeed, Justice White made this mistake in Guardians Ass’n v. Civil Service Commission, 463 U.S. 582 (1983), when he cited Bell for the proposition that “where legal rights have been invaded and a cause of action is available, a federal court may use any available remedy to afford full relief.” Guardians Ass’n, 463 U.S. at 595. However, the full opinion in Bell makes clear that the phrase refers to subject matter jurisdiction under the general federal question section of the United States Code, now 28 U.S.C. § 1331 (1994). Bell, 327 U.S. at 679. Plaintiffs claimed that federal agents violated their Fourth and Fifth Amendment rights. Id. The lower courts held that they lacked subject matter jurisdiction to hear such a claim. Id. at 680. The U.S. Supreme Court reversed on this point, holding that the case arose under federal law because the case necessarily turned on the scope of the protection afforded by the constitutional amendments. Id. at 684–85. The Court did not describe the issue facing the lower courts on remand as whether a private right of action could be inferred from the Fourth and Fifth Amendments. Instead, the Court stated that “[i]f the issue of law is whether federal courts can grant money recovery for damages said to have been suffered as a result of federal officers violating the Fourth and Fifth Amendments,” id. at 684, and the statements quoted in the text are some ruminations about that question. In any event, it is plain that the phrase “and a federal statute provides for a general right to sue for such invasion” could not refer to a section of the U.S. Code authorizing a damage action against federal agents for violation of the Constitution because no such section existed. Thus, the quoted language means that authority to grant any available remedy flows from a simple grant of subject matter jurisdiction. The Court does mention a cause of action earlier in the opinion, but in connection with the merits of plaintiffs’ claim:

Jurisdiction, therefore, is not defeated as respondents seem to contend, by the possibility that the averments might fail to state a cause of action on which petitioners could actually recover. For it is well settled that the failure to state a proper cause of action calls for a judgment on the merits and not for a dismissal for want of jurisdiction.

Id. at 682.
sue under §§ 14(a) and 27 of the Securities Exchange Act of 1934 for damages and to rescind a merger allegedly effected through false and misleading proxy statements. The Court brushed aside defendants’ contention that the Exchange Act made no reference to a private right of action under § 14(a), stating that among the Act’s “chief purposes is ‘the protection of investors,’ which certainly implies the availability of judicial relief where necessary to achieve that result.” The Court also noted that private enforcement of the proxy rules “provides a necessary supplement to Commission action” due to the volume of proxy statements submitted annually. Citing the liberal remedial language of several of the cases discussed above, the Court allowed the suit to proceed.

57. Section 14(a) of the Exchange Act, along with Rule 14a-9, which was promulgated pursuant thereto by the Securities and Exchange Commission, made it unlawful to issue false and misleading proxy statements. See Securities Exchange Act of 1934 § 14(a), 15 U.S.C. § 78j(b) (1994); 17 C.F.R. § 240.14a-9 (2000). Section 27 is a jurisdictional provision that gives the federal district courts exclusive jurisdiction “of all suits in equity and actions at law brought to enforce any liability or duty created by this title or the rules and regulations thereunder.” Securities Exchange Act of 1934 § 27, 15 U.S.C. § 78aa; J.I. Case Co., 377 U.S. at 427–33.

58. J.I. Case Co., 377 U.S. at 432.

59. Id.

60. Id. at 435. Two decisions in 1960 reaffirmed the Court’s broad power to order equitable remedies for violation of statutory duties. In United States v. Republic Steel Corp., 362 U.S. 482 (1960), the Court read § 10 of the Rivers and Harbors Act of 1899 to authorize the Attorney General to seek injunctive relief ordering the steel company to dredge a riverbed where it had deposited industrial solids. Id. at 485. The Act imposed a broad duty not to “obstruct” navigable waters and gave the Attorney General power to enforce the Act. Id. at 491–92. The Court stated: “Congress has legislated and made its purpose clear; it has provided enough federal law in § 10 from which appropriate remedies may be fashioned even though they rest on inferences. Otherwise we impute to Congress a futility inconsistent with the great design of this legislation.” Id. at 492.

61. Id. at 435. In Mitchell v. Robert DeMario Jewelry, 361 U.S. 288 (1960), the Court read §§ 15(a)(3) and 17 of the Fair Labor Standards Act of 1938 to allow the Secretary of Labor to bring an action on behalf of DeMario Jewelry employees seeking wages unpaid in violation of the Act. Id. at 289. Section 15(a)(3) made it unlawful for an employer to discharge or otherwise discriminate against an employee because the employee filed a complaint under the Act, but it did not specify a remedy. Id. Section 17 merely gave the district courts jurisdiction to restrain violations of § 15. Id. The Court held that the remedy sought by the Secretary was easily within the Court’s equitable power to enforce the statute:

“[A]ll the inherent equitable powers of the District Court are available for the proper and complete exercise of that jurisdiction . . . . [T]he court may go beyond matters immediately underlying its equitable jurisdiction . . . and give whatever other relief may be necessary under the circumstances. . . . Moreover, the comprehensiveness of this equitable jurisdiction is not to be denied or limited in the absence of a clear and valid legislative command.”

Id. at 291 (quoting Porter v. Warner Holding Co., 328 U.S. 395, 397–98 (1946)). The Court reasoned that an action for reimbursement of lost wages was necessary to achieve congressional purposes. Id. Congress relied on information and complaints from employees to enforce the Act. Id. at 292.
In the late 1960s, the Court also authorized judicial proceedings to enforce provisions of important civil rights statutes. The Court relied heavily on the traditional equation of rights and remedies in authorizing private suits to help achieve the broad remedial purposes of the legislation. None of these cases made reference to a private right of action as a separate procedural requirement. 62

As in the past, 63 the Court did not grant relief to every plaintiff who sought a remedy for an alleged violation of a statutory duty. The Court denied relief if it thought the statute did not actually confer the plaintiff’s claimed right 64 or that to grant relief would directly conflict with con-

employees could obtain only prospective relief reinstating them to their jobs, they would be reluctant to complain to officials about violations of the Act because they could not afford to lose their pay during the period while they sought reinstatement. Id. at 292–93.

62. See, e.g., Allen v. State Bd. of Elections, 393 U.S. 544 (1969). The Allen Court allowed private individuals to seek declaratory judgments that changes in state voting laws were subject to § 5 of the Voting Rights Act of 1965, and thus, could not be implemented until the state complied with the approval requirements of the Act. Allen, 393 U.S. at 554–57. The Court stated that the Act’s basic goal to make the Fifteenth Amendment a reality for all citizens would be “severely hampered” if citizens were required to rely solely on litigation instituted at the discretion of the Attorney General. Id. at 556. Similarly, Jones v. Alfred H. Mayer Co., 392 U.S. 409 (1968), allowed a private suit to enforce 42 U.S.C. § 1982, which provides that “[a]ll citizens . . . shall have the same right . . . as is enjoyed by white citizens . . . to . . . purchase . . . real and personal property.” The Court stated:


Jones, 392 U.S. at 414 n.13. Finally, Sullivan v. Little Hunting Park, 396 U.S. 229 (1969), authorized a private damage action under § 1982, which provides that “[a]ll citizens . . . shall have the same right . . . as is enjoyed by white citizens . . . to . . . purchase . . . real and personal property.” The Court quoted from Riggsby, Texas & New Orleans Railroad Co., and Bell. Id. at 238–39.

In another important decision in this period that applied traditional standards, Wyandotte Transportation Co. v. United States, 389 U.S. 191 (1967), the Court allowed the government to sue to recover its expenses in removing a sunken barge from the Mississippi River. Id. at 193, 204. The Rivers and Harbors Act of 1899 did not specifically authorize such a suit, but § 15 of the Act made it unlawful negligently to sink a vessel in navigable waters and imposed a duty on the owner to remove it. Id. at 197. The Court stated that allowing the government to sue for its expenses when the owner refused to remove the barge was consistent with the overall purposes of the statute and avoided an unfair windfall to the owner. Id. at 201–04.

63. See supra notes 41–42 and accompanying text.

64. For example, in Montana-Dakota Utilities Co. v. Northwestern Public Service Co., 341 U.S. 246 (1951), the Court held that a provision of the Federal Power Act requiring utilities to charge just and reasonable rates “creates no right which courts may enforce.” Id. at 251. The parties in the case were power companies, and one charged the other with fraud in rate setting that occurred when the companies had an interlocking directorate. See id. at 247–48. The district court found the rates unreasonable, determined what would have been reasonable rates, and gave judgment for the difference. See id. at 248. The U.S. Supreme Court stated the district court was wrong “to regard
gressional intent. But such cases did not purport to change the governing standards or to identify a cause of action as a separate component of the rights-remedies equation.

B. The Modern Standards

The late 1960s were the final days of the traditional standards. The Court made radical changes in the years that followed, dividing what previously had been a single, unified inquiry about whether a statute created judicially enforceable obligations into three separate and distinct questions. While traditionally the Court equated rights and remedies and rarely mentioned a cause of action, suddenly the Court sharply differentiated between rights, rights of action, and remedies. Moreover, the Court

reasonableness as a justiciable legal right rather than a criterion for administrative application in determining a lawful rate.” *Id.* at 251. Any remedy thus lay with the Federal Power Commission, not the courts. *See id.*

This decision evoked a spirited dissent. Justice Frankfurter, writing for himself and three other justices, thought that the traditional rights-remedies standards required a different result. *See id.* at 261–62 (Frankfurter, J., dissenting). He noted that the administrative remedies suggested by the majority were inadequate and that the “aim of Congress would be needlessly aborted if this ‘definite statutory prohibition of conduct’ did not impose civil liability . . . merely because no judicial relief was explicitly authorized.” *Id.* at 262 (Frankfurter, J., dissenting). He strongly reaffirmed the traditional standards:

*Texas & N.O. R. Co. v. Brotherhood of R. & S.S. Clerks, [281 U.S. 548]; Virginian R. Co. v. System Federation, No. 40, [300 U.S. 515 (1937)]; Deckert v. Independence Shares Corp., [311 U.S. 282 (1940)].* A duty declared by Congress does not evaporate for want of a formulated sanction. When Congress has “left the matter at large for judicial determination,” our function is to decide what remedies are appropriate in light of the statutory language and purpose and of the traditional modes by which courts compel performance of legal obligations. . . . If civil liability is appropriate to effectuate the purposes of a statute, courts are not denied this traditional remedy because it is not specifically authorized.


*Id.* at 261–62 (Frankfurter, J., dissenting); *see also T.I.M.E. Inc. v. United States, 359 U.S. 464, 469–72 (1959) (following *Montana-Dakota* approach in declining to imply cause of action for shippers against carriers under Motor Carrier Act and also noting that Congress had twice considered and rejected attempts to amend statute to grant shippers such cause of action).

65. A court generally would not conclude that the legislature intended to preclude a particular remedy unless it explicitly so stated. See, e.g., Porter v. Warner Holding Co., 328 U.S. 395, 398 (1946) (“Moreover, the comprehensiveness of this equitable jurisdiction is not to be denied or limited in the absence of a clear and valid legislative command.”); Hecht v. Bowles, 321 U.S. 321, 329–30 (1944) (“[I]f Congress had intended to make such a drastic departure from the traditions of equity practice, an unequivocal statement of its purpose would have been made.”).
soon adopted separate criteria to decide whether a statute should be read
to create a right, imply a cause of action, or provide a remedy.

Two cases in the mid-1970s signaled that change was afoot. In *National Railroad Passenger Corp. v. National Ass’n of Railroad
Passengers*66 and *Securities Investor Protection Corp. v. Barbour*,67 the Court focused solely on whether a private right of action in favor of the
plaintiffs could be inferred from the statute involved.68 In *National Railroad Passenger Corp.*, plaintiffs sought to enjoin discontinuance of
certain passenger trains, claiming that procedures required by the Rail
Passenger Service Act of 1970 had not been followed before terminating
service.69 In *Barbour*, some customers of a broker-dealer sought to
compel the Securities Investor Protection Corporation (SIPC) to exercise
its statutory authority for their benefit.70 In both cases, the Court
ultimately refused to imply a private right of action because plaintiffs’
suits might actually block achievement of the statutory goals.71 While
this problem would likely have caused the Court to deny relief under

68. *Natl R.R. Passenger Corp.*, 414 U.S. at 456. The Court stated:

[T]he threshold question clearly is whether the Amtrak Act or any other provision of law creates
a cause of action whereby a private party such as the respondent can enforce duties and
obligations imposed by the Act; for it is only if such a right of action exists that we need
consider [standing and jurisdiction].

Id. at 456; see also *Barbour*, 421 U.S. at 413–14 (“The question presented by this case is whether
such customers have an implied private right of action under the Securities Investor Protection Act . . . to compel [Securities Investor Protection Corporation (SIPC)] to exercise its statutory
authority for their benefit.”).

69. *Natl R.R. Passenger Corp.*, 414 U.S. at 455 n.3.
71. The legislative history of the Rail Passenger Service Act showed that Congress considered
and specifically rejected a provision that would have permitted suit to enforce the Act’s provisions
“by ‘any person adversely affected or aggrieved.’” *Natl R.R. Passenger Corp.*, 414 U.S. at 460–61
(quoting Supplemental Hearings on H.R. 17849 and S. 3706 before the Subcomm. on Transportation
and Aeronautics of the House Comm. on Interstate and Foreign Commerce, 91st Cong. 85 (1970)
(statement of Secretary of Transportation)). Moreover, Congress believed that “in order to achieve
economic viability in a basic rail passenger system,” it was necessary to pare uneconomic routes. *Id.*
at 461. If suits could be brought all over the country to block discontinuance, with injunctions
*pendente lite*, this could frustrate or at least severely delay passenger train discontinuance, with
possible dire economic consequences. *Id.* at 463–64. Similarly, in *Barbour* the Court thought that
“the overall structure and purpose of the SIPC scheme are incompatible with such an implied right.”
421 U.S. at 421. Intervention of SIPC in the affairs of a brokerage house was likely to put the house
out of business by driving away current customers and other brokers. *Id.* at 422–23. SIPC thus treats
an application for appointment of a receiver and liquidation of a firm as a last resort. *Id.* at 421. The
Court feared that customers could not be expected to consider the public interest in timing their
decision to apply to the courts. *Id.* at 422.
traditional standards, the singular focus on implication of a private right of action was new.

The Court reaffirmed the separate status of a cause of action in Cort v. Ash and set forth four criteria federal courts should use to decide whether to imply a private right of action from a federal statute. Plaintiff Ash, a stockholder in Bethlehem Steel Corporation, contended that the management of the company had authorized use of corporate funds for political advertisements in the 1972 presidential campaign in violation of a federal criminal statute. Ash sought injunctive relief against further expenditures and damages in favor of the corporation. “[T]he principal issue presented for decision,” the Court stated, “is whether a private cause of action for damages against corporate directors is to be implied in favor of a corporate stockholder under [the criminal statute].” The Court presented the following four criteria for deciding that question:

In determining whether a private remedy is implicit in a statute not expressly providing one, several factors are relevant. First, is the plaintiff “one of the class for whose especial benefit the statute was enacted,” Texas & Pacific R. Co. v. Rigsby—(emphasis supplied)—that is, does the statute create a federal right in favor of the plaintiff? Second, is there any indication of legislative intent, explicit or implicit, either to create such a remedy or to deny one? See, e.g., National Railroad Passenger Corp. v. National Assn. of Railroad Passengers ... (Amtrak). Third, is it consistent with the underlying purpose of the legislative scheme to imply such a remedy for the plaintiff? See, e.g., Amtrak, supra, Securities Investor Protection Corp. v. Barbour ... And finally is the cause of action one traditionally relegated to state law, in an area basically the concern of the States, so that it would be inappropriate to infer a cause of action based solely on federal law? ... cf. J. I. Case Co. v. Borak. ...

72. See supra notes 64–65 and accompanying text.
73. 422 U.S. 66 (1975).
74. Id. at 78.
75. Id. at 71–72.
76. Id.
77. Id. at 68.
Most commentators saw Cort not as effecting a major change in the law, but as an attempt to consolidate and harmonize precedent. But the significance of Cort depends on how one looks at precedent. If one accepts the Court’s implicit characterization of several hundred years of decisions as implied right of action cases, then Cort makes little change. But if one sees Cort as a major step in the separation of rights, rights of action, and remedies, then Cort makes a very large change indeed. Until Cort (and National Railroad Passenger Corporation and Barbour), the Court had never identified a cause of action as a separate and essential entity connecting a right and a remedy. Cort thus began a radical reconceptualization of the rights-remedies equation.

Davis v. Passman explicitly confirmed that the Court henceforth would consider rights, causes of action, and remedies as “analytically distinct.” Ms. Davis asserted that Congressman Otto Passman had discriminated against her on the basis of sex in violation of the Fifth Amendment. Congressman Passman countered that the Fifth Amendment...
ment did not provide a private right of action.\textsuperscript{84} The court of appeals agreed with Passman based on a \textit{Cort} analysis and affirmed dismissal of the suit.\textsuperscript{85} The U.S. Supreme Court reversed, holding that \textit{Cort}, with its heavy emphasis on legislative intent, did not provide the proper standard for deciding whether a cause of action should be implied from the Constitution.\textsuperscript{86} But the Court took pains to distinguish rights, causes of action, and remedies. After concluding that the plaintiff “assert[ed] a constitutionally protected right,”\textsuperscript{87} the Court went on to differentiate a cause of action and a remedy:

\begin{quote}
[\textit{C}ause of action is a question of whether a particular plaintiff is a member of the class of litigants that may, as a matter of law, appropriately invoke the power of the court; and \textit{r}elief is a question of the various remedies a federal court may make available. A plaintiff may have a cause of action even though he be entitled to no relief at all . . . . \textsuperscript{88}
\end{quote}

\textit{Davis} thus cemented the trifurcation of rights, rights of action, and remedies.\textsuperscript{89}

In the years following \textit{Davis}, the Court continued to treat rights, rights of action, and remedies as distinct, and to develop different criteria for each of the three parts of the new rights-remedies equation. The next section briefly reviews the Court’s developing doctrine for all three parts, beginning with implication of private rights of action.

\section{Rights of Action}

\textit{Cort v. Ash} itself soon came under attack by a Court increasingly hostile to judicial enforcement of federal statutory rights unless enforcement was explicitly authorized by Congress. In \textit{Touche Ross} &

\begin{footnotes}
\footnote{84. \textit{Id.} at 232.}
\footnote{85. \textit{Id.} at 232--33.}
\footnote{86. \textit{Id.} at 239--40.}
\footnote{87. \textit{Id.} at 234.}
\footnote{88. \textit{Id.} at 239 n.18. The Court also distinguished between the meaning of a cause of action in the rights context and in the pleading context. While a cause of action connects a right to a remedy, in pleadings a cause of action “refer[s] roughly to the alleged invasion of ‘recognized legal rights’ upon which a litigant bases his claim for relief.” \textit{Id.} at 237 (quoting Larson v. Domestic & Foreign Commerce Corp., 337 U.S. 682, 693 (1949)).}
\footnote{89. For good measure, the Court identified two other distinct questions: “[\textit{J}urisdiction is a question of whether a federal court has the power, under the Constitution or laws of the United States, to hear a case . . . ; standing is a question of whether a plaintiff is sufficiently adversary to a defendant to create an Art. III case or controversy . . . .” \textit{Id.} at 239 n.18.}
\end{footnotes}
Co. v. Redington Trustee\(^90\) and Transamerica Mortgage Advisors, Inc. v. Lewis,\(^91\) the Court abruptly elevated the second Cort criterion—whether Congress intended to create a private right of action—and downplayed the other three criteria. In Touche Ross, the Court refused to imply a private right of action under § 17(a) of the Securities Exchange Act of 1934 on behalf of brokerage firm customers against accountants who conducted a faulty audit of the firm’s records.\(^92\) The Court stated that “our task is limited solely to determining whether Congress intended to create the private right of action asserted” by the plaintiffs.\(^93\) The Court explicitly stated that the four Cort factors are not of equal weight and refused to consider plaintiffs’ argument that the third and fourth Cort factors favored implication, stating that “such inquiries have little relevance to the decision of this case.”\(^94\) Transamerica echoed Touche Ross: “[W]hat must ultimately be determined is whether Congress intended to create the private remedy asserted, as our recent decisions have made clear. . . . We accept this as the appropriate inquiry to be made in resolving the issues presented by the case before us.”\(^95\)

In the years following Touche Ross and Transamerica, the Justices bickered about whether Cort v. Ash retained any vitality. Cort became like a ghost attempting to regain solid form, appearing relatively substantial in one case only to fade to transparency in the next. Between April and June of 1981, the Court tried five times to define Cort’s status without reaching agreement, and while all of the Justices appeared to concede that congressional intent was now the key factor, the other Cort factors were given a large role in some cases and barely mentioned in others.\(^96\) In Daily Income Fund, Inc. v. Fox\(^97\) and Massachusetts Mutual

---

90. 442 U.S. 560 (1979).
92. See Touche Ross, 442 U.S. at 570.
93. Id. at 568 (emphasis added).
94. Id. at 575.
95. Transamerica, 444 U.S. at 15–16 (citing Touche Ross, 442 U.S. at 568).
96. In Universities Research Ass’n v. Coutu, 450 U.S. 754 (1981), the Court stated that to determine the ultimate question of whether Congress intended to create a private right of action, it would consider three factors set forth in Cort v. Ash, 422 U.S. 66 (1975), that it had traditionally relied upon in determining legislative intent: the language and focus of the statute, its legislative history, and its purpose. Univs. Research Ass’n, 450 U.S. at 770 (citing Touche Ross, 442 U.S. at 575–76). This language suggested that the Court would use the disfavored Cort factors to help discern the answer to the favored factor: whether Congress intended to create the private right of action. The Court’s language, however, is at best a very vague statement of the Cort factors. The fourth factor—whether the cause of action is one traditionally relegated to state law—is not mentioned at all. The Court subsequently indicated that it would not consider the fourth factor in this case. Id. at 770 n.21. In the Court’s second attempt to clarify Cort’s status, the Court restated the
Rights, Rights of Action, and Remedies

*Life Insurance Co. v. Russell,* both decided several terms later, the Court acknowledged that congressional intent is the main focus, but then proceeded to conduct a full-blown four-factor analysis. Justice Marshall made a Herculean attempt to harmonize all of these cases in *Thompson v.*
Thompson. 100 His language cannot be summarized or paraphrased, but must be quoted:

In determining whether to infer a private cause of action from a federal statute, our focal point is Congress’ intent in enacting the statute. As guides to discerning that intent, we have relied on the four factors set out in Cort v. Ash . . . along with other tools of statutory construction. See Daily Income Fund, Inc. v. Fox . . . California v. Sierra Club . . . Touche Ross & Co. v. Redington . . . Our focus on congressional intent does not mean that we require evidence that Members of Congress, in enacting the statute, actually had in mind the creation of a private cause of action. The implied cause of action doctrine would be a virtual dead letter were it limited to correcting drafting errors when Congress simply forgot to codify its evident intention to provide a cause of action. Rather, as an implied cause of action doctrine suggests, “the legislative history of a statute that does not expressly create or deny a private remedy will typically be equally silent or ambiguous on the question.” Cannon v. University of Chicago . . . We therefore have recognized that Congress’ “intent may appear implicitly in the language or structure of the statute, or in the circumstances of its enactment.” Transamerica Mortgage Advisors, Inc. v. Lewis . . . The intent of Congress remains the ultimate issue, however, and “unless this congressional intent can be inferred from the language of the statute, the statutory structure, or some other source, the essential predicate for implication of a private remedy simply does not exist.” Northwest Airlines, Inc. v. Transport Workers . . .


101. Id. at 179 (citing Daily Income Fund, 464 U.S. at 535–36; Northwest Airlines, 451 U.S. at 94; California v. Sierra Club, 451 U.S. 287, 293 (1981); Transamerica Mortgage Advisors, Inc. v. Lewis, 444 U.S. 11, 18 (1979); Touch Ross & Co. v. Redington, 442 U.S. 560, 575–76 (1979); Cannon v. Univ. of Chi., 441 U.S. 677, 694, (1979); Cort v. Ash, 422 U.S. 66, 78 (1975)). Justice Scalia erupted: “[T]he Court is not being faithful to current doctrine in its dicta denying the necessity of an actual congressional intent to create a private right of action, and in referring to Cort v. Ash . . . as though its analysis had not been effectively overruled by our later opinions.” Id. at 188 (Scalia, J., concurring). He continued: “I am at a loss to imagine what congressional intent to create a private right of action might mean, if it does not mean that Congress had in mind the creation of a private right of action.” Id. (Scalia, J., concurring). He also accused the Court of conveying “a misleading impression of current law when it proceeds to examine the ‘context’ of the legislation for indication of intent . . . .” Id. at 189 (Scalia, J., concurring). Justice O’Connor joined this part of Justice Scalia’s opinion, but not the final part where he suggested that “we should get out of the business of implied private rights of action altogether.” Id. at 192 (O’Connor, J., concurring).
Recently, Justice Kennedy acknowledged that “[t]he Court has encountered great difficulty in establishing standards for deciding when to imply a private cause of action under a federal statute which is silent on the subject.”102 While the standards are unclear, the results are not. By definition in these cases, the statute in question does not expressly create the cause of action the plaintiff wishes to employ. The legislative history is usually silent about whether Congress meant to create a private right of action. Thus, requiring clear evidence of congressional intent to create a private right of action ensures that few will be found.103 The Court rarely acknowledges the broader implications of its restrictive implication doctrine because it chooses to treat a cause of action as a separate procedural entity, apart from rights and remedies.104 But whether the Court acknowledges it or not, no right of action also means no right and no remedy.105

2. Rights

Whether a statute confers definable rights and duties is the logical starting point for any application of the rights-remedies principles. Traditionally, courts analyzed the issue in a facially intelligible way as a part of single, unified inquiry. Thus, the statute in Ashby v. White guaranteeing the right to vote plainly gave a right to vote to a qualified voter;106 the statute in Couch v. Steel requiring adequate medicines on board ship gave a right to seamen to have those medicines in place;107 the statute in Rigsby requiring secure handholds on trains gave train workers a right to have solid, not defective, handholds on the trains.108 The traditional approach was summed up in the RESTATEMENT OF TORTS: a statute conferred a right as long as it was intended for the benefit of the class of persons of which the plaintiff was a member, and the harm suffered was of a kind that the statute generally was intended to prevent.109

---

103. Thompson, 484 U.S. at 190 (Scalia, J., concurring) (pointing out that Court rejected claims of implied right of action in nine of eleven recent cases). For a listing of recent lower court decisions denying private rights of action, see Stabile, supra note 6, at 870 n.54.
104. See supra notes 66–89.
105. See infra notes 223–28 and accompanying text.
106. See supra notes 21–23 and accompanying text.
107. See supra note 23.
108. See supra notes 43–51 and accompanying text.
109. See supra note 36 and accompanying text.
Maine v. Thiboutot\footnote{448 U.S. 1 (1980).} precipitated the U.S. Supreme Court’s recent focus on rights as a separate part of the rights-remedies equation. Thiboutot held that 42 U.S.C. § 1983 encompasses claims against state and local officials for federal statutory violations as well as for federal constitutional violations.\footnote{See id. at 4.} This holding suggested that state or local officials could be sued in federal court for the violation of any federal law\footnote{See ERWIN CHEMERINSKY, FEDERAL JURISDICTION 527 (3d ed. 1999); Sunstein, supra note 6, at 394.} and raised the specter of an avalanche of new cases.\footnote{See Thiboutot, 448 U.S. at 22–23 (Powell, J., dissenting) (asserting that “[n]o one can predict the extent to which litigation arising from today’s decision will harass state and local officials; nor can one foresee the number of new filings” under literally hundreds of cooperative federal-state programs enacted by Congress).} Because § 1983 provided an explicit, all-purpose cause of action, to avoid the specter of a flood of new litigation, the Court had to find new ways to limit the right and remedy parts of the right/right of action/remedy triad.

The Court took the first step to limit rights in Pennhurst State School & Hospital v. Halderman.\footnote{451 U.S. 1 (1980). The Court also quickly moved to limit remedies in § 1983 cases. See infra notes 149–60 and accompanying text.} The plaintiff was a resident of Pennhurst, a Pennsylvania facility housing the mentally retarded.\footnote{Pennhurst, 451 U.S. at 5.} She brought a class action alleging that the “unsanitary, inhumane, and dangerous” conditions at Pennhurst violated, inter alia, § 6010 of the Developmentally Disabled Assistance and Bill of Rights Act.\footnote{Id. at 6.} She sought injunctive and monetary relief and asked that Pennhurst be closed and that “community living arrangements” be made for the residents.\footnote{Id. at 11, 31. Ultimately the Court remanded the case for consideration of other claims raised by the plaintiffs. Id. at 30–31.} The Court concluded that the § 6010 conferred no rights on the plaintiffs and thus denied relief.\footnote{Id. at 13.}

On its face, § 6010 appears to confer rights on persons like the plaintiffs in Pennhurst. Section 6010 is entitled a “bill of rights” provision, and states that “Congress makes the following findings,” including that “[p]ersons with developmental disabilities have a right to appropriate treatment, services, and habilitation” which “should be provided in the setting that is least restrictive of the person’s personal liberty.”\footnote{Id. at 13.} The Court nullified this seemingly straightforward language

110. 448 U.S. 1 (1980).
111. See id. at 4.
112. See ERWIN CHEMERINSKY, FEDERAL JURISDICTION 527 (3d ed. 1999); Sunstein, supra note 6, at 394.
113. See Thiboutot, 448 U.S. at 22–23 (Powell, J., dissenting) (asserting that “[n]o one can predict the extent to which litigation arising from today’s decision will harass state and local officials; nor can one foresee the number of new filings” under literally hundreds of cooperative federal-state programs enacted by Congress).
114. 451 U.S. 1 (1980). The Court also quickly moved to limit remedies in § 1983 cases. See infra notes 149–60 and accompanying text.
115. Pennhurst, 451 U.S. at 5.
116. Id. at 6.
117. Id.
118. Id. at 11, 31. Ultimately the Court remanded the case for consideration of other claims raised by the plaintiffs. Id. at 30–31.
119. Id. at 13.
Rights, Rights of Action, and Remedies

in several steps. The Court identified the issue as whether Congress intended § 6010 to create enforceable rights\textsuperscript{120} and then set forth some general guidelines to discern intent. First, the Court said that the case for inferring intent “is at its weakest where, as here, the rights asserted impose \textit{affirmative} obligations on the States to fund certain services, since we may assume that Congress will not implicitly attempt to impose massive financial obligations on the States.”\textsuperscript{121} Second, the Court established a “clear statement” requirement for statutes, like the Act in \textit{Pennhurst}, enacted pursuant to Congress’s spending power.\textsuperscript{122} While Congress can fix the terms on which it will disburse federal money to the states, such legislation is like a contract.\textsuperscript{123} For a state to voluntarily and knowingly accept the terms of the contract, the terms must be very clear.\textsuperscript{124} Thus, “if Congress intends to impose a condition on the grant of federal moneys, it must do so unambiguously.”\textsuperscript{125}

With these two principles in place, the Court turned to § 6010 and stated that it found nothing in the Act or in the legislative history to suggest that Congress meant to impose on the States the cost of providing appropriate treatment in the least restrictive setting for mentally retarded citizens.\textsuperscript{126} Because § 6010 only made a “finding” that persons with developmental disabilities have a right to treatment and habilitation in a least restrictive setting, it did not actually create the rights and obligations that plaintiffs claimed.\textsuperscript{127} After reviewing the legislative history, the Court concluded that the provisions of § 6010 “were intended to be hortatory, not mandatory.”\textsuperscript{128} As for the clear statement rule, the Court held that the phrases “appropriate treatment” and “least restrictive setting” were ambiguous, and thus did not provide sufficiently clear notice to the states of their obligations in exchange for the federal money.\textsuperscript{129}

The Court refined the criteria for deciding whether a statute creates a right in subsequent cases. In \textit{Wright v. City of Roanoke Redevelopment 

\begin{itemize}
  \item \textsuperscript{120} \textit{See id. at} 15.
  \item \textsuperscript{121} \textit{Id. at} 16–17.
  \item \textsuperscript{122} \textit{See id. at} 17. The spending power is contained in article I, section 8, clause 1 of the Constitution, which reads “Congress shall have Power to . . . provide for the . . . general Welfare of the United States.” U.S. CONST. art. I, § 8, cl. 1; \textit{Suter v. Artist M.}, 503 U.S. 347, 356 n.7 (1992).
  \item \textsuperscript{123} \textit{Pennhurst}, 451 U.S. at 17.
  \item \textsuperscript{124} \textit{Id.}
  \item \textsuperscript{125} \textit{Id.}
  \item \textsuperscript{126} \textit{Id. at} 18.
  \item \textsuperscript{127} \textit{Id. at} 18–19.
  \item \textsuperscript{128} \textit{Id. at} 24.
  \item \textsuperscript{129} \textit{See id. at} 24–25.
\end{itemize}
Housing Authority, plaintiffs claimed that the Housing Authority overbilled them for utilities in violation of the Brooke Amendment to the United States Housing Act of 1937. The Amendment limited rent paid by low-income families, including utilities, to thirty percent of their income. The Court held that the Brooke Amendment gave the tenants rights because the thirty percent limit was a mandatory provision intended to benefit the tenants. Moreover, the Housing and Urban Development (HUD) regulations requiring that a “reasonable” amount for utilities be included in the rent payment were not “too vague and amorphous” or “beyond the competence of the judiciary to enforce.” This reasoning suggested a three-part test that the Court subsequently articulated in Golden State Transit Corp. v. City of Los Angeles:

In deciding whether a federal right has been violated, we have considered whether the provision in question creates obligations binding on the governmental unit or rather “does no more than express a congressional preference for certain kinds of treatment.” Pennhurst . . . The interest the plaintiff asserts must not be “too vague and amorphous” to be “beyond the competence of the judiciary to enforce.” Wright . . . We have also asked whether the provision in question was “intend[ed] to benefit” the putative plaintiff. Id.

The Court has continued to apply the three-part Golden State test in recent years, sometimes finding that a statute creates an enforceable right, and sometimes not. The Court also has continued to treat the

131. Id. at 419.
132. Id. at 420–22.
133. Id. at 429–30.
134. Id. at 431–32.
136. Id. at 106 (quoting Wright, 479 U.S. at 431–32; Pennhurst State Sch. & Hosp. v. Halderman, 451 U.S. 1, 19 (1981)). Golden State arose from a labor dispute between the transit corporation and its union. The City of Los Angeles conditioned renewal of the transit corporation’s taxi franchise on settlement of the dispute. Id. at 104. In an earlier ruling, the Court held that the City had violated federal labor law by taking this action. See Golden State Transit Corp. v. City of L.A., 475 U.S. 608, 618–19 (1986). On remand the district court ordered the city to reinstate the franchise but declined to award damages under § 1983; the court of appeals affirmed. Id. at 104–05. Applying the criteria set forth in the quotation above, the U.S. Supreme Court held that the corporation was “the intended beneficiary of a statutory scheme that prevents governmental interference with the collective-bargaining process.” Id. The Court also held “that the NLRA [National Labor Relations Act] gives it rights enforceable against governmental interference in an action under § 1983,” id. at 109, and remanded the case for further proceedings, id. at 113.
question of whether a statute confers a right as analytically distinct from the question of whether a private right of action should be implied. 139

3. Remedies

After separating rights, rights of action, and remedies, the Court developed separate, relatively restrictive criteria for deciding whether a statute creates a right and whether to imply a right of action from a statute. The record regarding remedies is more mixed. While recent U.S. Supreme Court decisions articulate many reasons for limiting or denying remedies, the Court has on occasion reaffirmed its broad remedial powers. The Court may be hesitant to restrict its power to order an effective remedy when it believes circumstances warrant full relief.

Traditionally, courts were willing to make available all appropriate remedies to enforce rights. 140 Most modern cases, by contrast, are ringed about with reservations. 141 The Court developed several limitations on

---

138. See, e.g., Blessing v. Freestone, 520 U.S. 329, 341–42 (1997) (holding that Title IV-D of Social Security Act, which required states to take adequate steps to obtain child support payments from fathers of children, did not confer enforceable rights on custodial parents); Suter v. Artist M., 503 U.S. 347, 363 (1992) (holding that provision of Adoption Assistance and Child Welfare Act of 1980, requiring states to have plan providing that reasonable efforts will be made to keep children out of foster care, did not confer any rights on foster children).

139. The Court occasionally notes the interplay of the two questions, while treating them as separate inquiries. For example, in Suter, the district court and court of appeals both found that the statutory provisions in question could be enforced under § 1983 and by means of an implied right of action. Suter, 503 U.S. at 353–54. The U.S. Supreme Court first concluded that the statutory provisions did not create an enforceable right, and then separately and summarily concluded that if Congress had not meant to create an enforceable right, there was nothing to create a cause of action to enforce. Id. at 363–64. Wilder v. Virginia Hospital Ass’n, 496 U.S. 498 (1990), sparked a disagreement among the Justices about the relationship between rights and rights of action. The majority considered them to be separate, id. at 508 n.9, while the dissent sought to equate the question “Does the statute create a right?” with the first Cort criterion, id. at 525 (Rehnquist, C.J., dissenting). Wilder is discussed at infra notes 241–57 and accompanying text.

140. See supra Part I.A.

141. The Court signaled that it would be somewhat more restrictive in granting equitable remedies in Weinberger v. Romero-Barcelo, 456 U.S. 305 (1982). The restrictions were not all new; some merely restated traditional limits or preconditions. Id. at 311–13. But by compiling lists of limitations, and adding some new limits for good measure, the Court implicitly suggested that it would follow a more cautious, restrictive approach to equitable remedies. Id. In Weinberger, residents of Puerto Rico sued to enjoin the Navy from using Vieques, a small island off Puerto Rico’s coast, for target practice. See id. at 307. The plaintiffs claimed that discharging ordnance into the water around the island without first obtaining a permit from the Environmental Protection Agency violated several federal laws, including the Federal Water Pollution Control Act. Id. at 307–08. The district court refused to issue the injunction, although it did order the Navy to apply for a
remedies in cases where plaintiffs sought damages for violation of constitutional rather than statutory rights. For example, *Davis v. Passman*, 142 the first case stating that rights, rights of action, and remedies are “analytically distinct,” 143 also suggested that damages might be inappropriate for a violation of a constitutional right if: a) there were special factors counseling hesitation, b) the case presented unfocused remedial issues posing difficult questions of valuation or causation, or c) allowing damages would result in a flood of new cases. 144

The Court found special factors counseling hesitation in two cases brought by enlisted military personnel claiming constitutional violations by superior officers. 145 The Court refused a damage remedy in both cases because it did not want to interfere with military discipline or Congress’s permit. Id. at 309–10. The court of appeals held that the statute required an immediate injunction halting the bombing until a permit was obtained. Id. at 310–11. The U.S. Supreme Court reversed and remanded the case for the court of appeals to consider whether the district court abused its discretion in failing to order an injunction. Id. at 306, 320. Initially, the Court appeared to construe federal equitable powers broadly by stressing the flexibility of equity and need to mold decrees to fit the particular needs of each case. Id. at 312. However, the Court then listed reasons why injunctions should be denied. An injunction is not a remedy “which issues as of course” or to deal with trifling matters. Id. at 311 (quoting Harrisonville v. W.S. Dickey Clay Mfg. Co., 289 U.S. 334, 337–38 (1933)). An injunction should be denied unless the plaintiff can show irreparable injury and the inadequacy of legal remedies. Id. at 312. Finally, “courts of equity should pay particular regard for the public consequences” of ordering an injunction, and deny relief if it might adversely affect the public interest. Id. The Court appeared to believe that these principles supported denial of an injunction in this case. See id. The Navy’s technical violation of the federal statute was not causing any appreciable environmental harm. Id. at 307. Thus, the harm might be characterizing as “trifling” rather than “irreparable.” Moreover, the national interest might be harmed if the Navy could not use the island to train its gunners. Id. at 310. Ultimately, the Court reversed and remanded the case to the court of appeals to determine whether the district court abused its discretion when it declined to issue an injunction. Id. at 320.


---

142. 442 U.S. 228 (1979).
143. Id. at 239; see also supra notes 81–89 and accompanying text.
144. Id. at 245–48. The Court allowed a damage remedy in *Davis* because none of these conditions were met. Id. at 248. Although a suit against a Congressman claiming that he acted unconstitutionally in the course of his duties raised “special concerns counseling hesitation,” the Court held the concerns were coextensive with the protections afforded by the Speech and Debate Clause of the Constitution. Id. at 246. The Court also believed that “damages would be judicially manageable,” id. at 245, and did not see the potential for a deluge of new lawsuits, id. at 248.


plenary authority over military matters. The Court also found special factors counseling hesitation in two other cases claiming denial of constitutional rights where plaintiffs had obtained substantial relief through administrative remedies, even though the relief was incomplete. In those cases, the Court concluded that the elaborate and carefully thought-out remedial schemes set up by Congress should not be augmented with new remedies created by the judiciary.

The Court used similar reasoning to limit remedies in § 1983 actions against state and local officials for violation of federal statutes, holding that comprehensive statutory remedies raise a presumption that Congress intended to preclude use of § 1983. In Middlesex County Sewerage Authority v. National Sea Clammers Ass’n, for example, plaintiffs sued local, state, and federal officials responsible for dumping sewage and other waste into the Atlantic Ocean, claiming that the pollution was causing the “collapse of the fishing, clamming and lobster industries.” Plaintiffs claimed, inter alia, that defendants were violating the Federal Water Pollution Control Act and the Marine Protection, Research, and Sanctuaries Act of 1972 and sought declaratory and injunctive relief and damages. The Court first declined to imply a private right of action under the two statutes and then raised on its own the question whether

146. See Stanley, 483 U.S. at 681; Chappell, 462 U.S. at 304.
147. In Bush v. Lucas, 462 U.S. 367 (1983), plaintiff was returned to his job and got back pay for an improper retaliatory demotion, but was unable to obtain attorney’s fees or damages for emotional distress and mental anguish. Id. at 371–72, 388. In Schweiker v. Chilicky, 487 U.S. 412 (1988), plaintiffs were eventually restored to disabled status and were awarded full retroactive benefits after being wrongly terminated from a Social Security program, but were unable to obtain damages for emotional distress and other hardships suffered because of the delays in the receipt of their benefits. See id. at 425, 428–29.
148. Schweiker, 487 U.S. at 428–29; Bush, 462 U.S. at 389. Another factor may have made the Court particularly reluctant to grant a damage remedy in Schweiker. The lawsuit could be seen as an attempt by the plaintiffs to make an end run around the restrictive standards prevailing in the 1980s for implication of private rights of action from statutes. Schweiker, 487 U.S. at 418–19. Plaintiffs alleged that several federal officials and one state official who were “acting under color of federal law” took actions violating the Social Security Act, and that these actions denied them due process of law. Id. It was unlikely that a private right of action on their behalf would be implied under the Social Security Act; thus, the request for a remedy directly under the Fifth Amendment provided their only hope for relief (§ 1983 could not be used because the suit was against federal officials). Lawsuits against federal agencies often assert a failure to follow mandated procedures. If all or even many of these actions could be converted into constitutional actions under the Fifth Amendment and proceed on that basis, then the Touche Ross-Transamerica restrictions could be circumvented.
150. Id. at 4–5.
151. Id.
152. Id. at 11–18. The Court used the Touche Ross-Transamerica standards, see supra notes 90–95 and accompanying text, that focus on whether Congress intended to create a private right of
§ 1983 might provide plaintiffs with an alternate source of congressional authorization to sue.\textsuperscript{153} It quickly answered the question in the negative, stating that “[w]hen the remedial devices provided in a particular Act are sufficiently comprehensive, they may suffice to demonstrate congressional intent to preclude the remedy of suits under § 1983.”\textsuperscript{154} The Court found it “hard to believe that Congress intended to preserve the § 1983 right of action when it created so many specific statutory remedies, including the two citizen-suit provisions.”\textsuperscript{155} Consequently, it held that § 1983 was unavailable.\textsuperscript{156}

The Court reaffirmed this new limitation on judicial remedies in subsequent § 1983 cases. In Smith v. Robinson\textsuperscript{157} the Court held that Congress intended the comprehensive procedures and guarantees authorized in the Education of the Handicapped Act to be the exclusive means by which a handicapped child could raise an equal protection claim and that § 1983 could not be used.\textsuperscript{158} On the other hand, Wright v. City of Roanoke Redevelopment & Housing Authority\textsuperscript{159} made it somewhat easier to enforce federal statutes through § 1983 by placing the burden on the defendant state or local officials to demonstrate that Congress intended to foreclose § 1983 actions, and by holding that creation of general administrative remedies or fund cut-off procedures would not ordinarily suffice to show congressional intent to preclude use of § 1983.\textsuperscript{160}

\textit{Sea Clammers}, 453 U.S. at 14. Since the Acts contained elaborate enforcement mechanisms that conferred authority to sue on both government officials and private citizens, the Court concluded that Congress did not intend “to authorize by implication additional judicial remedies for private citizens.”\textit{Id.}

\textsuperscript{153} Id. at 19.

\textsuperscript{154} Id. at 20.

\textsuperscript{155} Id.

\textsuperscript{156} Id. at 21.

\textsuperscript{157} 468 U.S. 992 (1984).

\textsuperscript{158} Id. at 1011–12.

\textsuperscript{159} 479 U.S. 418 (1987).

\textsuperscript{160} Id. at 427–28. The Court noted that in both \textit{Sea Clammers} and Smith v. Robinson the statutes involved provided for some private judicial enforcement, thus evidencing an intent to supplant § 1983. Id. at 427. The statute at issue in Wright, by contrast, had no such provisions. Id.; see also Wilder v. Va. Hosp. Ass’n, 496 U.S. 498, 520–23 (1990) (holding that defendants had not carried their burden of showing that Medicaid Act contained comprehensive remedial scheme because Act contained no provisions for private judicial or administrative enforcement and fund cut-off mechanism was not sufficient to foreclose reliance on § 1983).

The Court has continued to reaffirm the “comprehensive remedial scheme” exception to the use of § 1983 in recent years. See Blessing v. Freestone, 520 U.S. 329, 341 (1997) (stating that Congress can specifically foreclose § 1983 remedy “expressly, by forbidding recourse to § 1983 in the statute itself, or impliedly, by creating a comprehensive enforcement scheme that is incompatible with individual enforcement under § 1983.”); Livadas v. Bradshaw, 512 U.S. 107, 133 (1994) (same).
Given the relatively restrictive remedies decisions in the 1970s and 1980s, it was a surprise in 1992 when the Court appeared to reaffirm the traditional generous remedial principles. *Franklin v. Gwinnett County Public Schools* accomplished this result by sharply distinguishing between rights, rights of action, and remedies, and by holding that the restrictive standards for creating rights and implying rights of action do not apply to remedies. In *Franklin*, plaintiff alleged that a high school teacher subjected her to continual sexual harassment and "coercive intercourse." She alleged that other teachers and administrators knew about this conduct and about similar incidents with other students but did nothing. The offending teacher finally resigned on the condition that the matter be dropped, and the school closed its investigation. Ms. Franklin brought suit under Title IX of the Education Amendments of 1972. The district court dismissed the complaint on the ground that Title IX does not authorize a damage award, and the court of appeals affirmed.

The U.S. Supreme Court began by reaffirming the decision in *Cannon v. University of Chicago*, which held that Title IX was enforceable through a private right of action. The Court then carefully distinguished the issue presented by *Franklin*, stating that "in this case we must decide what remedies are available in a suit brought pursuant to this implied right." The Court reiterated that "the question of what remedies are available under a statute that provides a private right of action is analytically distinct from the question of whether a private right of action exists in the first place." Thus, the implied right of action cases cited by the defendants with their emphasis on whether Congress

162.  Id. at 65–66.
163.  Id. at 66.
164.  Id. at 63.
165.  Id. at 64.
166.  Id.
167.  Id. The pertinent portion of Title IX reads: "No person in the United States shall, on the basis of sex, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any education program or activity receiving Federal financial assistance . . . ." 20 U.S.C. § 1681(a) (1994).
168.  See *Franklin*, 503 U.S. at 64.
171.  Id.
172.  Id. at 65–66 (citing *Davis v. Passman*, 442 U.S. 228, 239 (1979)). *Davis* is discussed supra notes 81–89 and accompanying text.
intended to create a private right of action were simply “inapposite.” The Court cited many of the venerable cases establishing a reciprocal relation between rights and remedies and stated that once a right and a cause of action are in place, “we presume the availability of all appropriate remedies unless Congress has expressly indicated otherwise.”

The Court found nothing in the legislative history of Title IX, which was enacted when the traditional norms were firmly in place, to indicate that Congress would not want a damage remedy for violation of Title IX. Moreover, any limitation that might flow from the fact that Title IX was enacted pursuant to the Spending Clause did not apply because the plaintiff alleged an intentional violation of the statute. Finally, the Court rejected the suggestion that relief under Title IX should be limited to back pay and prospective relief. Ms. Franklin would take nothing from either remedy. Ms. Franklin was a student, not a teacher; the offending teacher no longer taught at the school, and she was no longer a student there.

The Court recently retreated from its liberal statements of Franklin in Gebser v. Lago Vista Independent School District and Davis v. Monroe County Board of Education. Both cases limit the scope of the damage remedy available against a school district under Title IX and reaffirm that the Court considers rights, rights of action, and remedies to be distinct. Ms. Gebser had a sexual relationship with a teacher at her school but did not report the relationship to school officials. Subsequently, the parents of two other girls complained to the principal about comments the teacher made in class, and the teacher was cautioned to be careful in the future. A few months later a police officer

173.  Franklin, 503 U.S. at 69 n.6.  
175.  Franklin, 503 U.S. at 66. The Court asserted that “[t]his principle has deep roots in our jurisprudence.” Id.  
176.  Id. at 71–73.  
177.  Id. at 74–75. The Court believed that the school district plainly was on notice that Title IX prohibits discrimination on the basis of sex and that countenancing the teacher’s conduct amounted to a knowing violation of Title IX. Id.  
178.  Id. at 75–76.  
179.  Id. at 76.  
183.  Id. at 278.
discovered the plaintiff and the teacher having intercourse and arrested the teacher, who was subsequently fired. The case was removed to federal court and the Title IX claim was dismissed on the ground that the school district did not have sufficient notice of the teacher’s conduct. The court of appeals affirmed.

The U.S. Supreme Court acknowledged that past cases had established each of the three separate elements of the rights-remedies equation in Title IX cases: Students have a right to be free from sexual harassment, an implied private right of action exists, and damages are available in the implied private action. The Court made clear that this case concerned only the scope of the remedy: “We made no effort in Franklin to delimit the circumstances in which a damages remedy should lie.” Because the cause of action was judicially created, the Court believed that it had “a measure of latitude to shape a sensible remedial scheme that best comports with the statute.” The Court then restricted the circumstances in which damages could be recovered. The Court stated that statutory structure and purpose were “pertinent not only to the scope of the implied right [of action], but also to the scope of the available remedies.” While the general rule is that federal courts can use all available remedies to vindicate federal rights, the rule must yield “where necessary to carry out the intent of Congress or to avoid frustrating the purposes of the statute involved.”

The plaintiffs argued that the school district should be held to a respondeat superior standard of liability, or at least to a constructive-notice standard, making it liable for damages if officials knew or should have known about harassment.

184. Id.
185. Id. at 278–79.
186. Id. at 279.
187. Id.
188. Id. at 281–82. The Court noted that Meritor Savings Bank, FSB v. Vinson, 477 U.S. 57 (1986), held that a supervisor discriminates against an employee on the basis of sex when the supervisor sexually harasses the employee and that Franklin applied the same rule when a teacher sexually harasses a student. Gebser, 524 U.S. at 281–82.
189. Id. at 281 (citing Cannon v. Univ. of Chi., 441 U.S. 677 (1979)).
190. Id. (citing Franklin v. Gwinnett County Pub. Sch., 503 U.S. 60 (1992)).
191. Id. at 284.
192. Id.
193. Id.
194. Id. at 284–85.
195. Id. at 285 (quoting Guardians Ass’n v. Civil Serv. Comm’n, 463 U.S. 582, 595 (1983) (White, J.)).
but failed to address it. The Court thought that either of these standards would frustrate the purposes of Title IX. Because Title IX was enacted under the Spending Clause, the Court reasoned that the plaintiffs’ proposed standard would improperly result in liability without notice to an “appropriate person” and an opportunity to rectify any violation. To avoid this problem, the Court adopted the following standard of liability:

We conclude that damages may not be recovered [for the sexual harassment of a student by one of the district’s teachers] unless an official of the school district who at a minimum has authority to institute corrective measures on the district’s behalf has actual notice of, and is deliberately indifferent to, the teacher’s misconduct.

The Court’s analysis in *Davis v. Monroe County Board of Education* is similar to its analysis in *Gebser*. The issue was whether a school district can be liable for damages under Title IX for the sexual harassment of one student by another. The Court again separated the right, right of action, and remedy, stating that student-on-student sexual harassment can rise to the level of discrimination under Title IX, that a private right of action exists, and that money damages are available.

As in *Gebser*, the Court purported to consider only the scope of the remedy. The Court again relied on the Spending Power analysis to restrict the damage remedy. Private damage actions are available “only where recipients of federal funding [have] adequate notice that they could be liable for the conduct in issue.” The Court stated that the school district could only be liable for its own misconduct. The school district could only be said to “expose” its students to harassment or “cause” them to undergo it if two conditions were met. First, the school district must exercise “substantial control over both the harasser and the context in which the . . . harassment occurs,” and second, the school district must have actual knowledge of the harassment and be

---

196. *Id.* at 282–83.
197. *Id.* at 287–88.
198. *Id.* at 277, 290–91.
200. *Id.* at 632–33.
201. *Id.* at 639–40.
202. *Id.* at 639 (“[A]t issue here is the question whether a recipient of federal education funding may be liable for damages under Title IX under any circumstances for discrimination in the form of student-on-student sexual harassment.”).
203. *Id.* at 640.
204. *Id.*
205. *Id.* at 645.
Rights, Rights of Action, and Remedies

deliberately indifferent to it.\textsuperscript{206} The Court tacked on one further limitation for good measure. To recover damages, a victim of student-on-student sexual harassment must establish “harassment that is so severe, pervasive, and objectively offensive, that it effectively bars the victim’s access to an educational opportunity or benefit.”\textsuperscript{207}

Part I has shown that traditionally courts equated rights and remedies without considering whether the statute in question provided a private right of action. Recently, the Court divided what had previously been a unified inquiry into three separate questions and developed separate, relatively restrictive criteria to decide whether a statute should be read to create a right, imply a right of action, or provide a remedy. Part II of this Article explains why rights, rights of action, and remedies are interrelated\textsuperscript{208} and makes the case for development of a single, integrated

\textsuperscript{206} Id. at 646–47, 650.
\textsuperscript{207} Id. at 633.
\textsuperscript{208} Interestingly, there is a group of three cases in which the modern Court recognized the interrelation of rights, rights of action, and remedies. The issue in all three cases was whether a right of contribution existed on behalf of someone found liable under a particular federal statute. The Court seemed uneasy talking about all three elements at the same time, and the discussions are somewhat jumbled.

In \textit{Northwest Airlines, Inc. v. Transport Workers Union}, 451 U.S. 77 (1981), the airline sued for contribution from two unions that it claimed were partly to blame for a policy of differential pay for male and female cabin attendants that resulted in a substantial money judgment against the airline in an earlier action. \textit{Id.} at 79–80. The airline claimed that it had an implied cause of action under the Equal Pay Act, or alternately, a federal common law right to contribution. \textit{Id.} at 82. The Court assumed arguendo that all of the elements of a typical contribution claim were established. \textit{Id.} at 88–90. It then made the following statement that appears to recognize the interrelation between the three elements, although the Court did not elaborate:

None of these assumptions, however, provides a sufficient basis for recognizing the right petitioner is asserting in this proceeding.

That right may have been created in either of two ways. First, it may have been created by statute when Congress enacted the Equal Pay Act or Title VII. Even though Congress did not expressly create a contribution remedy, if its intent to do so may fairly be inferred from either or both statutes, an implied cause of action for contribution could be recognized on the basis of the analysis used in cases such as \textit{Cort v. Ash}, [422 U.S. 68 (1975)]. . . . Second, a cause of action for contribution may have become a part of the federal common law through the exercise of judicial power to fashion appropriate remedies for unlawful conduct. \textit{Id.} at 90. The Court went on to deny the right of contribution. \textit{Id.} at 98–99.

The Court in \textit{Texas Industries, Inc. v. Radcliff Materials, Inc.}, 451 U.S. 630 (1981), considered whether a defendant found liable in an antitrust case could seek contribution from other participants in the unlawful conspiracy. \textit{Id.} at 632. It is hard to tell from the Court’s language whether it is equating the right and the right of action, or simply mixing them up. The Court stated:

There is no allegation that the antitrust laws expressly establish a right of action for contribution. Nothing in these statutes refers to contribution, and if such a right exists it must be by implication. Our focus, as it is in any case involving the implication of a right of action, is on the intent of Congress.
standard for deciding whether a statutory provision should be judicially enforceable.

Id. at 639. The Court concluded Congress did not intend to authorize a right of contribution. Id. at 639–40.

Finally, in *Musick, Peeler & Garrett v. Employers Insurance of Wausau*, 508 U.S. 286 (1993), the Court considered whether defendants in a suit based on the implied right of action under § 10(b) of the Securities Exchange Act of 1934 and Rule 10b-5 of the Securities and Exchange Commission may seek contribution from joint tortfeasors. Id. at 288. Justice Kennedy's opinion for the Court does not present an integrated analysis of rights, rights of action, and remedies. Instead, it hopelessly jumbles them. The problems begin with the question on which certiorari was granted: “‘Whether federal courts may imply a private right to contribution in Section 10(b) . . . and Rule 10b-5 . . . .’” Id. at 289–90 (quoting petition for writ of certiorari). It is unclear from this statement whether the Court was deciding if it should imply a right or imply a right of action. Although rights, rights of action, and remedies are inextricably related, they are not the same thing. Justice Kennedy made several statements that treated the three parts of the equation as synonymous:

The private right of action under Rule 10b-5 was implied by the Judiciary on the theory courts should recognize private remedies to supplement federal statutory duties, not on the theory Congress had given unequivocal direction to the courts to do so. . . . Thus, it would be futile to ask whether the 1934 Congress also displayed a clear intent to create a contribution right collateral to the remedy. . . .

[Touche Ross, Transamerica, and their progeny] caution against the creation of new causes of action. . . . They teach that the creation of new rights ought to be left to legislatures, not courts. . . . Whether the right of a tortfeasor to seek contribution from those who share, or ought to share, joint liability is recognized by statute . . . or as a matter of common law, in both instances the right is thought to be a separate or independent cause of action. Id. at 291–92. The Court ultimately granted a right of contribution in *Musick*. Id. at 298.

It is intriguing to ponder why the Court talks of rights, rights of action, and remedies together in *Northwest Airlines, Texas Industries*, and *Musick*, while it considers them separately in so many other cases. There are two possible reasons. First, in the contribution cases all three parts of the rights-remedies equation were at issue. The cases are unlike those in the *Thiboutot* line, where § 1983 created a cause of action and the focus was on the “rights” part of the equation, or the recent Title IX cases, where the Court took the right and right of action as established, and focused only on the remedy. Thus, the procedural setting in the contribution cases invited a discussion of all three elements. Second, in the contribution cases, the right is defined in terms of the remedy. The right is a right to a remedy: contribution. This makes it almost impossible to talk about the right and the remedy as separate entities. In most cases the right is not a right to a remedy, but rather a right to have another act or forbear from acting in a particular manner—for example, to tell the truth in a proxy statement, or avoid discriminating on the basis of race—and this facilitates separate discussions of rights and remedies.

Although the discussions in *Northwest Airlines, Texas Industries*, and *Musick* are somewhat confused, on the positive side they suggest that the Court might be moving, however haltingly, in the right direction.
II. TOWARD A NEW TEST FOR DECIDING WHETHER A STATUTORY PROVISION SHOULD BE JUDICIALLY ENFORCEABLE

Rights, rights of action, and remedies are inextricably related. A court cannot make a decision about one of them without necessarily affecting the other two. A decision that focuses just on one element without acknowledging the impact on the other two is incomplete and may distort the meaning of the statute in issue and undermine congressional intent. Thus, the Court has been wrong to consider separately whether a statute confers a right, implies a right of action, or authorizes a remedy. All three questions are really part of a single, broader inquiry: Assuming the plaintiff can prove the case, does the applicable statutory provision entitle the plaintiff to the particular judicial remedy he or she seeks? One inquiry should have one test for answering it.

The further questions “what should the test be?” and “how should it be applied?” cannot be answered simply or mechanically. How one answers these questions depends on one’s views on such matters as the role of the federal courts in the federal system, separation of powers, the Erie doctrine, judicial activism, and on several pragmatic concerns. Part II of this Article groups together positions on these issues to form two general models governing the judicial enforcement of federal statutes. It denotes one model the adversarial model and the other the cooperative model. This sets the stage for Part III of this Article, which seeks to establish a middle ground between the two models.

A. The Case for a Single, Integrated Standard

1. The Interrelation of Rights, Rights of Action, and Remedies

Rights, rights of action, and remedies are interrelated conceptually and practically. The interrelation flows first from the very definitions of the terms. A legal right imposes a correlative duty on another to act or to refrain from acting for the benefit of the person holding the right. 209 From this definition it follows that

[A] right without a remedy is not a legal right; it is merely a hope or a wish. . . . Unless a duty can be enforced, it is not really a duty; it is only a voluntary obligation that a person can fulfill or not at his whim. In such circumstances, the holder of the correlative “right”

209. See Hohfeld, supra note 2, at 35–38.
can only hope that the act or forbearance will occur. Thus, a right without a remedy is simply not a legal right. 210

Franklin v. Gwinnett County Public Schools 211 illustrates this point. If the plaintiff could not obtain damages from the school district for sexual harassment by a teacher, then the school district had no legal duty to protect her. It could protect her or not as it wished. Thus, without a remedy, plaintiff had no legal right as against the school district to be free of sexual harassment by teachers.

The kind of remedy that one may obtain also defines the scope and meaning of the right. Bivens v. Six Unknown Named Agents of the Federal Bureau of Narcotics 212 provides an example. The Fourth Amendment grants people a right “to be secure in their persons [and] houses . . . against unreasonable searches and seizures.” 213 Bivens alleged that federal agents violated this right by entering his apartment without a warrant or probable cause, terrorizing his family, searching the entire premises, and using unreasonable force to arrest him. 214 It is unclear from the Court’s opinion whether the agents found the narcotics they were searching for, but assume that they did. If Bivens’ only remedy was to raise the Fourth Amendment violation in defense of his criminal prosecution, then his Fourth Amendment right meant that Bivens could successfully defend a criminal prosecution. This is obviously important; however, because the Court decided that damages could be obtained for a

210. Zeigler, supra note 6, at 678 (footnotes omitted). Many jurisprudential scholars accept this conclusion. See, e.g., THEODORE M. BENDITT, RIGHTS 51 (1982) (“[O]ne can’t say that he has a right, then and there, to do something if he is not permitted, then and there, to do it. Another way of putting this is to say that a right doesn’t exist if one can’t act on it (or insist on it).”). Joel Feinberg states:

Why is the right to demand recognition of one’s rights so important? The reason, I think, is that if one begged, pleaded, or prayed for recognition merely, at best one would receive a kind of beneficent treatment easily confused with the acknowledgment of rights, but in fact altogether foreign and deadly to it. JOEL FEINBERG, RIGHTS, JUSTICE, AND THE BOUNDS OF LIBERTY 141 (1980); see also IREDELL JENKINS, SOCIAL ORDER AND THE LIMITS OF LAW 254–55 (1980) (“The doctrine of legal rights teaches us that declarations of rights are vain without an effective apparatus to implement them.”). Some writers include the element of enforceability in their definition of legal rights. See, e.g., MORRIS GINSBERG, ON JUSTICE IN SOCIETY 74 (1965) (“Legal rights are claims enforceable at law.”); JENKINS, supra, at 247 (“Natural rights are merely claims, regardless of the intellectual justification and emotional fervor with which they are pressed.”); BLACK’S LAW DICTIONARY 1324 (6th ed. 1990) (defining “right” as “legally enforceable claim of one person against another, that the other shall do a given act, or shall not do a given act”).

211. 503 U.S. 60 (1992); see supra notes 161–79 and accompanying text.

212. 403 U.S. 388 (1971).

213. U.S. CONST. amend. IV.

Fourth Amendment violation, the right also meant that Bivens could obtain compensation for the intrusion, terror, and property damage. Thus, the kind of remedy a person may obtain defines both the scope and meaning of the right.

Decisions about rights also directly affect remedies. If a court decides that a statutory provision does not confer a right on the plaintiff because the provision is not mandatory or because it is too vague and amorphous, then the plaintiff obviously gets no remedy at all. For example, in Pennhurst State School & Hospital v. Halderman, the Court concluded that the Bill of Rights provision in the statute at issue only made “findings” that were intended to be hortatory, not mandatory, and that the phrases “appropriate treatment” and “least restrictive setting” were too ambiguous to provide sufficiently clear notice to the States of what they agreed to if they took federal money. Because the statute conferred no rights on the plaintiffs, they were entitled to no remedy.

In other cases a court may decide that although a statute creates rights and duties, they are not as broad as the plaintiff’s claim. If plaintiffs are unable to fit their case within the court’s narrow definitions, they are effectively remediless. For example, in Central Bank of Denver, N.A. v. First Interstate Bank of Denver, N.A., an issuer of bonds defaulted, and bondholders sued the issuer and several other parties under § 10(b) of the Securities Exchange Act of 1934, which imposes private civil liability on those who commit a manipulative or deceptive act in connection with the purchase or sale of securities. Plaintiffs sought to hold one of the defendants, Central Bank, “secondarily liable under § 10(b) for its conduct in aiding and abetting the fraud.” The Court

215. Id. at 390–91.
216. Bush v. Lucas, 462 U.S. 367 (1983), and Schweiker v. Chilicky, 487 U.S. 412 (1988), discussed supra notes 147–148 and accompanying text, also illustrate this point. In Bush, the plaintiff was returned to his job and got backpay for an illegal retaliatory demotion, but was unable to obtain attorney’s fees or damages for the emotional distress and mental anguish. 462 U.S. at 371–72, 388. Similarly, in Schweiker, the plaintiffs eventually were restored to disabled status and were awarded full retroactive benefits after being wrongly terminated from a Social Security disability program, but they were unable to obtain damages for the emotional distress and other hardships suffered because of the delays in the receipt of their benefits. 487 U.S. at 425, 428–29. In both cases, the limits on the remedy limited the scope and meaning of the rights.
217. 451 U.S. 1 (1980); see supra notes 114–129 and accompanying text.
218. See supra notes 126–29 and accompanying text.
220. Id. at 168.
221. Id. (quoting plaintiffs’ complaint). The bonds were issued by a local Colorado building authority to finance public improvements at a planned residential and business center. Id. at 167. The bonds were secured by landholder assessment liens. Id. The bond covenants mandated that the land
declined to extend § 10(b) to impose a duty on parties who do not directly engage in the illegal deception, but merely aid or abet the violation.222 Plaintiffs thus had no rights against Central Bank, and no remedy.

A decision about a cause of action directly affects both rights and remedies. A cause of action in this context, whether created explicitly or implicitly, is something that says, in effect, “If someone violates your legal right you can sue him for relief.”223 Section 1983 creates what is probably the best-known federal cause of action.224 In awkward phrasing, it says that if a person acting under color of state law deprives someone of a federally protected right, the injured party can sue the wrongdoer for damages or equitable relief.225 A cause of action thus connects a right and a remedy. It is an essential link between a right and a remedy that enables the right to be enforced. Plainly, if you have no cause of action, you have no right and no remedy.226 A cause of action also may be

that was subject to the liens be worth at least 160% of the bonds’ total outstanding principal and interest. Id. The covenants also required the developer to give Central Bank, the indenture trustee for the bond issues, an annual report containing evidence that the 160% requirement was met. Id. Central Bank did not diligently investigate or ensure that the requirement was met. Id. at 168–69.

222. Id. at 177–78.

223. The meaning of the phrases “cause of action” and “right of action” has been elusive historically, see supra Part I.A., and it continues to be so today. Justice Powell defined “private cause of action” to mean “the right of a private party to seek judicial relief from injuries caused by another’s violation of a legal requirement.” Cannon v. Univ. of Chi., 441 U.S. 677, 730 n.1 (1979) (Powell, J., dissenting). Consider Justice Harlan’s definition in Bivens v. Six Unknown Named Agents of the Federal Bureau of Narcotics, 403 U.S. 388, 402 n.4 (1971) (Harlan, J., concurring): “The notion of ‘implying’ a remedy, therefore, as applied to cases like Bork, can only refer to a process whereby the federal judiciary exercises a choice among traditionally available judicial remedies according to reasons related to the substantive social policy embodied in an act of positive law.” Justice Brennan defined “cause of action” in Davis v. Passman, 442 U.S. 228, 239 n.18 (1979), as “a question of whether a particular plaintiff is a member of the class of litigants that may, as a matter of law, appropriately invoke the power of the court” and then separated that question from the question of what remedies should be available. Justice Brennan’s definition of cause of action is so incomplete as to be almost meaningless. The definition begs the question: For what purpose does the plaintiff want to invoke the power of the court?


225. Section 1983 reads:

Every person who, under color of any statute, ordinance, regulation, custom, or usage, of any State or Territory or the District of Columbia, subjects, or causes to be subjected, any citizen of the United States or other person within the jurisdiction thereof to the deprivation of any rights, privileges, or immunities secured by the Constitution and laws, shall be liable to the party injured in an action at law, suit in equity, or other proper proceeding for redress. . . .

Id.

226. For example, consider California v. Sierra Club, 451 U.S. 287 (1981), where the Court held that a fisherman had no cause of action under § 10 of the Rivers and Harbors Appropriations Act of 1899, which forbids obstruction to the navigable capacity of any waters of the United States unless
Rights, Rights of Action, and Remedies

defined in a way that limits the available remedy and thus limits the right as well. For example, Congress sometimes creates a cause of action for injunctive relief only. As explained above, this sort of limitation on the remedy necessarily limits the scope and meaning of the right.

In sum, the questions whether the statute confers a right, whether a cause of action should be implied, and whether a remedy can be granted do not have much meaning in isolation. A right without a remedy is not a legal right because it is not enforceable. A right without a cause of action is not a right for the same reason. A cause of action without a right or remedy is meaningless. Because a cause of action acts to connect a right and a remedy, it has no meaning if there is nothing to connect. Finally, a remedy obviously cannot stand alone either. One is entitled to a remedy only for the violation of a right.

2. Problems Caused by Trying To Separate the Inseparable

The Court’s attempts to separate these inseparable concepts have caused several problems. The Court’s analysis is necessarily incomplete when it discusses only one part of the rights-remedies equation. By focusing on only one part of the equation, the Court does not acknowledge the impact of its decision on the other parts. The Justices sometimes appear to confuse themselves by their overly narrow focus on

---


228. See supra notes 212–16 and accompanying text.

229. The Court’s separate treatment of rights, rights of action, and remedies is reminiscent of one of the Court’s tests for deciding whether a case arises under federal law. Under the so-called Mottley rule, a case arises under federal law only if the federal issue necessarily appears on the face of the plaintiff’s well-pleaded complaint, Louisville & Nashville R.R. Co. v. Mottley, 211 U.S. 149, 152 (1908), even if the case turns on an important issue of federal law that is the only real issue in the case. Id. The Court essentially puts on blinders and refuses to see what is actually going on. For a critique of the Mottley rule, see Donald L. Doernberg, There’s No Reason for It; It’s Just Our Policy: Why the Well-Pleaded Complaint Rule Sabotages the Purposes of Federal Question Jurisdiction, 38 Hastings L.J. 597 (1987).
a single factor. In other cases the Court either disguises or actually misstates what it is doing.

The continuing controversy over the status of \textit{Cort v. Ash}\textsuperscript{230} is one example of the confusion caused by separating rights, rights of action, and remedies. Several of the Justices appear genuinely perplexed and a bit angry that the ghost of \textit{Cort v. Ash} will not go quietly.\textsuperscript{231} Ironically, even those Justices most adamant that it be banished forever continue to cite it and discuss the four factors.\textsuperscript{232} Many of the Justices want to limit the inquiry in implied right of action cases solely to the second \textit{Cort} criteria: Did Congress intend to create the cause of action that the plaintiff asserts?\textsuperscript{233} It is very difficult, however, to focus only on the cause of action and ignore the rights being asserted and the remedy sought. As a result, the first and third \textit{Cort} criteria\textsuperscript{234} tend to creep back in, forcing even the anti-\textit{Cort} Justices to concede that these factors remain the means to determine the key issue of congressional intent.\textsuperscript{235}

The first \textit{Cort} factor continues to be important because the Court must consider the purpose of a cause of action in deciding whether

\textsuperscript{230} 422 U.S. 66 (1975); see supra notes 96–101 and accompanying text.

\textsuperscript{231} See supra notes 96, 101.


\textsuperscript{233} See, e.g., Middlesex County Sewerage Auth. v. Nat’l Sea Clammers Ass’n, 453 U.S. 1, 13 (1981) (“The key to the inquiry is the intent of the Legislature. . . . We look first, of course, to the statutory language . . . [and] then we review the legislative history and other traditional aids of statutory interpretation to determine congressional intent.”); Transamerica Mortgage Advisors, Inc. v. Lewis, 444 U.S. 11, 15–16 (1979):

While some opinions of the Court have placed considerable emphasis upon the desirability of implying private rights of action in order to provide remedies thought to effectuate the purposes of a given statute, . . . what must ultimately be determined is whether Congress intended to create the private remedy asserted, as our recent decisions have made clear.

\textit{Id.}; Touche Ross & Co. v. Redington, 442 U.S. 560, 575–76 (1979) (“The central inquiry remains whether Congress intended to create, either expressly or by implication, a private cause of action.”).

\textsuperscript{234} \textit{Cort}, 422 U.S. at 78. The Court stated:

First, is the plaintiff ‘one of the class for whose \textit{especial} benefit the statute was enacted,’ . . . that is, does the statute create a federal right in favor of the plaintiff? . . . Third, is it consistent with the underlying purpose of the legislative scheme to imply such a remedy for the plaintiff?

\textit{Id.}

Congress intended to create it. A plaintiff does not come to court asserting a cause of action in the abstract. Instead, a plaintiff asserts that a particular federal statute grants a right that the defendant has violated, and asks the Court to imply a cause of action to enforce that right. The Court must then look at the provision and ask whether “the statute create[s] a federal right in favor of the plaintiff.” It is almost impossible to answer the question of whether Congress intended to create a cause of action to enforce the right the plaintiff asserts without deciding whether the statute actually confers the right. Consequently, the first Cort factor continues to be central in implied right of action cases, even when the Court does not specifically identify it as such, but instead disguises it as an inquiry into the language of the statute or the legislative history.

Similarly, it is very difficult for the Court to avoid considering the third Cort factor—“is it consistent with the underlying purposes of the legislative scheme to imply [a particular] remedy for the plaintiff?” in deciding implied right of action cases. By definition, the statute does not explicitly create the cause of action, and the legislative history is silent as to whether Congress actually intended to create it. If a statutory provision appears to confer a right on the plaintiff, the Court would have difficulty deciding whether Congress wanted the right to be judicially enforceable without looking at whether it would foster or subvert the underlying statutory purposes to allow such lawsuits. Thus, unless the Court looks at the overall statutory purposes, it denies itself access to the information it needs to answer the question whether Congress intended to create the cause of action.

---

236. Cort, 422 U.S. at 78.
237. See, e.g., Transamerica, 444 U.S. at 16–17 (“[W]e begin with the language of the statute itself. . . . It is apparent that the two sections were intended to benefit the clients of investment advisers . . . .”); Touche Ross, 442 U.S. at 568–69 (“[O]ur analysis must begin with the language of the statute itself. . . . [Section] 17(a) neither confers rights on private parties nor proscribes any conduct as unlawful.”).
238. Cort, 422 U.S. at 78.
239. See Thompson, 484 U.S. at 179 (noting that “the legislative history of a statute that does not expressly create or deny a private remedy will typically be equally silent or ambiguous on the question” of whether Congress actually had in mind creation of private cause of action) (quoting Cannon v. Univ. of Chi., 441 U.S. 677, 694 (1979)); Transamerica, 444 U.S. at 18 (noting that it is not surprising that legislative history is silent on whether Congress intended to create private right of action given that Act does not explicitly provide any private remedies).
240. As with the first Cort factor, the Court sometimes discusses the third Cort factor without explicitly acknowledging that it is doing so. See, e.g., Mass. Mut. Life Ins. Co. v. Russell, 473 U.S. 134, 142 (1985) (“A fair contextual reading of the statute makes it abundantly clear that its draftsmen were primarily concerned with the possible misuse of plan assets, and with remedies that would protect the entire plan, rather than with the rights of an individual beneficiary.”) (footnote omitted); Transamerica, 444 U.S. at 18 (explaining that congressional intent to create private right
The Court’s separation of rights, rights of action, and remedies sometimes leads it to disguise what it is doing. *Wilder v. Virginia Hospital Ass’n*, one of the § 1983 cases in the *Thiboutot* line, provides such an example. In the guise of deciding whether the statute conferred a right and whether the private remedy the plaintiff sought was foreclosed by the existence of a comprehensive remedial scheme, the Court essentially performed a *Cort v. Ash* analysis. The majority claimed that its inquiry was different from deciding whether to imply a private right of action. The dissenting Justices disagreed in part, noting the overlap between the criteria for determining whether a statutory provision confers a right and the first *Cort v. Ash* criterion. Closer examination, however, reveals that the majority’s analysis actually overlaps with the first three *Cort* criteria.

In *Wilder*, a Virginia health care provider brought a § 1983 suit against state officials claiming that Virginia’s reimbursement plan violated a provision of the federal Medicaid Act requiring payment rates that “are reasonable and adequate to meet the costs which must be incurred by efficiently and economically operated facilities.” The plaintiff sought declaratory and injunctive relief requiring defendants to promulgate a new state plan and to provide interim rates commensurate with payments under the Medicare program. The defendants argued the Act did not create any enforceable rights and Congress had foreclosed enforcement of the Act under § 1983.

To evaluate the defendant’s first argument, the Court used the three-part test articulated in *Golden State Transit Corp. v. City of Los Angeles*; namely, whether Congress intended the provision to benefit someone like the plaintiff, whether it is mandatory, as opposed to merely reflecting a congressional preference, and whether it is too vague and amorphous for judicial enforcement. As the dissent noted, the first of these questions is very like the first *Cort* factor: “[I]s the plaintiff ‘one of the class for whose especial benefit the statute was enacted’?” The

---

242. *Id.* at 508 n.9.
243. *Id.* at 524–27 (Rehnquist, C.J., dissenting, joined by O’Connor, Scalia, and Kennedy, J.J.).
244. *Id.* at 502–04.
245. *Id.* at 504.
246. *Id.* at 514–22.
Court easily concluded that the requirement was met, stating that “[t]here can be little doubt that health care providers are the intended beneficiaries” of the provision requiring reasonable and adequate reimbursement rates.\(^{250}\)

The second *Golden State* question—whether the provision is mandatory or merely reflects a congressional preference—overlaps the second *Cort* inquiry: “[I]s there any indication of legislative intent, explicit or implicit, either to create [a private] remedy or to deny one?”\(^{251}\) If the provision is mandatory, that means the state must comply, which in turn suggests that the provision should be enforceable. The Court read the statutory language as saying the state reimbursement plan “must” provide for rates of payment that are reasonable and adequate as “a congressional command.”\(^{252}\) While this command need not necessarily be enforceable by a private lawsuit, the fact that Congress made the provision mandatory clearly is relevant to deciding whether there is “any indication of legislative intent, explicit or implicit,” to create a private remedy.\(^{253}\) Plainly, the Court’s inquiry into whether the provision was mandatory overlaps substantially with the second *Cort* question.

Finally, whether a private remedy was foreclosed by the existence of a comprehensive enforcement scheme overlaps the third *Cort* question: “[I]s it consistent with the underlying purposes of the legislative scheme to imply such a remedy for the plaintiff?”\(^{254}\) The *Wilder* Court noted the Medicaid Act contains few enforcement provisions.\(^{255}\) The Secretary of Health and Human Services can withhold approval of state plans or withhold federal funds from states whose plans do not comply with the Act.\(^{256}\) Because the amendments to the Act were intended to provide even less oversight by the Secretary,\(^{257}\) the existence of a private judicial remedy for health care providers clearly was consistent with Congress’s

\(^{250}\) *Wilder*, 496 U.S. at 510.

\(^{251}\) *Cort*, 422 U.S. at 78.

\(^{252}\) *Wilder*, 496 U.S. at 512. The Court rejected defendant’s argument that the statute only required the state to make findings and give assurances its rates are reasonable and adequate and did not require the state to adopt rates that actually are reasonable and adequate. *Id.* at 513–15.

\(^{253}\) The Court also cited additional evidence of congressional intent to allow private enforcement of the provision. The current provision amended an earlier one that allowed health care providers to sue in federal court for injunctive relief to ensure that they were reimbursed according to reasonable and adequate rates. *Id.* at 516. During debate on the amendment, legislators stated that it should not be construed to limit the rights of health care providers to seek injunctive relief in federal or state court. *Id.* at 517.

\(^{254}\) *Cort*, 422 U.S. at 78.

\(^{255}\) *Wilder*, 496 U.S. at 521.

\(^{256}\) *Id.*

\(^{257}\) *Id.* at 522.
intent to require states to reimburse them at reasonable and adequate rates. How else could the requirement be enforced? Indeed, although the Court does not say so, the existence of the private remedy is probably necessary to effectuate the underlying congressional purpose. In sum, although the Court purported to talk only about whether the statute confers a right and whether the private remedy the plaintiff sought was foreclosed by the existence of a comprehensive remedial scheme, the discussion is very similar to a Cort v. Ash analysis.

The Court’s separate treatment of rights, rights of action, and remedies sometimes leads it to misstate what it is doing. For example, in Gebser v. Lago Vista Independent School District and Davis v. Monroe County Board of Education, the two recent Title IX sexual harassment decisions, the Court purports only to delimit the remedy that a student can obtain from a school district, but in fact it also circumscribes students’ rights and rights of action under Title IX. Students do not have a general right to be free of sexual harassment at school because the Court imposes several conditions on recovery. Gebser holds that a damage remedy may not be recovered from a school district for sexual harassment of a student by a teacher unless a school district official with authority to intervene has actual notice of the teacher’s misconduct and does nothing. Davis also imposes an “actual notice-deliberate indifference” condition on recovery for student-on-student harassment, along with the further conditions that the district have “substantial control over both the harasser and the context in which the . . . harassment occurs” and that the harassment be so severe as to deny the victim equal access to school programs. A school district has a duty to protect students from sexual harassment only if these conditions are satisfied. If these conditions are not satisfied, the school district has no duty and the students have no right to be protected. If students have no right and no remedy, the Title IX implied right of action is of no use. Because the decisions in Gebser and Davis so plainly affect

---

260. These decisions are discussed supra notes 180–207 and accompanying text.
261. See Davis, 526 U.S. at 639 (“[A]t issue here is the question whether a recipient of federal education funding may be liable for damages under Title IX under any circumstances for discrimination in the form of student-on-student sexual harassment.”); Gebser, 524 U.S. at 284 (“[W]e have a measure of latitude to shape a sensible remedial scheme . . . .”).
262. See Gebser, 524 U.S. at 277, 290–91.
263. Davis, 526 U.S. at 645–47.
264. Id. at 633, 650.
265. Ms. Gebser’s complaint did not allege actual knowledge of harassment by school officials. Gebser, 524 U.S. at 278–79. Consequently, she had no right against the school district to be free of
students’ rights and rights of action, it is very misleading for the Court to contend that it is only addressing the scope of the damage remedy.

B. Two Models Governing Judicial Enforcement of Federal Statutes

Rights, rights of action, and remedies cannot be separated. When a court makes a decision about one of them it necessarily affects the other two. The Court’s attempt to separate the inseparable has caused confusion, distortion, and misstatement. The Court could integrate rights, rights of action, and remedies if it used one test for deciding whether a federal statutory provision should be judicially enforceable. To assist in framing a test, this section presents the adversarial and cooperative models for the judicial enforcement of federal statutes.

1. The Adversarial Model

The basic premise of the adversarial model is that the federal courts should play a limited role in the interpretation and implementation of federal legislation. The federal courts should not invade state prerogatives or perform legislative functions. As Justice Powell stated in Cannon v. University of Chicago,266 “[t]he dangers posed by judicial arrogation of the right to resolve general societal conflicts have been manifest to this Court throughout its history.”267 The federal courts must be modest, cautious, and even passive.

Why then is this model called “adversarial?” This term is used because the attitudes of some who support this model are not actually passive but instead are passive-aggressive.268 U.S. Supreme Court

that harassment, and her implied right of action under Title IX was entirely disabled. Ms. Davis’s complaint probably could survive the “substantial control,” “actual knowledge,” and “deliberate indifference” requirements because the harassment occurred at school. Ms. Davis’s mother complained repeatedly to school officials, and they failed to take any action. Davis, 526 U.S. at 633–35. It is unclear whether the harassment she alleged was so severe and pervasive that it barred her access to school. Id. at 635–36. If it was not, then she too had no right against the district to be free of that harassment, and her implied right of action was of no use.

266. 441 U.S. 677 (1979).
267. Id. at 744 (Powell, J., dissenting).
268. See Stewart & Sunstein, supra note 6, at 1199, 1230 (“[T]he formalist thesis [that federal courts must have some textual warrant, constitutional or statutory, for adding new remedies to administrative systems] denies the respect for the sovereign lawgiver upon which it is supposedly based and is thus internally inconsistent.”). Lack of respect for the legislatures is not new. See, e.g., Roscoe Pound, Common Law and Legislation, 21 HARV. L. REV. 383, 383 (1908) (“Not the least notable characteristics of American law today are the excessive output of legislation in all our jurisdictions and the indifference, if not contempt, with which that output is regarded by courts and lawyers.”).
Justices sometimes reveal their resentment of Congress. In *Middlesex County Sewerage Authority v. National Sea Clammers Ass’n*, Justice Stevens admitted that “a Court that is properly concerned about the burdens imposed upon the federal judiciary, the quality of the work product of Congress, and the sheer bulk of new federal legislation, has been more and more reluctant to open the courthouse door to the injured citizen.” Justice Powell believed that the *Cort v. Ash* test encouraged Congress to be irresponsible: “Congress is encouraged to shirk its constitutional obligation and leave the issue to the courts to decide. When this happens, the legislative process with its public scrutiny and participation has been bypassed, with attendant prejudice to everyone concerned.”

Proponents of the adversarial model seek to shift decision-making responsibility back to Congress. They believe that if Congress wants a statute to confer a right, create a cause of action, or provide a remedy, it must say so expressly, fully, and unambiguously. The Court recently has made plain that if legislation is vague or incomplete, it is Congress’s job to fix it, not the Court’s. As then-Justice Rehnquist put it...

---

270. *Id.* at 24–25 (Stevens, J., concurring in the judgment and dissenting in part); *see also* Merrill Lynch, Pierce, Fenner & Smith, Inc. v. Curran, 456 U.S. 353, 374 (1982) (“Our approach to the task of determining whether Congress intended to authorize a private cause of action has changed significantly, much as the quality and quantity of federal legislation has undergone significant change.”).
271. *Cannon*, 441 U.S. at 743 (1979) (Powell, J., dissenting). Hans Linde suggests that crass political maneuvering may underlie a legislative decision to define rights with vague and amorphous language and to omit express rights of action and remedies from a statute:

[As to] social legislation that places the burden of progress on those whom it regulates, a very low prospect of effectiveness may be the sine qua non of winning enactment of the law at all. . . . Contrary to the instrumentalist canon, the ineffectiveness of a law to achieve its goal may be itself a policy, a policy shared by the act’s opponents and some of its supporters, and may be the price for permitting the law to reach enactment. . . .

People have reasons for wanting a law, and the lawmaker will see a value in meeting their wishes, quite apart from any practical good it may do.

272. *See, e.g.*, Suter v. Artist M., 503 U.S. 347, 357–60 (1992) (holding that Congress did not unambiguously confer on child beneficiaries of Adoption Assistance and Child Welfare Act of 1980 right to enforce requirement that state make “reasonable efforts” to prevent child from being removed from his family, and once removed, to reunify him with his family); Pennhurst State Sch. & Hosp. v. Halderman, 451 U.S. 1, 18–19, 24–25 (1981) (holding that Bill of Rights provision of Developmentally Disabled Assistance and Bill of Rights Act of 1975 did not confer right on Pennhurst residents to “appropriate treatment” in “least restrictive” setting because Congress listed this “right” as “finding” and because phrases are too ambiguous to impose specific obligations).
Rights, Rights of Action, and Remedies

in *Touche Ross & Co. v. Redington*,\(^{273}\) the Court should not attempt to “improve upon the statutory scheme that Congress enacted into law.”\(^{274}\)

While it thus is appropriate to call this model “adversarial,” I do not mean to suggest that proponents of the adversarial model are unprincipled. To the contrary, several weighty arguments support their stance. Separation-of-powers principles preclude the federal courts from assuming legislative functions. It is not proper for the Court to transform aspirational provisions of statutes into mandatory provisions, thereby creating rights that Congress did not create. As the Court stated in *Rosado v. Wyman*,\(^{275}\) “Congress sometimes legislates by innuendo, making declarations of policy and indicating a preference while requiring measures that, though falling short of legislating its goals, serve as a nudge in the preferred directions.”\(^{276}\)

Similarly, Justice Powell argued eloquently in *Cannon v. University of Chicago*\(^{277}\) that judicial creation of a cause of action violated separation-of-powers principles. He believed that in addition to encouraging Congress to act irresponsibly, the *Cort* test “too easily may be used to deflect inquiry away from the intent of Congress, and to permit a court instead to substitute its own views as to the desirability of private enforcement.”\(^{278}\) In addition, he argued that “[d]etermining whether a private action would be consistent with the ‘underlying purposes’ of a legislative scheme permits a court to decide for itself what the goals of a scheme should be, and how those goals should be advanced.”\(^{279}\) According to this view, Congress knows how to create private rights of action explicitly when it wants to, and the Court should not take over the legislative role.\(^{280}\) Moreover, when a statute provides some remedies, courts should be “chary of reading others into it.”\(^{281}\)


\(^{274}\). *Id.* at 578; accord Univs. Research Ass’n v. Coutu, 450 U.S. 754, 770 (1981).


\(^{276}\). *Id.* at 413.

\(^{277}\). 441 U.S. 677 (1979).

\(^{278}\). *Id.* at 740 (Powell, J., dissenting).

\(^{279}\). *Id.* (Powell, J., dissenting). Similarly, in *Thompson v. Thompson*, 484 U.S. 174, 192 (1988), Justice Scalia stated in his concurring opinion that it was “dangerous to assume that, even with the utmost self-discipline, judges can prevent the implications they see from mirroring the policies they favor.”


\(^{281}\). Mass. Mut. Life Ins. Co. v. Russell, 473 U.S. 134, 147 (1985) (quoting *Transamerica*, 444 U.S. at 19); see also Northwestern Airlines, Inc. v. Transp. Workers Union, 451 U.S. 77, 94 n.30 (1981) ("A frequently stated principle of statutory construction is that when legislation expressly provides a particular remedy or remedies, courts should not expand the coverage of the statute to subsume

117
Proponents of the adversarial model also rely on principles of federalism. Modern federal legislation often regulates matters that traditionally were left to the states. Congress does not usually preempt state law, but instead legislates against the backdrop of state law. Reading federal statutes liberally to maximize the creation of rights, rights of action, and remedies may invade state prerogatives in ways that Congress did not intend.\(^{282}\) In many instances state remedies are available to cure wrongdoing and federal remedies are unnecessary. Federalism concerns are particularly important when plaintiffs sue state and local government. The Spending Clause cases that require Congress to use clear and unambiguous language when it imposes conditions on the grant of federal moneys to states\(^{283}\) protect the states from unexpected federal intrusion and control.

Some proponents of the adversarial model believe that \textit{Erie Railroad Co. v. Tompkins}\(^{284}\) and its progeny limit the power of the federal courts to create private remedies for violations of federal statutes. \textit{Erie}, of course, explicitly restricted the power of the federal courts to declare rules of general common law.\(^{285}\) Consequently, while it may have been permissible in the freewheeling days before 1938 to create causes of action and use any available remedy to redress a wrong, such practices are improper today.\(^{286}\)

Several pragmatic arguments also support the adversarial model. Congress’s lawmaking powers are superior to those of the courts because it can gather facts, set priorities, and accommodate different viewpoints in providing rights, rights of action, and remedies.\(^{287}\) Congress can

\(^{282}\) See, e.g., \textit{Daily Income Fund, Inc. v. Fox}, 464 U.S. 523, 541 (1984) ("[A] corporation’s rights against its directors or third parties with whom it has contracted are generally governed by state, not federal, law."); \textit{Cort v. Ash}, 422 U.S. 66, 84 (1975) ("Corporations are creatures of state law, and investors commit their funds to corporate directors on the understanding that, except where federal law expressly requires certain responsibilities of directors with respect to stockholders, state law will govern the internal affairs of the corporation.").

\(^{283}\) See \textit{supra} notes 114–29 and accompanying text.

\(^{284}\) 304 U.S. 64 (1938).

\(^{285}\) \textit{Id.} at 78.


\(^{287}\) See \textit{Tex. Indus., Inc. v. Radcliff Materials, Inc.}, 451 U.S. 630, 647 (1981) (noting that legislative bodies can provide "the kind of investigation, examination, and study that courts cannot," and that legislative process "involves the balancing of competing values and interests, which in our democratic system is the business of elected representatives." (quoting \textit{Diamond v. Chakrabarty}, 447 U.S. 303, 317 (1980)).
legislate prospectively and comprehensively, and can create administrative agencies to implement legislation. Courts, by contrast, lack the “self-starting investigatory and analytical capacities” needed to address broad social and economic problems, and are unable to ensure that enforcement will be centralized and coordinated. In addition, the complexity of modern legislation requires greater reliance on Congress. Intricate policy calculations may be necessary to adjust and coordinate enforcement of complex legislation, and judicial creation of additional causes of action or remedies may disrupt a delicate balance. Thus, when a statute contains a comprehensive remedial scheme including an integrated set of enforcement mechanisms, a court should presume that Congress deliberately omitted any additional remedies. A court should not imply any additional private rights of action, grant additional remedies not specifically authorized by the statute, or allow plaintiff to rely on § 1983 or other general causes of action that may exist outside the statute itself. Finally, and of ever-increasing importance,

---

288. See Stewart & Sunstein, supra note 6, at 1208–09; see also Carlson, 446 U.S. at 36 (Rehnquist, J., dissenting). Judge Henry J. Friendly once summarized a legislature’s advantages in making law:

[i]n the legislature’s superior resources for fact gathering; its ability to act without awaiting an adventitious concatenation of the determined party, the right set of facts, the persuasive lawyer, and the perceptive court; its power to frame pragmatic rules departing from strict logic, and to fashion a broad new regime or to bring new facts within an existing one; its practice of changing law solely for the future in contrast to the general judicial reluctance so to proceed; and, finally, the greater assurance that a legislative solution is not likely to run counter to the popular will: all these give the legislature a position of decided advantage, if only it will use it.


289. See Texas Industries, 451 U.S. at 646 (noting that “range of factors to be weighed in deciding whether a right to contribution should exist demonstrates the inappropriateness of judicial resolution of this complex issue”); see also Merrell Dow Pharm. Inc. v. Thompson, 478 U.S. 804, 811–12 (1986); Merrill Lynch, 456 U.S. at 377.

290. See Frankel, supra note 6, at 570–84 (arguing that private compensatory actions are ill-suited to deterrence system of securities laws and may hamper central purposes of those statutes); Stewart & Sunstein, supra note 6, at 1206–07 (arguing that judicial creation of private rights of action may usurp administrative agency’s responsibility for enforcement of statute and decrease legislative control over enforcement activity).

291. See Northwest Airlines, Inc. v. Transp. Workers Union, 451 U.S. 77, 97 (1981) (“The presumption that a remedy was deliberately omitted from a statute is strongest when Congress has enacted a comprehensive legislative scheme including an integrated system of procedures for enforcement.”); see also Texas Industries, 451 U.S. at 645 (same).


293. See Smith v. Robinson, 468 U.S. 992, 1011 (1984). The Court stated:
tighter standards would help reduce the flood of cases to the federal courts.294

2. **The Cooperative Model**

The basic premise of the cooperative model is that the federal courts should play a constructive and supportive role in interpreting and implementing federal legislation. Proponents of this model view the drafting, enactment, interpretation, and implementation of legislation as a single, ongoing process. All actors in this process, whether legislative, administrative, or judicial, must work together to help achieve the underlying goals of federal statutes. The process should not be a minefield where judges wait to pounce on legislators’ missteps, omissions, or vague language to nullify statutory provisions.295 Instead, the federal courts should construe and enforce federal law to protect basic personal rights and help ensure that government benefits and entitlements are distributed in a fair, timely, and complete manner. The cooperative model embraces the traditional equating of rights and remedies. Proponents believe the federal courts’ role “is to ensure completeness and consistency in the whole fabric of the law, including those legal norms established through regulatory schemes.”296

Proponents of the cooperative model disagree with proponents of the adversarial model about separation-of-powers and federalism principles. Separation-of-powers principles do not preclude a federal court from reading a statutory provision to entitle a plaintiff to the remedy sought, even if the provision does not confer a right on the plaintiff with crystal clarity or expressly designate the appropriate remedy.297 Under the

In light of the comprehensive nature of the procedures and guarantees set out in the [Education of the Handicapped Act] . . . we find it difficult to believe that Congress also meant to leave undisturbed the ability of a handicapped child to go directly to [federal] court with an equal protection claim to a free appropriate public education.

*Id.*; see also *Middlesex County Sewerage Auth. v. Nat’l Sea Clammers Ass’n*, 453 U.S. 1, 21 (1981) (concluding that comprehensiveness of remedies provided “demonstrates not only that Congress intended to foreclose implied private actions but also that it intended to supplant any remedy that otherwise would be available under § 1983”).


295. *See* Stewart & Sunstein, *supra* note 6, at 1230–31 (“If a judge insists that the sovereign use a particular verbal formula to confer authority, he restricts the sovereign’s lawmaking authority by precluding other approaches, such as reliance on background understandings.”) (footnote omitted).

296. *Id.* at 1228.

297. *See Merrill Lynch*, 456 U.S. at 375–76 (“Because the Rigsby approach has prevailed throughout most of our history, there is no merit to the argument . . . that judicial recognition of an implied remedy violates the separation-of-powers doctrine.”) (footnote omitted).
traditional standards, federal courts performed precisely this role from the founding of the Republic until the early 1970s. The founders recognized that separation of powers could not be rigid or absolute. Instead, they considered some blending of powers as necessary for effective government. Congress cannot see all of the situations to which a law may apply. Therefore, the federal courts must be free to fill statutory interstices to ensure completeness of consistency in federal law and to provide appropriate remedies. This exercise of judicial power is not a usurpation of legislative authority, but rather is a necessary supplement to it. Finally, if Congress believes that the courts have made a serious error in interpreting a federal statute, Congress can amend it.

Proponents of the cooperative model do not believe that principles of federalism preclude a federal court from playing a constructive and supportive role in interpreting and implementing federal legislation. Federal-state relations are involved only peripherally in the relationship between two branches of the federal government. It does not follow from the fact that Congress usually legislates against the backdrop of state law that Congress does not want federal statutes liberally and beneficially

298. See supra Part I.A.

299. See Youngstown Sheet & Tube Co. v. Sawyer, 343 U.S. 579, 635 (1952) (Jackson, J., concurring) (“While the Constitution diffuses power the better to secure liberty, it also contemplates that practice will integrate the dispersed powers into a workable government. It enjoins upon its branches separateness but interdependence, autonomy but reciprocity.”); JAMES O. FREEDMAN, CRISIS AND LEGITIMACY 18 (1978) (explaining that separation of powers has “not preclude[d] one branch of government from participating in functions assigned primarily to another”).


301. See Stewart & Sunstein, supra note 6, at 1228–31; Albert Tate, Jr., The Law-Making Function of the Judge, 28 LA. L. REV. 211, 222 (1968).


303. See Zeigler, supra note 6, at 720 (noting that traditional criteria for formulating remedies “allow a sympathetic, cooperative effort by the courts to work with Congress in effectuating underlying congressional purposes and goals”).
construed to achieve its goals. Moreover, state law often does not provide adequate remedies.\textsuperscript{304}

For somewhat similar reasons, proponents of the cooperative model believe that the \textit{Erie} doctrine is almost irrelevant to the question of whether a federal statutory provision entitles a plaintiff to the federal judicial remedy the plaintiff seeks.\textsuperscript{305} The \textit{Erie} doctrine addresses the question of whether federal or state law should apply in a federal lawsuit and has little to do with the relations between the federal courts and Congress.\textsuperscript{306} The federal courts may not create general federal common law in areas constitutionally left to state control, but the federal courts plainly may create specific common law in areas properly within federal cognizance.\textsuperscript{307} As Paul J. Mishkin stated, “At the very least, effective Constitutionalism requires recognition of power in the federal courts to declare, as a matter of common law or ‘judicial legislation,’ rules which may be necessary to fill in interstitially or otherwise effectuate the statutory patterns enacted in the large by Congress.”\textsuperscript{308}

Proponents of the cooperative model also reject many of the pragmatic arguments put forth to support limiting the role of the federal courts in interpreting and enforcing statutes. Congress plainly is better situated than the courts to make law, but this argument misses the point. A court does not legislate at large when it decides to enforce a specific statutory provision. Instead, its discretion is channeled and hedged about on all sides by Congress’s words. At this stage of the law making and enforcing process, a court’s perspective often will be superior. Congress cannot anticipate all of the situations to which a law may apply, and thus it cannot always specify in advance the precise remedy that justice requires.\textsuperscript{309} A court, by contrast, can assess the actual effectiveness of the

\begin{thebibliography}{99}

\bibitem{304} See, \textit{e.g.}, \textit{Gebser v. Lago Vista Indep. Sch. Dist.}, 524 U.S. 274, 301 n.11 (1998) (Stevens, J., dissenting) (responding to majority’s suggestion that plaintiff may have right of recovery against school district as matter of state law by noting that state law immunizes school district from liability).

\bibitem{305} See \textit{Brown}, \textit{supra} note 79, at 622–27; \textit{Foy}, \textit{supra} note 6, at 583 (“Nor does the doctrine of \textit{Erie Railroad v. Tompkins} have anything to do with the question of implied private remedies.”); \textit{Frankel}, \textit{supra} note 6, at 563–66; \textit{Hazen}, \textit{supra} note 6, at 1375–82.

\bibitem{306} \textit{See generally} \textit{Erie R.R. Co. v. Tompkins}, 304 U.S. 64 (1938).


\bibitem{308} Mishkin, \textit{supra} note 300, at 799–800.

\bibitem{309} See Felix Frankfurter, \textit{Some Reflections on the Reading of Statutes}, 47 COLUM. L. REV. 527, 529 (1947) (“The intrinsic difficulties of language and the emergence after enactment of situations not anticipated by the most gifted legislative imagination, reveal doubts and ambiguities in statutes that compel judicial construction.”); Stewart & Sunstein, \textit{supra} note 6, at 1229 (“It is a
remedies expressly authorized by Congress\textsuperscript{310} and decide whether granting the remedy plaintiff seeks would foster or frustrate congressional purposes. Thus, when faced with the question whether the federal statutory provision the plaintiff relies on creates an entitlement to the judicial remedy sought, courts should ask whether Congress, if it had considered this specific situation, would have said “aye” or “nay.”\textsuperscript{311}

Courts also should not be overly concerned about upsetting the delicate balance of statutory enforcement provisions. Deciding whether a particular statutory provision confers an enforceable right on the plaintiff is not intrinsically more difficult than other tasks courts routinely undertake.\textsuperscript{312} A court can delve as deeply as necessary into legislative history to determine whether granting plaintiff the remedy sought will seriously interfere with congressional purposes. Moreover, many statutory-enforcement mechanisms are neither comprehensive nor finely tuned.\textsuperscript{313} Congress often leaves a great deal of discretion to administrative agencies in selecting sanctions because the legislators cannot anticipate the exact activity that should be proscribed.\textsuperscript{314} Thus, the fact that Congress expressly authorizes remedies for violation of some statutory provisions should not deter a court from granting a remedy for violation of another provision.\textsuperscript{315} Finally, if the workload of the federal courts is too heavy, there are better ways to deal with the problem than by denying relief to people who otherwise deserve it.

III. A PROPOSED NEW STANDARD AND ITS APPLICATION

Proponents of the “pure” version of each model are not likely to agree on a test for deciding whether a statutory provision should be judicially enforceable. Advocates of the adversarial model would likely favor a standard that combined Justice Scalia’s position in Thompson v.
Thompson—no more implied rights of action—and a clear-statement rule requiring Congress to confer rights and authorize remedies in express, unambiguous terms. Advocates of the cooperative model, by contrast, would probably favor a return to the traditional standards that granted any available judicial remedy for a statutory violation as long as the statute was intended for the benefit of the class of persons of which the plaintiff was a member and the harm suffered was of a kind that the statute generally was intended to prevent.

Neither course is appropriate. It is impractical to return to the traditional standards. While “[f]ew principles of the American constitutional tradition resonate more strongly than . . . for every violation of a right, there must be a remedy,” the principle states an ideal that was never fully achieved and cannot be today. In addition, today Congress enacts an enormous amount of legislation, much of it vague, sprawling, and amorphous. It no longer seems workable to allow people to sue to enforce any provision of a federal statute that arguably is intended for their benefit. Nor, on the other hand, should courts adopt the overly restrictive standards suggested by the adversarial model. Even if legislators could write with perfect clarity, they simply cannot anticipate all of the situations to which their words may apply. Thus, the restrictive standards place an unrealistic responsibility on Congress and unfairly disadvantage those whom Congress seeks to help by legislating.

Instead, the courts should chart a middle course between these two extremes. The test should incorporate the generous, constructive impulses of the cooperative model and the cautionary concerns of the adversarial model. Courts may properly play a cooperative and supportive role in implementing federal legislation. A majority of the Justices have rejected the arguments that separation-of-powers principles and the Erie doctrine prohibit a federal court from implying a private right of action from a federal statute or granting a damage remedy to

317. See supra Part I.A.
319. See supra notes 15–17 and accompanying text.
320. See DAVID SCHOENBROD ET AL., REMEDIES: PUBLIC AND PRIVATE 50 (1990) (“[I]t is probably more accurate to say that where there’s a right, there may be many remedies, or none.”) If there is no remedy, of course, what Professor Schoenbrod, et alia, call a “right” is only a hope or a wish. See also Fallon & Meltzer, supra note 16, at 1777 (“[N]ot every violation of constitutional rights is centrally relevant to the question of enforceable duties for which the jurisdiction of an enforcement court is invoked.”); Zeigler, supra note 6, at 680–81 (suggesting many practical reasons why “rights” cannot have remedies).
enforce an existing cause of action.\textsuperscript{321} Courts should work with Congress to try to achieve the underlying goals of federal statutes. At the same time, courts should be sensitive to the pragmatic concerns of proponents of the adversarial model. They should not seek to convert every vague, aspirational provision of a federal statute into a font of enforceable rights. They should carefully examine how the remedy the plaintiff seeks would fit into the overall remedial scheme of the statute to ensure that the litigation does not undercut Congress’s purposes. They should also consider whether the plaintiff’s lawsuit will be manageable and whether it might lead to a flood of similar cases.

Courts should not take any part of the rights-remedies equation as a given in deciding whether the federal statutory provision the plaintiff relies on entitles the plaintiff to the judicial remedy he or she seeks. It is not helpful to assume a priori that the right, the right of action, or the remedy is secure, and then to focus only on the remaining elements. This sort of assumption is misleading, as demonstrated above.\textsuperscript{322} Because rights, rights of action, and remedies cannot be separated, a decision concerning one of them necessarily affects the other two. Thus, there is simply no point in pretending that one or more elements of the rights-remedies equation is secure in any lawsuit where an element is in play.

Because the elements are interrelated, the separate criteria developed by the Court for assessing each element are not sufficient to answer the overarching question of whether the statutory provision on which a plaintiff relies creates the rights and duties claimed and entitles the plaintiff to the particular remedy sought. The three-part \textit{Golden State} test for deciding whether a statute confers a right\textsuperscript{323} cannot alone answer whether Congress wanted a person such as this plaintiff to be able to enforce the right or to obtain a particular remedy. The \textit{Touche Ross-Transamerica} test, which uses only the second \textit{Cort} factor\textsuperscript{324} to decide

\begin{itemize}
\item \textsuperscript{321} See Franklin v. Gwinnett County Pub. Sch., 503 U.S. 60, 73–74 (1992) (rejecting argument that separation-of-powers principles prohibit “the authority to award appropriate relief”); Merrill Lynch, Pierce, Fenner & Smith v. Curran, 456 U.S. 353, 375–76 (1982) (“Because the \textit{Rigsby} approach prevailed throughout most of our history, there is no merit to the argument advanced by petitioners that the judicial recognition of an implied private remedy violates the separation-of-powers doctrine.”). Only Justices Powell and Rehnquist have stated that implication of a private right of action violates the \textit{Erie} doctrine. Carlson v. Green, 446 U.S. 14, 36–38 (1980) (Rehnquist, J., dissenting); Cannon v. Univ. of Chi., 441 U.S. 677, 732 (1979) (Powell, J., dissenting) (stating that \textit{Rigsby} was decided under regime of \textit{Swift v. Tyson}, 41 U.S. (6 Pet.) 1 (1842), and thus “cannot be taken as authority for the judicial creation of a cause of action not legislated by Congress”).
\item \textsuperscript{322} See supra notes 229–65 and accompanying text.
\item \textsuperscript{323} See supra notes 135–36 and accompanying text.
\item \textsuperscript{324} \textit{Cort} v. \textit{Ash}, 422 U.S. 66, 78 (1975) (“[I]s there any indication of legislative intent, explicit or implicit, either to create such a remedy or to deny one?”).
\end{itemize}
whether to create a private right of action, does not work very well in
making that decision, as is shown by the Court’s continuing reliance on
the other *Cort* factors.325 Finally, the presumption that where rights have
been invaded and a federal court has jurisdiction to hear the case, all
appropriate remedies are available to make good the wrong done,326 does
not help a court decide whether the statutory provision in question
actually confers the right the plaintiff claims or whether Congress would
want this plaintiff to obtain the specific remedy sought.

What is needed in place of these tests is a single, integrated test. This
Article suggests taking parts of all of the tests and combining them in a
new way that will allow the courts to play a cooperative, positive role in
interpreting and enforcing federal legislation while avoiding the pitfalls
identified by proponents of the adversarial model. The proposed test is
more an analytic process than a set of fixed criteria. The test has several
steps that focus on key questions and consider important cautionary
factors.327 The steps are set forth below.

**A. The Language of the Statute**

Courts routinely begin with the statutory language when interpreting
a statute,328 and the actual words are a logical starting point in this
context as well.329 The plaintiff claims the defendant violated a right

---

325. *See supra* notes 230–240 and accompanying text.
678, 684 (1946).
327. Preliminarily, remember that the analysis that follows is necessary only in cases where the
statutory provision fails expressly to provide one or more of the three elements. If the provision
clearly confers a right on the plaintiff, specifically creates a cause of action, and authorizes the exact
remedy the plaintiff seeks, then the court should simply proceed to determine whether the plaintiff
can prove the case.
[Congress’s] purpose is the statutory text adopted by both Houses of Congress and submitted to the
point for interpretation of a statute ‘is the language of the statute itself.’”) (quoting Consumer Prod.
Ass’n, 310 U.S. 534, 542 (1940) (“[T]he function of the courts is . . . to construe the language so as
to give effect to the intent of Congress.”).
329. The Court has routinely begun its inquiry with the language of the statute in implied right of
whether § 36(b) confers a right that could be judicially enforced by an investment company, we look
first, of course, at the language of the statute.”); Transamerica Mortgage Advisors, Inc. v. Lewis, 444
Redington, 442 U.S. 560, 568–69 (1979) (“[O]ur analysis must begin with the language of the
statute.”). And in cases in the *Thiboutot* line, the Court inquires whether a statutory provision
granted by a particular federal statutory provision. What does the provision say? Is there a clear and logical fit between the language and the defendant’s conduct? In other words, does it expressly or specifically prohibit the defendant from doing what he has done or require him to do something he has failed to do?330 In addition, does the provision appear to have been enacted to protect someone like the plaintiff? The standard for evaluating statutory language suggested in Cannon v. University of Chicago331 is helpful. Does "the language of the statute explicitly confer[] a right directly on the class of persons that include[] the plaintiff in the case," or does it only "create duties on the part of persons for the benefit of the public at large"?332 Congress is more likely to have intended the provision to be enforced in a private suit by plaintiff in the former instance than in the latter.

The first Cort criteria—"Is the plaintiff ‘one of the class for whose especial benefit the statute was enacted’”333—is too restrictive in one way and not restrictive enough in another. Use of the word “especial” suggests that a statute cannot confer enforceable rights on several groups of people in the same statutory provision. This reading is too restrictive. Congress can confer rights of primary and secondary importance in one provision and still intend that they all be enforced. If plaintiff can only show that he is a part of the class for whose secondary benefit the statute was enacted, that should not automatically disqualify him from enforcing the provision. In Cort itself, the Court denied a private right of action to stockholders under a federal criminal provision prohibiting corporations from making contributions to candidates in presidential elections because the provision’s primary purpose was to benefit the public at large by freeing elections from the power of corporate money, while protecting stockholders’ interests in seeing that corporate funds were not given to candidates whom they did not support “was at best a secondary concern.”334 It is not clear why Congress’s desire to protect stockholders did not support implication of a private right of action on their behalf.

---

330. If the answer to either of these questions is a clear no, the case is at an end. E.g., Central Bank of Denver, N.A. v. First Interstate Bank of Denver, N.A., 511 U.S. 164, 177 (1992) (noting that absence of text in Securities Exchange Act of 1934 imposing duty to refrain from aiding and abetting § 10(b) violation “resolves the case”).


332. Id. at 690–92 & n.13.


334. Id. at 81.
The first Cort factor also is not restrictive enough because it does not specifically require that the provision be mandatory rather than precatory. This requirement, which is borrowed from the Golden State test, is important. In enacting social legislation, Congress often begins with a lengthy policy statement explaining why the legislation is needed and what it is intended to accomplish. Such provisions often are only statements of aspirations written to introduce the legislation and to make Congress look good. While such provisions may indicate that the statute was enacted for the benefit of a particular group, Congress may not have intended that these general provisions be judicially enforceable.

Professor Henry Monaghan suggests that the Golden State test distinguishes “between those who are ‘incidental’ beneficiaries of federal programs imposing duties and those who are intentionally protected.” The especial benefit language of Cort may be intended to make this distinction as well. If Congress intended to benefit a group of which plaintiff is a member, even though this group is only the secondary or tertiary beneficiary of the provision, then courts should probably grant plaintiff a remedy. On the other hand, if Congress truly did not intend to benefit the group and the seeming benefit the plaintiff claims results from careless or imprecise drafting, a court probably should not grant relief. The problem, of course, is deciding where to draw the line.

Some examples help illustrate the distinction and the difficulty in drawing the line. In Massachusetts Mutual Life Insurance Company v. Russell, the plaintiff was a beneficiary under two employee benefit plans administered by Massachusetts Mutual. She alleged that the fiduciaries administering the plans improperly cut off her benefits for a back ailment and violated federal regulations by taking 132 days to process her claim. Although plaintiff’s benefits were restored and full retroactive benefits were paid, she sued for the financial losses suffered because her disabled husband was forced to cash out a retirement plan to tide them over and for psychological injury caused by the stress.

---

335. See, e.g., Individuals With Disabilities Education Act Amendments of 1997, Pub. L. No. 105-17, § 651, 111 Stat. 37, 123–24 (1997) (listing, inter alia, several goals that effective educational systems must seek to achieve in addressing needs of children with disabilities).
336. Monaghan, supra note 6, at 250.
338. Id. at 136.
339. Id. at 136–37.
340. Id. at 137.
The court of appeals held that § 409(a) of the Employee Retirement Income Security Act of 1974 (ERISA) entitled plaintiff to relief. Section 409(a) reads:

Any person who is a fiduciary with respect to a plan who breaches any of the responsibilities, obligations, or duties imposed upon fiduciaries by this title shall be personally liable to make good to such plan any losses to the plan resulting from each such breach, and to restore to such plan any profits of such fiduciary which have been made through use of assets of the plan by the fiduciary, and shall be subject to such other equitable or remedial relief as the court may deem appropriate . . . .

The Court reasoned that the long delay in processing plaintiff’s claim breached defendant’s fiduciary duty to process claims in a timely and diligent manner, that § 409(a) created a cause of action for a plan beneficiary, and that the language authorizing “such other equitable or remedial relief as the court may deem appropriate” gave the court wide discretion in ordering compensatory and punitive damages.

The U.S. Supreme Court read § 409(a) differently. The Court pointed out that § 409(a) only requires a fiduciary who breaches his duty “to make good to such plan any losses to the plan resulting from each such breach.” It does not make the fiduciary liable to individual beneficiaries. The next clause, which requires the fiduciary to restore to the plan any profits the fiduciary made through the misuse of plan assets, suggests that Congress’s main goal in writing § 409(a) was to prevent misuse of plan assets by plan administrators, not to provide recompense to beneficiaries injured by bureaucratic delays. The Court concluded that the lower court erred in jumping from the opening clause of § 409(a), which makes a fiduciary personally liable, to the catch-all remedial clause at the end, “skipping over the intervening language establishing remedies benefiting, in the first instance, solely the plan [and] divorc[ing] the phrase being construed from its context and construct[ing] an entirely new class of relief available to entities other than the plan.”
The U.S. Supreme Court probably has the stronger argument, but it is a close call. Section 409(a) by its terms confers rights on the plan, and not on individual beneficiaries, and imposes a duty on the fiduciary to compensate the plan for any breach rather than individuals. In a broader sense, however, § 409(a) is intended to benefit plan beneficiaries. As the Court admits, the purpose of § 409(a) is to ensure the plan remains financially sound and its assets are not stolen or squandered by plan administrators so that moneys are available for plan beneficiaries. 348 Indeed, the whole purpose of the plan is to ensure that employees get disability benefits to which they are entitled when they are entitled to them. 349 Thus, in writing § 409(a) Congress intended to benefit the group of which plaintiff is a member, even if only in a derivative or secondary way. Arguably, therefore, plaintiff should have been given relief under the final clause of § 409(a) that makes the defendant “subject to such other equitable or remedial relief as the court may deem appropriate.” 350

California v. Sierra Club 351 provides another example of the difficulty in distinguishing between the intentional and incidental beneficiaries of a statutory provision. In this case the U.S. Supreme Court interpreted the statute incorrectly and denied a remedy to people Congress clearly intended to protect. California transported water from the wetter north to the more arid central and southern parts of the state. 352 Fresh water was gathered and stored behind dams and then released as needed into the Sacramento-San Joaquin Delta. 353 From there the water was diverted into canals and aqueducts to carry it south. 354 The fresh water released into the Delta could be degraded by salt water from the Pacific, so the state proposed to build a canal that would by-pass the Delta. 355 The Sierra Club, joined by a commercial fisherman and a Delta landowner, filed suit to stop the state from building the canal. They alleged that present and proposed diversions of fresh water from the Delta made the Delta water too salty. 356 Plaintiffs claimed that the canal project would violate their rights under § 10 of the Rivers and Harbors Act of 1899, which prohibits “[t]he creation of any obstruction not

---

348. Id. at 142.
349. Id.
350. Id. at 139.
352. Id. at 290.
353. Id.
354. Id.
355. Id. at 291.
356. Id. at 292.
Rights, Rights of Action, and Remedies

affirmatively authorized by Congress, to the navigable capacity of any of
the waters of the United States . . . ." Section 10 also makes it unlawful
“to excavate or fill, or in any manner to alter or modify the course,
location, condition, or capacity of, any port, roadstead, haven,
harbor, canal, lake, harbor of refuge . . . or of the channel of any
navigable water of the United States, unless the work has been
recommended by the Chief of Engineers and authorized by the
Secretary of the Army . . . .”

The lower courts concluded that plaintiffs could avail themselves of a
private cause of action to enforce § 10. The Ninth Circuit Court of
Appeals reasoned that § 10 was “designed for the especial benefit of
private parties who may suffer ‘special injury’ caused by an unauthorized
obstruction to a navigable waterway.” The U.S. Supreme Court
disagreed. The Court said the question was not whether plaintiffs would
benefit from enforcement of § 10, “but whether Congress intended to
confer federal rights upon those beneficiaries.” In this case, the
language of the Act “states no more than a general proscription of certain
activities; it does not unmistakably focus on any particular class of
beneficiaries whose welfare Congress intended to further.” Thus,
Congress was not concerned with the rights of individuals but rather with
benefiting the public at large.

The court of appeals has the stronger argument. The statutory
language strongly suggests Congress intended to confer rights on anyone
who was directly harmed by an unauthorized obstruction, and this
certainly includes the plaintiffs. Section 10 states that it is unlawful “to
excavate or fill, or in any manner to alter or modify the course, location,
condition, or capacity of, any . . . haven, harbor, canal, lake, harbor of
refuge . . . or of the channel of any navigable water of the United States”
without the required permission. Building a forty-mile canal to divert
large amounts of fresh water would clearly alter the condition or capacity
of the Delta, a navigable water of the United States. Plaintiffs claimed
that this change would cause them measurable harm. Plaintiffs thus

357. Id. at 289 n.2.
358. Id. (quoting Rivers and Harbors Appropriation Act of 1899 § 10, 33 U.S.C. § 403 (1994)).
359. Id. at 292. The court of appeals held that the necessary approvals had been obtained as to one
part of the project. Id.
360. Id. at 293–94.
361. Id. at 294.
362. Id.
appears to be members of the class for whose especial benefit the statute was enacted. Indeed, people harmed by obstructions or changes in the condition of navigable waters would seem to be the primary group the statute was enacted to protect, while members of the general public are only incidental beneficiaries.

The U.S. Supreme Court stressed that the statute does not expressly state that § 10 was enacted for the benefit of people harmed by violation of the provision. 365 This distinguished *Sierra Club* from a case like *Cannon v. University of Chicago*, 366 where the statute explicitly identified the intended beneficiaries. 367 But it is easy to determine who Congress intended to protect in § 10. Although Congress did not say expressly it intended to benefit people harmed by obstructions, its intentions are clearly and fairly inferable from the statutory language. Requiring a certain form of words frustrates, rather than furthers, Congress’s intent. 368

---

365. *Id.* at 294.
367. *Id.* at 682 n.3 (discussing statute requiring that “[n]o person . . . on the basis of sex [shall] be . . . subjected to discrimination under any education program . . . receiving federal financial assistance”).
368. In *Suter v. Artist M.*, 503 U.S. 347 (1992), the Court engaged in a tortured reading of a statutory provision to deny plaintiffs relief. The plaintiffs brought a class action against an Illinois state agency charged with caring for abused and neglected children. They claimed the defendants were violating a provision of the Adoption Assistance and Child Welfare Act of 1980. *Id.* at 351–52. The Act was passed pursuant to Congress’s spending power and required that to be eligible for payments, a state “shall have a plan approved by the Secretary.” *Id.* at 351. The statute required, inter alia, that as of October 1, 1983, the plan provide “(A) prior to the placement of a child in foster care, to prevent or eliminate the need for removal of the child from his home, and (B) to make it possible for the child to return to his home . . . .” *Id.* The plaintiffs claimed that the agency was failing to comply with this provision because it did not assign caseworkers promptly to children placed in agency custody. *Id.* The district court ordered that a caseworker be assigned within three working days and the court of appeals affirmed. *Id.* at 353.

The U.S. Supreme Court reversed, holding that the provision did not unambiguously confer a right upon children to enforce the “reasonable efforts” language. *Id.* at 363. The Court admitted that the language was mandatory, but held that the provision merely required that states have a plan that provides that reasonable efforts will be made to keep children out of foster care, not that states actually make reasonable efforts to achieve this goal. *Id.* at 358. Moreover, the statute gave no guidance as to how reasonable efforts were to be measured, and thus, the meaning of the directive would vary from case to case. *Id.* at 359–60. The Court concluded that the provision imposed “only a rather generalized duty on the State, to be enforced not by private individuals,” but by a cut-off of federal funds if the state is not complying with its own plan. *Id.* at 363.

*Suter* shows the adversarial model operating at its cynical worst. Did Congress really mean to require only that states have a plan stating that reasonable efforts would be made to keep children out of foster care, but not mean to require that states actually make those efforts? It seems unlikely that Congress wanted to make state officials go through the meaningless ritual of writing things down on a piece of paper that was a dead letter before the ink dried. In addition, the reasonable efforts standard does not seem any less judicially enforceable than the myriad other statutory and...
If it is unclear whether the language of the provision that a plaintiff relies on entitles the plaintiff to the particular remedy sought, a court should broaden the inquiry beyond that provision. A court should look at the overall structure and purpose of the statute for guidance. Several questions are pertinent. Why did Congress enact this statute? What was it trying to accomplish? What problem was the legislature responding to, and how did it seek to cure or correct it? What are people required to do under the law, and what are they forbidden from doing? By answering these questions, a court can provide a context for interpreting the provision. It can determine whether granting plaintiff (and others similarly situated) the remedy sought would be consistent with the underlying purposes of the statute, or helpful in accomplishing them, or perhaps even necessary to accomplish them.

The suggestion that courts consider the overall statutory context is controversial because of the U.S. Supreme Court cases that purport to reject the third Cort criterion,369 which asks whether it is consistent with underlying statutory purposes to imply a private right of action for the plaintiff.370 As explained above, however, the Court often cannot avoid this inquiry because it needs the information to decide whether the legislature intended either to create or deny a private right of action.371 In addition, the Court expressly and routinely considers overall statutory structure and purpose when it decides whether a statute creates a right or authorizes a remedy.372 As Professors Stewart and Sunstein point out, “Because reliance on these background understandings is inevitable . . . [i]n interpreting statutes that are silent on the existence of private enforcement rights . . . such reliance should be viewed as an established and legitimate device of lawmaking through statutory construction rather than a controversial intrusion on legislative prerogatives.”373

common law requirements that people act reasonably. See, e.g., Wilder v. Va. Hosp. Ass’n, 496 U.S. 498, 519–20 (1990) (finding judicially enforceable statutory requirement that state plan for medical assistance provide rates of payment that “are reasonable and adequate to meet the costs which must be incurred by efficiently and economically operated facilities”). In sum, the Suter Court went out of its way to undercut Congress’s intent to require that states make reasonable efforts to keep children out of foster care.

369. See supra notes 90–96 and accompanying text.
371. See supra notes 239–40 and accompanying text.
372. See infra notes 383–96 and accompanying text.
373. Stewart & Sunstein, supra note 6, at 1231.
The Court candidly admitted it was considering the underlying statutory purposes and legislative history in deciding whether to imply a private right of action in *Thompson v. Thompson*. David and Susan Thompson engaged in a dispute over custody of their son Matthew that ended with conflicting orders from courts in Louisiana and California. To resolve the impasse, David brought suit in federal court under the Parental Kidnapping Prevention Act of 1980 to determine which state decree was valid. The Act did not explicitly create a private right of action in favor of a parent, so the Court looked at “the context of the *Act* with an eye toward determining Congress’ perception of the law that it was shaping or reshaping.” Congress passed the statute in response to a perceived epidemic of parental kidnapping. Because custody decrees generally are modifiable to further the best interests of the child, the constitutional and statutory full faith and credit requirements did not reliably ensure national enforcement of custody decisions. The Act was passed to make the Full Faith and Credit Clause more effective in custody cases. Given this focus, the Court thought it would be incompatible “with the purpose and context of the legislative scheme” to create a private right of action that would substitute a federal court determination for state court application of the enhanced full faith and credit requirements. Thus, the Court concluded that the overall statutory structure and purposes counseled against creation of a private right of action.

The U.S. Supreme Court often considers statutory structure and purpose in deciding whether a statutory provision confers a right. For example, in *Musick, Peeler & Garrett v. Employers Insurance of Wausau*, the Court relied on statutory structure and purpose in finding in a suit brought pursuant to § 10(b) of the Securities Exchange Act of 1934 and Rule 10b-5 of the Securities and Exchange Commission that a

---

375. Id. at 178.
376. Id.
377. Id. at 180.
378. Id. at 180–81.
379. Id. at 181.
380. Id. at 181–82.
381. Id. at 183. Moreover, explicit statements in the legislative history indicated that Congress did not want the federal courts to play the enforcement role that plaintiff sought. Id. at 183–87. Congress wanted more full faith and credit among state courts and did not want the federal courts “to play Solomon” where two state courts issued conflicting custody orders. Id. at 186.
382. Id. at 187.
defendant could seek a right to contribution from joint tortfeasors. The Court examined two other sections of the Act—§§ 9 and 18—that were “close in structure, purpose, and intent to the 10b-5 action.” Both sections contained an express right to contribution. The Court concluded: “We think that these explicit provisions for contribution are an important, not an inconsequential, feature of the federal securities laws and that consistency requires us to adopt a like contribution rule for the right of action existing under Rule 10b-5.”

The Court also routinely relies on overall statutory context in deciding whether a statute should be read to authorize the remedy the plaintiff seeks. In *Gebser v. Lago Vista Independent School District* and *Davis v. Monroe County Board of Education*, the two recent Title IX sexual harassment cases, the Court generally examined Title IX “to ensure that we do not fashion the parameters of an implied right in a manner at odds with the statutory structure and purpose.” In *Gebser*, the Court found “important clues” in Title IX’s express enforcement provisions that Congress did not intend to allow a damage remedy where liability was based on constructive notice or vicarious liability. The express means of enforcement by administrative agencies assumed that recipients would be on actual notice of violations before any action was taken because an agency could not initiate enforcement proceedings until it had notified the recipient of its failure to comply and determined that compliance could not be achieved by voluntary means. Similarly, the *Davis* Court looked to Title IX’s other prohibitions to help give content to the term “discrimination” in the sexual harassment context. Students are protected not only from discrimination, but also from being

---

384. Id. at 288.
385. Id. at 295.
386. Id. at 297.
387. Id. Similarly, in *Golden State Transit Corp. v. City of Los Angeles*, 493 U.S. 103 (1989), the Court looked beyond the statutory language in deciding that the National Labor Relations Act (NLRA) created rights that forbade the City to condition renewal of petitioner’s taxicab franchise on settlement of a labor dispute between the petitioner and its union. Id. at 104–05. The City argued that because the duties of government are not expressly set forth in the Act, the Act did not create rights against the City or the State. Id. at 111. The Court disagreed, stating that the “language, structure, and history of the NLRA” all show that Congress meant to protect “certain rights of labor and management against governmental interference.” Id.
390. See supra notes 180–207 and accompanying text.
392. Id. at 288.
393. Id.
“excluded from participation in” or “denied the benefits of” any program receiving federal financial assistance. Based on these other prohibitions, the Court concluded that sexual harassment amounts to prohibited discrimination only when it is so severe that students are effectively denied equal access to school programs.

As yet another way of examining overall statutory context, courts sometimes consider which constitutional power Congress used in enacting the legislation for guidance in interpreting it. As to statutes enacted pursuant to the spending power, Pennhurst State School & Hospital v. Halderman held that Congress must impose any conditions on a state accepting federal money in clear and unambiguous language or the Court will not enforce them. While the Pennhurst rule may be reasonable, in recent years the Court has transformed it into a new standard that governmental entities receiving federal funds are liable only for their own knowing and intentional wrongdoing. The Court signaled the change in Franklin v. Gwinnett County Public Schools when it stated that Pennhurst “observed that remedies were limited under . . . Spending Clause statutes when the alleged violation was unintentional” because the entity receiving federal funds “lacks notice that it will be liable for a monetary award.” The Court applied the new standard in Gebser v. Lago Vista Independent School District, holding that a school district can be liable under Title IX for a teacher’s sexual harassment of a student only when district officials have actual knowledge of the harassment and are deliberately indifferent to it.

Indeed, the rule appears to have taken on a life of its own and is no longer anchored to the Pennhurst rationale. Saying that a fund recipient can be held liable only when it is put on clear notice of what it must do if it takes federal funds is not that same thing as saying that it can be held liable only for its own intentional wrongdoing. Assume Congress enacts

395. Id.
396. Id.
397. 451 U.S. 1 (1980).
398. Id. at 17. Moreover, a state must be given the option of either assuming the costs of complying with federal regulations or withdrawing from the program. See Guardians Ass’n v. Civil Serv. Comm’n, 463 U.S. 582, 596–97 (1983) (White, J.); see also Rosado v. Wyman, 397 U.S. 397, 420–21 (1970).
400. Id. at 74.
402. Id. at 292. The Court also applied the new standard in Davis v. Monroe County Board of Education, 526 U.S. 629, 640–46 (1999) (holding school district liable under Title IX for student-on-student sexual harassment only when school officials have actual knowledge of harassment and deliberately fail to stop it).
a statute under the spending power that imposes a duty on states to supervise their employees to ensure they comply with the requirements of the new law and specifically imposes a respondeat superior standard of liability on the states for any wrongdoing by an employee. A state could be held liable for employee misconduct under the original Pennhurst rule because the states are clearly and unambiguously on notice of their duties. But states could not be held liable under the later cases that refuse to grant make-whole remedies except for knowing and intentional misconduct on the part of the fund recipient itself.

There may be good reasons not to impose liability on state and local government or supervisory officials for the actions of their subordinates unless the government or the official authorizes the action or at least knowingly allows it to continue, despite the fact that such limitations conflict with normal tort principles of respondeat superior. Holding local government or government officials liable on a respondeat superior or constructive-notice basis may conflict with sovereign-immunity principles or impose unfair burdens on the officials who are often not in a position to supervise subordinates closely. But these reasons have little to do with the rationale of the Pennhurst rule that it is unfair to hold a party to terms of a contract that are unclear or ambiguous. The Court has gone beyond assessing the statutory context of the provision on which plaintiff relies to making its own affirmative substantive policy choices about the scope of liability under statutes enacted pursuant to the spending power.

In sum, the overall structure and purpose of a statute often are crucial in determining whether the statutory provision on which a plaintiff relies creates the rights and duties claimed and entitles the plaintiff to the particular judicial remedy sought. The U.S. Supreme Court has looked to the statutory context in determining whether a statutory provision confers a right, creates a right of action, or authorizes a remedy. Reliance on these background understandings is a necessary and legitimate device for

---

403. The Court has imposed such limitations in § 1983 cases. The Court has refused to make municipalities liable under § 1983 for injuries inflicted by city employees unless they were acting pursuant to official government policy or custom. See Monell v. Dep’t of Soc. Servs., 436 U.S. 658, 692–95 (1978). Similarly, the Court has refused to impose § 1983 liability on city officials unless they affirmatively direct their subordinates to violate citizens’ civil rights. See Rizzo v. Goode, 423 U.S. 362, 371, 376 (1976). As Professors Doernberg and Wingate observe, “[t]he Court’s insistence in Rizzo on a direct, causal connection to establish individual liability compares closely with its insistence in Monell on finding a municipal statute, ordinance, custom, usage or policy in order to establish municipal liability.” DONALD L. DOERNBERG & C. KEITH WINGATE, FEDERAL COURTS, FEDERALISM AND SEPARATION OF POWERS 505–06 (2d ed. 2000).

construing unclear or ambiguous statutory provisions. The Court should continue to look to overall statutory structure and purpose, although it should be careful not to use the inquiry, as it does in the Spending Clause cases, as a means to impose its own ideas about the proper scope of liability.

C. Possible Reasons for Caution

Even if the language of the provision the plaintiff relies on and the overall statutory structure and purpose strongly support granting the plaintiff the particular remedy sought, a court should nonetheless consider several cautionary factors before allowing the case to proceed. A court should consider whether the action, and others like it, will be judicially manageable, whether granting a private remedy might interfere with the remedial scheme that Congress has expressly enacted, and whether allowing plaintiff to proceed will result in a flood of new lawsuits. There is, of course, no easy or mechanical way to determine when one or more of these factors justifies dismissing a lawsuit. But if federal courts may abstain entirely from hearing actions that Congress has explicitly directed them to hear, 405 surely they may decline for pragmatic reasons to hear lawsuits that Congress has not explicitly authorized.

1. Will the Plaintiff’s Action, and Others Like It, Be Judicially Manageable?

An action seeking to enforce a provision of a complex regulatory or entitlement statute may be unmanageable for several reasons. The action may simply be too complicated, raising scientific or economic issues that courts are not well-equipped to resolve.406 These problems may be particularly acute in class actions. As Professors Stewart and Sunstein point out, “Not only must the court resolve the question whether a violation of regulatory requirements occurred, but it also must attempt to measure the economic impact of noncompliance. In complex areas of economic regulation, such measurements may approach the

405. See, e.g., Younger v. Harris, 401 U.S. 37, 54 (1971) (ordering abstention in action seeking to interfere with pending state criminal prosecution); R.R. Comm’n v. Pullman Co., 312 U.S. 496, 500–01 (1941) (ordering plaintiff to proceed in state court where definitive interpretation of ambiguous provision of state law may accord plaintiff relief and obviate need to reach federal constitutional question raised by plaintiff’s complaint).

406. See Stewart & Sunstein, supra note 6, at 1293.
impossible.”407 An action also may require a federal court to resolve sensitive issues of human relations that have traditionally been resolved in the state courts. In Thompson v. Thompson,408 for example, the Court refused to authorize an action under the Parental Kidnapping Prevention Act of 1980 to determine which of two conflicting state custody decisions was valid because it did not want the federal judges to become entangled in state domestic-relations questions “that they have little expertise to resolve.”409

An action to enforce a broad statutory provision also may pose manageability problems if the plaintiff seeks correspondingly broad injunctive relief that would be very expensive or would involve substantial restructuring of a government program. Moreover, as Professors Stewart and Sunstein point out, some “statutory norms that allow considerable flexibility in regulating conduct” do not create clear “bipolar right-duty legal relations.”410 Judicial enforcement is difficult in such cases unless the parties or the court can redefine and narrow the lawsuit to make judicial enforcement manageable.

Blessing v. Freestone411 provides a good example of both of these problems. The case involved Title IV-D of the Social Security Act, a complex statute that requires a state, in exchange for federal monies, to “establish a comprehensive system to establish paternity, locate absent parents, and help families obtain support orders.”412 The statute also requires states to be in “substantial compliance” with its provisions.413 Plaintiffs were custodial parents of children eligible to receive child support services from the State of Arizona under the Title IV-D.414 They claimed the state agency charged with enforcing the statute had not taken adequate steps to obtain support payments from the fathers of their children because of “staff shortages, high caseloads, unmanageable backlogs, and deficiencies in the State’s accounting methods and record-keeping.”415 The plaintiffs brought a class action on behalf of all children and custodial parents residing in Arizona who were or would be entitled to services under Title IV-D seeking a declaratory judgment that the state was not in “substantial compliance” with the statute and injunctive relief

407. Id.
409. Id. at 186.
410. Stewart & Sunstein, supra note 6, at 1305.
412. Id. at 333–34.
413. Id. at 335.
414. Id. at 332.
415. Id. at 337.
requiring affirmative measures to achieve substantial compliance “throughout all programmatic operations at issue.”\footnote{416}

The Court initially expressed alarm that the plaintiffs’ broadly worded request for relief “essentially invited the District Court to oversee every aspect of Arizona’s Title IV-D program.”\footnote{417} The Court held the substantial-compliance provision of Title IV-D did not create a federal right.\footnote{418} Instead, it was merely a “yardstick” for the Secretary of Health and Human Services to use in measuring the system-wide performance of the State’s program.\footnote{419} The substantial-compliance provision and other provisions of Title IV-D requiring adequate staffing levels did not give rise to federal rights because:

[The link between increased staffing and the services provided to any particular individual is far too tenuous to support the notion that Congress meant to give each and every Arizonan who is eligible for Title IV-D the right to have the State Department of Economic Security staffed at a “sufficient” level.\footnote{420}]

In Stewart and Sunstein’s terms, the statutory provisions plaintiffs relied on did not “create bipolar right-duty legal relations.”\footnote{421} The Court ultimately remanded the case to the district court to give the plaintiffs an opportunity to narrow their claims and to identify specifically the rights they claimed defendants had violated.\footnote{422}

While the decision in Blessing seems correct, courts should not exaggerate manageability problems. Statutes often contain “reasonableness” or “adequacy” standards that may be enforced without getting a court into a quagmire. For example, in \textit{Wright v. Roanoke Redevelopment & Housing Authority},\footnote{423} the Court believed that HUD regulations requiring that a “reasonable” amount for utilities be included in the rent that a public housing authority could charge was not “beyond the competence of the judiciary to enforce” because the regulations set out guidelines for the authority to follow in making utility allowances.\footnote{424} Similarly, in \textit{Wilder v. Virginia Hospital Ass’n},\footnote{425} the Court held that a

\footnotesize{\begin{itemize}
  \item \textit{Id.}\footnote{416}
  \item Id. at 341.\footnote{417}
  \item Id. at 343.\footnote{418}
  \item Id.\footnote{419}
  \item Id. at 345.\footnote{420}
  \item Stewart & Sunstein, supra note 6, at 1305.\footnote{421}
  \item Freestone, 520 U.S. at 342, 349.\footnote{422}
  \item 479 U.S. 418 (1987).\footnote{423}
  \item Id. at 431–32.\footnote{424}
  \item 496 U.S. 498 (1990).\footnote{425}
\end{itemize}}
statutory provision giving state health care providers a right to a state medical plan that provided “reasonable and adequate” payment rates could be manageably enforced because the statute provided the benchmark of an “efficiently and economically operated facilit[y]” to help assess rates.426

The Court exaggerated the manageability problems in Suter v. Artist M. 427 when holding a statutory provision requiring caseworkers to make “reasonable efforts” to keep children out of foster care could not be workably enforced because the meaning of the directive would vary from case to case.428 The district court had entered a carefully limited injunction requiring only that the state assign a caseworker to each child placed in state custody within three working days of the time the case was first heard in juvenile court, and assign a new caseworker to a child within three working days of the time the original caseworker relinquished responsibility of the case.429 The efforts a caseworker should make to keep children with their biological parents were left up to the Agency and the caseworker. The district court apparently did not plan to review individual cases to see if efforts to keep children out of foster care were reasonable or to engage in elaborate, ongoing supervision of the details of the program. The district court order may not have ensured full compliance with the statutory provision, but the relief ordered clearly was judicially manageable.

2. Will Granting a Private Remedy Interfere with the Remedial Scheme that Congress Expressly Enacted?

Granting a plaintiff a remedy not authorized by a statute may occasionally interfere with the express remedial scheme that Congress enacted. In such a case, a court should be very reluctant to grant the remedy if it concludes the interference actually will occur. A court should not simply presume from the fact that Congress has enacted a comprehensive remedial scheme that any additional remedies are inappropriate. A court should deny relief only if the interference is obvious, or if Congress has expressly indicated that it does not want additional remedies.

426. Id. at 519. The Court concluded: “Although some knowledge of the hospital industry might be required to evaluate a State’s findings with respect to the reasonableness of its rates, such an inquiry is well within the competence of the Judiciary.” Id. at 520.
428. Id. at 359–60.
429. Id. at 353.
Professors Stewart and Sunstein assert that a mix of public and private enforcement is often the best means of enforcing federal regulatory or entitlement programs. They find “largely unpersuasive” the objection that judicial authorization of private enforcement might upset legislative fine-tuning of remedies:

Statutory language is often so general or ambiguous that the legislature could not confidently predict the specific kind of activity against which enforcement action will be brought. The fine-tuning argument is even less convincing when, as is common, an agency has a lump-sum budget to cover enforcement of numerous provisions. Moreover, statutes often give agencies considerable discretion in selecting among alternative sanctions and penalty levels. These considerations severely weaken the argument that the legislature anticipates and controls enforcement levels in a way that would be undermined by judicial creation of private rights of action.

Consequently, a court should not assume that additional remedies are inappropriate simply because a remedial scheme is comprehensive or already includes some private remedies. As Stewart and Sunstein point out, refusing to grant an additional private remedy in such circumstances begs the question by assuming that congressional silence reflects an intention to foreclose judicial creation of such rights. It is equally plausible that sanctions and appropriations have been tailored in contemplation of possible private enforcement. Congress may wish to take advantage of the courts’ ability to draw upon experience with implementation of a particular regulatory program and to judge the impact and desirability of private rights of action.

Thus, a court should deny relief only if the defendant can demonstrate that the private remedy plaintiff seeks would actually undermine or work at cross purposes with the remedies Congress has authorized. This analysis suggests that the Court was wrong in *Middlesex County Sewerage Authority v. National Sea Clammers Ass’n* to deny summarily a private remedy under federal environmental statutes to fisher-

---

430. Stewart & Sunstein, *supra* note 6, at 1202. Although they recognize that there is some danger private actions may usurp an agency’s responsibility for regulatory implementation, *id.* at 1206–07, they also state private actions can be a useful supplement to public enforcement, which is often inadequate because of budget restraints. *Id.* at 1214.  
431. *Id.* at 1290 (footnotes omitted).  
432. *Id.* at 1291 (footnotes omitted).  
men injured by the dumping of sludge and sewage into the ocean off the New York and New Jersey coasts. The Court denied relief simply because the statutes contained “elaborate enforcement provisions” without even considering whether the remedy the plaintiff sought would interfere with the statutory enforcement scheme or otherwise undermine legislative intent.\textsuperscript{434} The Court also was wrong in \textit{Massachusetts Mutual Life Insurance Co. v. Russell}\textsuperscript{435} in refusing to authorize a private remedy for extra-contractual damages on behalf of a beneficiary of an employee benefit plan against a fiduciary of the plan without considering whether the remedy would undermine or be inconsistent with other remedies allowed by the plan. The Court wrongly presumed from ERISA’s “interlocking, interrelated, and interdependent remedial scheme”\textsuperscript{436} that Congress did not intend any additional remedies without pointing to any express statements to support that conclusion.\textsuperscript{437}

In \textit{Smith v. Robinson}\textsuperscript{438}, by contrast, the Court appeared to apply the correct standards in deciding that a private remedy would actually interfere with express statutory remedies. Plaintiff, a child suffering from cerebral palsy and a variety of physical and emotional handicaps, brought suit charging that a Rhode Island school district had denied him the right to a free, appropriate public education.\textsuperscript{439} He claimed a denial of due process and equal protection under the Fourteenth Amendment and § 1983, and a violation of the Education of the Handicapped Act (EHA).\textsuperscript{440} The plaintiff eventually won on the merits and claimed attorneys fees.\textsuperscript{441} Fees were recoverable on the § 1983 claim, but not under the Act.\textsuperscript{442} Whether plaintiff was entitled to fees thus depended on whether the § 1983 claim was viable or had been supplanted by the Act. The Court concluded that both the provisions of the statute and the

\begin{footnotesize}
\begin{itemize}
\item[434.] \textit{Id.} at 14–15. The Court in a footnote claimed that the Senate Reports on the two statutes emphasized the limited nature of the citizen suits authorized, although it did not explain what the limitations were. \textit{Id.} at 17 n.27. It also quoted from the remarks of a single legislator during debates on another statute, the Clean Air Act, to the effect that a private damage remedy was not authorized by that statute. \textit{Id.} Because the citizen-suit provision of one of the statutes was modeled on the parallel provision of the Clean Air Act, the Court concluded that Congress would not have approved the fishermen’s suit for damages. \textit{Id.} These fragmentary references do not amount to a serious and thorough consideration of whether the plaintiffs’ lawsuit would have interfered with the statutory enforcement scheme or undermined legislative intent. \textit{Id.}
\item[435.] 473 U.S. 134 (1985).
\item[436.] \textit{Id.} at 146.
\item[437.] \textit{Id.} at 146–48.
\item[438.] 468 U.S. 992 (1984).
\item[439.] \textit{Id.} at 995.
\item[440.] \textit{Id.} at 994.
\item[441.] \textit{Id.} at 998.
\item[442.] \textit{Id.} at 1002.
\end{itemize}
\end{footnotesize}
legislative history “indicate that Congress intended handicapped children with constitutional claims to a free appropriate public education to pursue those claims through the carefully tailored administrative and judicial mechanism set out in the statute.”

In reaching this conclusion, the Court appeared to consider whether independent § 1983 claims would actually interfere with these mechanisms. The Court stated that allowing § 1983 claims would “render superfluous most of the detailed procedural protections outlined in the statute.” Moreover, “it would also run counter to Congress’ view that the needs of handicapped children are best accommodated by having the parents and the local education agency work together to formulate an individualized plan for each handicapped child’s education.” Finally, the Court stated that “[a]llowing a plaintiff to circumvent the EHA administrative remedies would be inconsistent with Congress’ carefully tailored scheme.”

Thus, in Smith the Court appeared to assess the actual negative impact of an unauthorized private remedy on the Act’s remedial mechanism, rather than merely assuming from the existence of a comprehensive remedial scheme that no additional private remedies should be allowed.

The Court also was wrong in Touche Ross & Co. v. Redington and Transamerica Mortgage Advisors, Inc. v. Lewis to deny private remedies simply because Congress provided private remedies under other sections of the statute involved in each case. In Touche Ross, the Court refused to allow a private damage remedy under § 17(a) of the Securities Exchange Act of 1934 against an accounting firm that conducted an allegedly improper audit and certification of a broker-dealer’s financial statement. The Court noted that several nearby sections of the Act created private rights of action and concluded from this that Congress did not want a private remedy under § 17(a): “Obviously, then, when Congress wished to provide a private damage remedy, it knew how to do so and did so expressly.”

The Court drew a similar inference in Transamerica. It refused to allow a private remedy under § 206 of the Investment Advisers Act of 1940 against the investment advisers of a trust charged with fraud and breach of fiduciary duty because other sections of the Act expressly authorized private remedies.

443. Id. at 1009.
444. Id. at 1011.
445. Id. at 1012.
446. Id. (emphasis added).
449. Touche Ross, 442 U.S. at 579.
450. Id. at 571–72.
remedies while § 206 did not.\textsuperscript{451} In both cases, the Court should instead have asked whether allowing a damage remedy would have interfered with other enforcement provisions, or whether there were any statements in the legislative history showing that Congress did not want a private remedy under the sections of the acts on which plaintiffs relied.

The Court has followed the correct approach in other cases. For example, in \textit{Musick, Peeler & Garrett v. Employers Insurance of Wausau},\textsuperscript{452} the Court considered whether allowing a right of contribution from joint tortfeasors in a 10b-5 action would “conflict with Congress’ own express rights of action.”\textsuperscript{453} The Court noted that §§ 9 and 18 of the Securities Exchange Act, which are close in structure and purpose to § 10(b), both expressly provide for a right of contribution.\textsuperscript{454} Instead of inferring from the existence of these other private remedies that it should not create a contribution remedy under § 10(b), the Court concluded that it should.\textsuperscript{455} The Court also considered whether the contribution remedy would “detract[] from the effectiveness of the 10b-5 implied action or interfere[] with the effective operation of the securities laws,” and concluded that it would not.\textsuperscript{456} Similarly, in \textit{Cannon v. University of Chicago},\textsuperscript{457} the Court stated: “The fact that other provisions of a complex statutory scheme create express remedies has not been accepted as a sufficient reason for refusing to imply an otherwise appropriate remedy under a separate section.”\textsuperscript{458} Instead, the Court required some “other, more convincing, evidence that Congress meant to exclude the remedy.”\textsuperscript{459}

In sum, a court should not assume additional remedies are inappropriate simply because Congress has enacted a comprehensive remedial scheme or has included some private remedies in a statute. If the other factors courts consider point in favor of authorizing the remedy plaintiff seeks, the court should grant it unless it would interfere with the statute’s express remedial scheme or Congress has expressly ruled out the remedy.

\begin{footnotesize}
\begin{enumerate}
\item Transamerica, 444 U.S. at 19–20.
\item 508 U.S. 286 (1993).
\item Id. at 295.
\item Id. at 295–97.
\item Id. at 297.
\item Id. at 298.
\item 441 U.S. 677 (1979).
\item Id. at 711.
\item Id.
\end{enumerate}
\end{footnotesize}
3. **Will Allowing the Plaintiff To Proceed Result in a Flood of New Lawsuits?**

A court may properly consider whether authorizing a remedy for the plaintiff will inundate the courts with lawsuits from other people similarly situated. In a time when federal judicial resources are stretched thin, the federal courts must not take on more cases than they can competently adjudicate. Many times in our history the U.S. Supreme Court has adopted rules to help keep caseloads manageable. For example, following the 1875 grant of general civil federal question jurisdiction to the lower federal courts, the U.S. Supreme Court adopted the rule that a federal question must appear on the face of the plaintiff’s well-pleaded complaint, thus greatly restricting the number of cases that could be brought in federal court. Similarly, the traditional equitable doctrine that a federal court will not intervene in a pending state criminal proceeding was given new life in *Younger v. Harris* by a Court presumably fearful that the federal courts were about to be engulfed by thousands of lawsuits from state criminal defendants alleging constitutional violations in their state proceedings. In addition, the decisions of the Burger Court restricting the availability of federal habeas corpus responded to the dramatic increase in habeas corpus filings in the late 1960s and early 1970s. And indeed, the Court has acknowledged that the “increased complexity of federal legislation and the increased volume of federal litigation” prompted the adoption of the strict new standards for implying private rights of action in the late 1970s and early 1980s.

While caseload is a legitimate concern, the world is increasingly busy, and the federal courts have become more efficient and have

---

461. For a discussion of the cases adopting this rule, see Doernberg, supra note 229, at 611–18.
463. See Stefanelli v. Minard, 342 U.S. 117, 123–24 (1951) (asserting that allowing state defendants to come to federal court to raise constitutional challenges to procedures followed in their pending state criminal cases “would invite a flanking movement against the system of State courts by resort to the federal forum”); see generally Donald H. Zeigler, *Federal Court Reform of State Criminal Justice Systems: A Reassessment of the Younger Doctrine from a Modern Perspective*, 19 U.C. DAVIS L. REV. 31, 92–98 (1985) (discussing whether flood of litigation actually would occur if Younger doctrine were abandoned).
adopted substantial reforms to enable them to keep pace with increasing caseloads.\textsuperscript{466} Finally, as Justice Harlan once said:

Judicial resources, I am well aware, are increasingly scarce these days. Nonetheless, when we automatically close the courthouse door solely on this basis, we implicitly express a value judgment on the comparative importance of classes of legally protected interests. And current limitations upon the effective functioning of courts arising from budgetary inadequacies should not be permitted to stand in the way of the recognition of otherwise sound constitutional principles.\textsuperscript{467}

Justice Harlan’s remarks may sound quaint today, but the maxim “where there is a right, there is a remedy” should still have enough force that courts will not deny remedies unless the practical consequences of granting a remedy would seriously hamper their ability to function.

CONCLUSION

In \textit{Middlesex County Sewerage Authority v. National Sea Clammers Ass’n},\textsuperscript{468} Justice Stevens stated: “When should a person injured by a violation of federal law be allowed to recover his damages in a federal court? This seemingly simple question has recently presented the Court with more difficulty than most substantive questions that come before us.”\textsuperscript{469} This Article suggests that the Court caused much of the difficulty itself by separating rights, rights of action, and remedies, and talking about them as though they were independent of one another. Cases asking whether a statute confers a right, implies a right of action, or authorizes a remedy all involve the same basic question: Does the applicable statutory provision entitle a plaintiff to the particular judicial remedy he or she seeks? This Article has proposed a single, integrated test to answer this question. If the Court were to follow this standard, it would take the first step in putting its rights-remedies jurisprudence in order.


\textsuperscript{468} 453 U.S. 1 (1981).

\textsuperscript{469} Id. at 22–23 (Stevens, J., joined by Blackmun, J., concurring in judgment and dissenting in part).
QUALIFYING IMMUNITY: PROTECTING STATE EMPLOYEES’ RIGHT TO PROTECT THEIR EMPLOYMENT RIGHTS AFTER ALDEN v. MAINE

Raymond J. Farrow

Abstract: Recent U.S. Supreme Court decisions have barred state employees from bringing private suits against their state employers to recover back wages due them as a result of having been paid in violation of the Fair Labor Standards Act (FLSA). This Comment proposes that the only method by which state employees may protect their FLSA rights on their own behalf is to bring suits against responsible state supervisory personnel in their individual capacities. Although such actions are not barred by sovereign immunity, the potential ability of state agents to invoke a defense of “qualified immunity” would severely impair state employees’ ability to protect their FLSA rights. This Comment therefore argues that there is no basis either in the language of the FLSA or at common law for applying the doctrine of qualified immunity to state officers liable for violations of the FLSA.

In December of 1993, John Alden and ninety-five current and former state probation officers sued the state of Maine seeking to enforce their right under the Fair Labor Standards Act (FLSA) to receive time-and-one-half pay for overtime hours worked. The district court determined that Maine had failed to pay the officers appropriate overtime compensation and in a subsequent decision ordered the computation of damages due each officer in the form of two years of backpay. Five years later, after an odyssey through the federal and state court systems, John Alden and his fellow officers found themselves unable to collect the sums awarded them under the FLSA. The U.S. Supreme Court determined that no court in the land had the power to force Maine to defend itself against Alden’s attempt to enforce his statutory rights. A

1. Mills v. Maine, 839 F. Supp. 3, 4 (D. Me. 1993) aff’d 118 F.3d 37, 41 (1st Cir. 1997). The original named plaintiff Jon Mills was later replaced as the named plaintiff by John Alden. To avoid confusion, this Comment refers to Alden as the named plaintiff. See Mills v. Maine, 118 F.3d 37, 41 (1st Cir. 1997) (explaining that ninety-six officers were included in original suit).


controversial pair of U.S. Supreme Court decisions regarding the scope of state sovereign immunity from suit had denied Congress the ability to subject states to private suit for violating federal laws enacted pursuant to Congress’s Commerce Clause powers. As a result, the state of Maine would forever be able to keep the thousands of dollars in wages that had been wrongfully withheld from Alden and his colleagues.

Despite the U.S. Supreme Court’s dramatic re-interpretation of the doctrine of sovereign immunity in the 1990s, state employees such as Alden may have the ability to protect their right to recover back wages due to them, as specified by the FLSA. The FLSA provides an expansive definition as to who may be considered liable for a violation of the FLSA, specifically allowing complaining parties to seek relief from “any person acting directly or indirectly in the interest of an employer.” The FLSA thereby provides a cause of action allowing state employees to sue appropriate state supervisory personnel responsible for allegedly violative employment practices. While sovereign immunity does not protect individual state agents sued in their individual capacities, the doctrine of qualified immunity is available to government officials sued for violating a plaintiff’s constitutional rights when acting within the

6. In Kimel v. Fla. Bd. of Regents, 528 U.S. 62 (2000), Justice Stevens harshly criticized one of the Court’s recent sovereign immunity cases:

I remain convinced that . . . the decision of five Justices in Seminole Tribe . . . was profoundly misguided. Despite my respect for stare decisis, I am unwilling to accept Seminole Tribe as controlling precedent . . . . [T]he reasoning of that opinion is so profoundly mistaken . . . that it has forsaken any claim to the usual deference or respect owed to decisions of this Court.

Id. at 97–98 (Stevens, J., concurring in part, dissenting in part).

7. Alden, 527 U.S. at 754 (holding that Congress lacks power to subject states to private suit in state court when legislating under its Article I powers); Seminole Tribe, 517 U.S. at 72–73 (same result in federal court).

8. The exact amount owed had not been determined when the Seminole Tribe decision was issued. See Mills v. Maine, 118 F.3d 37, 41 (1st Cir. 1997).

9. Only seven years before Seminole Tribe, the Court stated that Congress had the power to abrogate sovereign immunity when acting pursuant to its Commerce Clause powers. See Pennsylvania v. Union Gas Co., 491 U.S. 1, 14–15 (1989).


course of their official duties. Qualified immunity protects government officers from liability if their conduct was objectively reasonable in light of clearly established law. If qualified immunity were available to defendants facing FLSA claims, state employees’ ability to protect their FLSA rights would be consider-ably weakened. However, an examination of the basis for granting government officials qualified immunity from civil rights claims leads to the conclusion that this defense should not be available to bar an FLSA action brought against state agents in their individual capacities.

Part I of this Comment outlines the substantive features of the FLSA and examines the basis on which courts have imposed individual liability for FLSA violations. The historical basis for granting state employees limited immunity from suit when acting within the scope of their official duties is developed in Part II. Parts III and IV explain the two major recent U.S. Supreme Court decisions regarding sovereign immunity and their impact on the ability of state employees to protect their rights under the FLSA. Part V analyzes why the FLSA should be interpreted to provide a cause of action against state supervisory personnel responsible for FLSA violations and proposes that the common law defense of qualified immunity for governmental officials has no application to such actions.

I. THE FAIR LABOR STANDARDS ACT

The FLSA establishes minimum standards for employment practices within both the private and public sector. The FLSA establishes a cause of action for aggrieved employees against their “employer,” a term that has a specific interpretation in the context of the FLSA.

15. The U.S. Supreme Court has noted that qualified immunity protects “all but the plainly incompetent or those who knowingly violate the law.” Malley v. Briggs, 475 U.S. 335, 341 (1986).
A. The FLSA Protects State Employees from Underpayment for Overtime Hours Worked

Enacted in 1938, the FLSA \textsuperscript{16} applies to all enterprises in the nation in which interstate activities take place.\textsuperscript{17} A series of amendments from 1966 to 1985 extended FLSA coverage to states in their roles as employers.\textsuperscript{18} The FLSA establishes a federal minimum wage to be paid to all employees\textsuperscript{19} and requires that an overtime wage of time-and-a-half apply for all hours worked in excess of forty hours per week.\textsuperscript{20} Specific exemptions to the overtime requirement apply to certain classes of state employees,\textsuperscript{21} as well as to executive, administrative, and professional employees.\textsuperscript{22} The FLSA imposes strict liability on any employer who knowingly or unknowingly pays an employee less than the minimum wage or fails to pay a non-exempt employee time-and-a-half for hours worked beyond the statutory level.\textsuperscript{23} The U.S. Supreme Court has affirmed that the FLSA represents a valid exercise of Congress’s constitutional powers under the Commerce Clause,\textsuperscript{24} and that Congress

\textsuperscript{17} See 29 U.S.C. § 206(a) (1994).
\textsuperscript{19} 29 U.S.C. § 206.
\textsuperscript{20} Id. § 207(a)(1).
\textsuperscript{21} Id. § 207(k), (o), (p) (providing exceptions for employees of public agencies engaged in fire protection and law enforcement activities, and providing special rules for compensatory time in lieu of overtime for certain public-sector employees).
\textsuperscript{22} Id. § 213(a)(1).
\textsuperscript{23} But see id. § 259 (providing one exception if employer has made “good faith . . . reliance on any administrative regulation, order, ruling, approval, or interpretation” of Wage and Hour Division of Department of Labor).
\textsuperscript{24} See Maryland v. Wirtz, 392 U.S. 183, 188–99 (1968).
has the power to require that state employment practices conform to the FLSA.25

Even though the FLSA permits the Department of Labor (DOL) to pursue FLSA actions on behalf of employees,26 Congress intended that aggrieved parties have the power to pursue relief under the FLSA on their own initiative.27 With regard to state employees, Congress has noted that “[s]ince the 1974 Amendments extend FLSA coverage to additional state government employees, it is now all the more necessary that employees in this category be empowered themselves to pursue vindication of their rights.”28 Indeed, the United States’ brief before the U.S. Supreme Court in Alden v. Maine29 noted that “the Department of Labor[‘s] . . . recent experience confirms Congress’s judgment that private enforcement is necessary to ensure that state employees receive the wages to which they are entitled by federal law.”30 One commentator has noted that the DOL has averaged less than fifty cases per year brought under the FLSA since 1950.31 In contrast, between 1993 and 1997 alone, plaintiffs filed more than 61,000 civil FLSA actions in federal courts.32

B. The FLSA Contemplates an Expansive Definition of Employer Liability

The use of an expansive definition as to who may be held liable for relief owed to employees arising from an FLSA violation supports the broad remedial objectives of the FLSA.33 The FLSA states that “any

25. See Auer v. Robbins, 519 U.S. 452, 457 (1997) (“In 1974 Congress extended FLSA coverage to virtually all public-sector employees . . . and in 1985 we held that this exercise of power was consistent with the Tenth Amendment.”) (citations omitted).
26. 29 U.S.C § 216(c) (“The Secretary may bring an action . . . to recover the amount of unpaid minimum wages or overtime compensation and an equal amount as liquidated damages . . . . Any sums thus recovered . . . shall be paid . . . directly to the employee or employees affected.”).
27. See S. REP. NO. 93-690, at 26–27 (1974) (“[T]he enforcement capability of the Secretary of Labor is not alone sufficient to provide redress in all or even a substantial portion of the situations where compliance is not forthcoming voluntarily.”).
28. Id at 27.
33. See, e.g., Herman v. RSR Sec. Servs. Ltd., 172 F.3d 132, 139 (2d Cir. 1999).
employer who violates the . . . [FLSA] shall be liable to the employee or employees affected in the amount of their . . . unpaid overtime compensation . . . and in an additional equal amount as liquidated damages.\textsuperscript{34} An employer is defined as “includ[ing] any person acting directly or indirectly in the interest of an employer in relation to an employee and includes a public agency.”\textsuperscript{35} Courts have construed this definition broadly,\textsuperscript{36} determining those with “managerial responsibilities” and “substantial control of the terms and conditions of the [employee’s] work” may be employers for the purpose of establishing FLSA liability.\textsuperscript{37}

I. Involvement in the Activity Giving Rise to the FLSA Violation Determines Liability, Even Where No Personal Benefit Accrues to the Individual Liable for the Violation

The critical determinant of liability for an FLSA violation is involvement in the activity that gives rise to the violation.\textsuperscript{38} The alleged employer must have sufficient control over the employees’ working conditions to ensure compliance with the FLSA.\textsuperscript{39} In \textit{Bonnette v. California Health \\& Welfare Agency},\textsuperscript{40} the Ninth Circuit established a four-part test that has been widely adopted to help evaluate whether an individual or organization may be considered an employer within the terms of the FLSA.\textsuperscript{41} Factors weighing in favor of a ruling that an

\begin{itemize}
\item 34. 29 U.S.C. § 216(b) (1994). Senator (later Justice) Hugo Black described this as “the broadest definition that has ever been included in any one act.” United States v. Rossenwasser, 323 U.S. 360, 363 n.3 (1945) (quoting 81 Cong. Rec. 7657, July 27, 1937) (statement of Senator Black)).
\item 35. 29 U.S.C. § 203(d).
\item 36. E.g., Dole v. Elliott Travel & Tours, Inc., 942 F.2d 962, 965 (6th Cir. 1991) (“The remedial purposes of the FLSA require the courts to define ‘employer’ more broadly than the term would be interpreted in traditional common law applications.”) (quoting McLaughlin v. Seafood Inc., 867 F.2d 875, 877 (5th Cir. 1989) (per curiam)).
\item 38. See, e.g., House v. Cannon Mills Co., 713 F. Supp. 159, 161 (M.D.N.C. 1988) (holding individuals accountable due to their authority over employment decisions under ADEA, relying on FLSA precedents).
\item 40. 704 F.2d 1465 (9th Cir. 1983).
\item 41. See id. at 1470 (applying four-factor test to determine that public welfare agencies were employers of in-home service providers for blind and disabled). The Second Circuit adopted this test in \textit{Herman v. RSR Security Services Ltd.}, 172 F.3d 132, 139 (2d Cir. 1999). A number of district courts have also adopted this test. See Robertson v. Bd. of County Comm’rs, 78 F. Supp. 2d 1142, 1151 (D. Colo. 1999); Baker v. Stone County, 41 F. Supp. 2d 965, 979–80 (W.D. Mo. 1999);
\end{itemize}
individual has employer status include (1) the power to hire and fire, (2) supervision and control of work schedules or conditions of employment, (3) involvement in determining the rate and method of payment, and (4) involvement in maintaining employment records. No one factor standing alone is dispositive; rather courts look to the totality of the circumstances. Further, not all factors must be present: the factors are to be treated as merely a guide to an examination of the facts of each case.

Whether the putative employer stands to benefit financially from the underpayment of the employees is not a factor to be considered. For example, within the private sector the courts distinguish ownership of an enterprise from involvement in the activity. Although there are many cases imposing individual liability on substantial shareholders of closely held corporations, a substantial ownership interest alone, without some degree of control, is insufficient to justify liability for FLSA violations. Significantly, courts have imposed individual liability against individuals with no ownership interest but with supervisory authority over employees’ working conditions.


42. Bonnette, 704 F.2d at 1470.
43. See id.
44. See RSR Sec. Servs., 172 F.3d at 140 (finding three of four factors sufficient).
45. Falk v. Brennan, 414 U.S. 190, 195 (1973) (holding that partners in real-estate-management partnership were employers of maintenance personnel despite employee's pay being determined independently of management fee going to partnership).
46. See Donovan v. Sabine Irrig. Co., 695 F.2d 190, 195 (5th Cir. 1983) (noting that stock ownership in corporate employer is not sine qua non of employer status under FLSA, and looking to control of employment relationship to determine status as employer).
47. See, e.g., United States Dep’t of Labor v. Cole Enters., Inc., 62 F.3d 775, 778 (6th Cir. 1995) (imposing liability on fifty percent shareholder and president).
48. Patel v. Wargo, 803 F.2d 632, 637–38 (11th Cir. 1986) (deciding that principal stockholder and president was not employer because he had no operational control of matters relating to employees); Wirtz v. Pure Ice Co., 322 F.2d 259, 262–63 (8th Cir. 1963) (holding that seventy-five percent shareholder who played no active role in company was not employer); see also House v. Cannon Mills Co., 713 F. Supp. 159, 161 (M.D.N.C. 1988) (“A position as an officer or shareholder is not a condition for liability, but merely an indicia of authority.”).
49. See, e.g., Schultz v. Falk, 439 F.2d 340, 344–45 (4th Cir. 1971) (holding that rental agents were employers of maintenance personnel, despite lack of ownership interest in buildings managed, and fact that agents were fully reimbursed by building owners for all wages paid to personnel); Cannon Mills, 713 F. Supp. at 161 (relying on FLSA definition of employer to hold two non-director, non-officer supervisors with no ownership interest liable as employers under Age Discrimination in Employment Act); Brock v. VAFLA Corp., 668 F. Supp. 1516, 1520 (M.D. Fla. 1987) (finding that general manager of amusement park with no apparent ownership interest is
2. **Courts Have Imposed Individual Liability on State Officials.**

Cases imposing individual liability for FLSA violations within the public sector are extremely rare. This partly reflects the more recent coverage of state employees within the terms of the FLSA\(^{50}\) and partly reflects that until the U.S. Supreme Court’s recent decisions on sovereign immunity,\(^{51}\) plaintiffs did not need to look to private parties for relief when the state’s resources were available to satisfy any judgment. However, a number of district courts have declined to dismiss claims against individual state officers who hold positions that give them responsibility for the employment practices giving rise to the alleged FLSA violations.\(^{52}\)

Only three circuit courts have had the opportunity to consider whether individual liability for FLSA violations applies to public officials. The Seventh Circuit affirmed without comment an FLSA damage award against the director of the Illinois Department of Central Management in his individual capacity.\(^{53}\) In *Lee v. Coahoma County*,\(^{54}\) the Fifth Circuit determined that a sheriff who had responsibilities for appointing

---

50. See *supra* note 18.

51. *Supra* note 7 and accompanying text.


53. *Bankston v. Illinois*, 60 F.3d 1249, 1257 (7th Cir. 1995); see also *Riordan v. Kempiners*, 831 F.2d 690, 694 (7th Cir. 1987) (suggesting that state employee could be considered “employer” under FLSA “provided the [official] had supervisory authority over the complaining employee and was responsible in whole or part for the alleged violation”).

54. 937 F.2d 220 (5th Cir. 1991).
deputies and fixing their compensation fell “within the class of managerial personnel considered employers by the FLSA”\(^{55}\) and hence could be held jointly and severally liable for damages resulting from the FLSA violation.\(^{56}\) The *Lee* court relied on U.S. Supreme Court and Fifth Circuit precedent determining the scope of employer liability in the private sector as support for its holding.\(^{57}\)

Contrary to the opinions of the Seventh and Fifth Circuits are two Eleventh Circuit panel decisions denying that the definition of “employer” used in the FLSA justifies holding individual state employees liable for violations of statutes based on the FLSA.\(^{58}\) In *Welch v. Laney*,\(^{59}\) an Eleventh Circuit panel asserted, in a single conclusory sentence supported by no authority,\(^{60}\) that the defendant sheriff was an employer in his official capacity\(^{61}\) but *not* in his individual capacity\(^{62}\) for the purposes of the Equal Pay Act, a part of the FLSA.\(^{63}\) In *Wascura v. Carver*,\(^{64}\) a subsequent Eleventh Circuit panel relied on *Welch* in dismissing an individual-capacity claim against a state official under the Family and Medical Leave Act (FMLA).\(^{65}\) The FMLA incorporates the FLSA definition as to who may be considered an employer held liable for violations of the Act.\(^{66}\) The *Wascura* court stated that “we are bound by the *Welch* decision regardless of whether

\(^{55}\) Id. at 226.

\(^{56}\) Id.

\(^{57}\) Id. at 226 (citing Falk v. Brennan, 414 U.S. 190, 194 (1973) & Donovan v. Grim Hotel, 747 F.2d 966 (5th Cir. 1984)).

\(^{58}\) Wascura v. Carver, 169 F.3d 683, 687 (11th Cir. 1999) (dismissing claims against various public officials in their individual capacities); Welch v. Laney, 57 F.3d 1004, 1011 (11th Cir. 1995) (dismissing individual-capacity claim against sheriff).

\(^{59}\) 57 F.3d 1004 (11th Cir. 1995).

\(^{60}\) Id. at 1011 (“Sheriff Laney in his individual capacity had no control over Welch’s employment and does not qualify as Welch’s employer under the Act.”).

\(^{61}\) Id. at 1010.

\(^{62}\) Id. at 1011.

\(^{63}\) The 1963 Equal Pay Act is a part of the FLSA. It amended the FLSA minimum-wage provisions to prohibit wage differentials based solely on sex. Pub. L. No.88-38, 77 Stat 56 (codified at 29 U.S.C. § 206(d) (1994)). The EPA relies on the same definition of employer as the FLSA. *Wascura*, 169 F.3d at 686 (noting that EPA is extension of FLSA and incorporates FLSA’s definition of employer).

\(^{64}\) 169 F.3d 683 (11th Cir. 1999).

\(^{65}\) Id.

\(^{66}\) See id. at 686–87.
we agree with it”67 and made no attempt to rebut the claim that the Welch decision was “unclear and inadequate.”68

Despite the split in the circuits, the definition of who may be held liable as an employer under the FLSA is sufficiently broad to warrant consideration of the doctrine of qualified immunity.

II. THE DOCTRINE OF QUALIFIED IMMUNITY

The doctrine of qualified immunity presents a potential impediment to FLSA individual-capacity suits against state officials. Qualified immunity grants limited personal immunity to state employees when sued in their individual capacities.69 “Shield[ing the official] from liability for civil damages insofar as their conduct does not violate clearly established statutory or constitutional rights of which a reasonable person would have known.”70 Qualified immunity is distinct from sovereign immunity, which attaches to states or to state employees sued in their official capacities.71 To determine whether qualified immunity protects state officials from FLSA liability it is necessary to understand the historical development and underlying policy rationales for the granting of such immunities to government officials.72

A. State Employees’ Protection from Constitutional Tort Liabilities Protec t the Public from the Consequences of Impairing Those Officials in the Performance of Their Public Duties

Qualified immunity is a personal defense available to state employees when sued in their individual capacities.73 Qualified immunity shields public officials from pretrial procedural obligations74 and from personal

67. Id. at 687.
68. Id.
71. See Hafer, 502 U.S. at 26–27 (stating that actions against state officials are not individual-capacity complaints simply because complained-of conduct was undertaken in course of their official duties); Kentucky v. Graham, 473 U.S. 159, 166–67 (1985) (noting that state’s sovereign immunity only restricts suits against public officials in their official capacities).
73. See Hafer, 502 U.S. at 25.
liability for money damages. Immunity for public servants has been justified by the desire to prevent harm to the public that may result if government officials are distracted from the effective performance of their public duties by harassing lawsuits. The defense is designed to protect public officials from undue interference with their duties, to prevent public officials from acting with an excess of caution, and to ensure that talented candidates are not deterred from public service.

The qualified immunity defense imposes an extraordinarily high burden on plaintiffs seeking civil liability for public officials accused of violating a plaintiff’s constitutional rights. Determining whether a government official has violated a constitutional right generally requires the application of a balancing test or a determination of whether the official’s acts were reasonable under the circumstances. Courts’ willingness to grant qualified immunity in civil rights cases reflects the difficulty of determining that an official knowingly violated a right defined in such ephemeral terms. As Professor Pillard notes,

because of the common-law, case-by-case method through which constitutional standards develop . . . and the high level of specificity at which the clearly-established-law inquiry is conducted, most fact-intensive constitutional claims can reasonably be characterized as new. A nonfrivolous defense based on the merits of a constitutional issue will generally suffice to support immunity.

77. Harlow, 457 U.S. at 806.
82. Pillard, supra note 80, at 79.
Understanding the development of the doctrine of qualified immunity is essential to determining the scope of its application.

B. The Historical Development of Qualified Immunity Shapes the Scope of Its Application

The U.S. Supreme Court developed the doctrine of qualified immunity for public officials in response to the explosion in the number of suits brought against state and federal officials following the Court’s expansion of liability for constitutional torts. In *Monroe v. Pape*, the Court expanded the ability of plaintiffs to enforce constitutional rights against state officials through damages actions under 42 U.S.C. § 1983. Rejecting the argument that the “under color of law” clause reached only unconstitutional acts taken by state officials when authorized by the state, the Court held that § 1983 reached unconstitutional acts taken by state officials even if they were acting without authority. Having thus exposed state officials to civil liability, the Court subsequently attempted to limit the exposure of government officials to such actions through the development of an immunity doctrine calculated to counter *Monroe*’s expansive statement of individual liability.

---

83. Crawford-El v. Britton, 93 F.3d 813, 830 (D.C. Cir. 1996) (noting that while there were only twenty-one cases decided under 42 U.S.C. § 1983 in its first fifty years, in 1988 alone more than 40,000 civil-rights actions were filed against government officials).
84. The term “constitutional torts” refers to damages actions brought against public officials in their individual capacities for the deprivation of federal constitutional rights. *See* Berry v. Funk, 146 F.3d 1003, 1013 (D.C. Cir. 1998). When brought against state officials, such actions are authorized by 42 U.S.C. § 1983 (1994). Federal agents may be sued directly under the Constitution as established by *Bivens* v. Six Unknown Named Agents of Federal Bureau of Narcotics, 403 U.S. 388, 397 (1971). Such suits have become known as *Bivens* actions.
86. *Id.* at 187 (holding that § 1983 provides cause of action against state officials). Section 1983 states: “[E]very person who, under color of any statute, ordinance, regulation, custom, or usage, of any State or Territory, subjects . . . any citizen . . . to the deprivation of any rights, privileges, or immunities secured by the Constitution and laws, shall be liable to the party injured . . . for redress.” 42 U.S.C. § 1983.
87. See *Monroe*, 365 U.S. at 172.
88. *Id.*; see also *id.* at 194–98 (Harlan, J., concurring) (explaining need for federal remedy even if unauthorized acts by state officials would constitute state tort).
Although § 1983 contains no language granting immunity to defendants, the Court developed an immunity doctrine to protect officials from § 1983 claims. The U.S. Supreme Court first addressed the existence of immunity for government officials from constitutional tort claims in *Tenney v Brandhove*. The *Tenney* Court granted immunity to members of a committee of the California State Assembly on Un-American Activities, determining that the privilege of legislative immunity was so deeply entrenched in the common law that it was impossible to believe that Congress would have intended to abrogate it in § 1983 without doing so explicitly. *Tenney* granted the legislators in question absolute immunity from suit, protecting them from civil liability irrespective of fault or motive. Such “absolute immunity” has a limited reach, only protecting officials when acting within the scope of their legislative, judicial, or prosecutorial capacities.

In contrast to absolute immunity, the Court has also recognized a form of “good faith,” or qualified immunity, a more limited form of immunity available to a broader class of government officials than are protected by *Tenney’s* absolute immunity. Relying on *Tenney*’s assertion that common law immunities survived passage of § 1983, the Court in *Pierson v. Ray* noted that the common law provided a good-faith defense for police officers accused of false arrest and held that this defense was also available to officers charged under § 1983.

Later U.S. Supreme Court decisions have extended the reach of the qualified-immunity doctrine to include a broader class of government officials.
officials. In *Scheuer v. Rhodes*, the Court expanded the class of officials able to rely on qualified immunity beyond police officers, noting that government officials with a broad range of duties face the same need for swift, firm action as police officers. *Scheuer* concerned the liability of the Governor of Ohio for acts taken during the Kent State University riots, a situation calling for rapid, intuitive judgments akin to those undertaken by police officers in the course of their duties. Unable to cite any specific historical basis for qualified immunity for officials in Sheuer’s position, the Court instead relied on the need to protect such officials from harassment.

Subsequently, in *Harlow v. Fitzgerald* the Court abandoned the subjective good-faith aspect of the qualified-immunity doctrine as developed in *Pierson* and *Scheuer*. The *Harlow* Court adopted an objective standard for evaluating a defendant’s entitlement to qualified immunity because of concerns that the subjective features of the good-faith defense were failing to prevent insubstantial claims from proceeding to trial, thus failing to free officers from time-consuming court proceedings. As a result, qualified immunity protects “government officials performing discretionary functions . . . from ‘liability for civil damages insofar as their conduct [does] not violate clearly established statutory or constitutional rights of which a reasonable person would have known.’”

---

100. See *Anderson v. Creighton*, 483 U.S. 635, 645 (1987) (noting that qualified immunity has been expanded beyond its common law scope).
102. *Id.* at 246–47.
103. *Id.*
104. *Id.* at 248.
105. *Id.* at 239–49.
107. See *id.* at 815–16.
108. See *id.* at 818.
III. **SEMINOLE TRIBE AND ALDEN GRANT THE STATES IMMUNITY FROM PRIVATE SUITS ALLEGING VIOLATIONS OF THE FLSA**

Two recent U.S. Supreme Court decisions have severely curtailed the ability of state employees to protect their rights under the FLSA.110 These controversial decisions111 held that sovereign immunity shields the states against private actions brought to enforce federal laws enacted under Congress’s Commerce Clause powers.112 These decisions reversed previous rulings as to the ability of state employees to sue states under the FLSA. Prior to 1996, the U.S. Supreme Court had established that Congress had the power to abrogate a state’s sovereign immunity from suit when acting pursuant to its Commerce Clause powers as long as Congress explicitly stated its intent to do so.113 In particular, the Commerce Clause empowered Congress to specify that employee FLSA rights may be enforced by private suit even if the defendant is a state or state agency.114

In 1996, the U.S. Supreme Court, in *Seminole Tribe v. Florida*,115 reconsidered whether the Commerce Clause afforded Congress the power to abrogate the states’ Eleventh Amendment immunity116 from suit in federal court.117 Chief Justice Rehnquist, speaking for a five-justice majority, announced that Article I of the Constitution did not confer on Congress the power to expand the jurisdiction of the federal courts over the states, overruling the Court’s *Union Gas* decision of just seven years earlier.118

111. Supra note 6.
112. *Alden*, 527 U.S. at 754 (holding that Congress lacks power to subject states to private suit in state court when legislating under its Commerce Clause powers); *Seminole Tribe*, 517 U.S. at 47 (same result in federal court).
114. See, e.g., *Hale v. Arizona*, 993 F.2d 1387, 1391–92 (9th Cir. 1993) (applying *Union Gas* to allow private FLSA action against state in federal court).
116. *U.S. Const.* amend. XI (“The Judicial power of the United States shall not be construed to extend to any suit in law or equity, commenced or prosecuted against one of the United States by Citizens of another State, or by Citizens or Subjects of any Foreign State.”).
117. See *Seminole Tribe*, 517 U.S. at 76.
118. See id. at 72–73.
Following the *Seminole Tribe* decision, a number of FLSA suits brought by state employees were dismissed from federal court on the basis of the defendant state’s Eleventh Amendment immunity.\(^{119}\) Having been denied access to the federal courts, many of the defeated plaintiffs turned to state courts for protection of their rights.\(^{120}\) The state courts disagreed as to whether the states were immune from FLSA suits in their own courts,\(^{121}\) leading the U.S. Supreme Court to grant certiorari to resolve this question.\(^{122}\) On June 23, 1999, the Court announced its decision in *Alden v. Maine*.\(^{123}\) Again by a five-to-four margin, the Court announced that “the States retain immunity from private suit in their own courts, an immunity beyond the congressional power to abrogate by Article I legislation.”\(^{124}\) Without any plain constitutional language upon which to base its extension of sovereign immunity to actions in state court, the majority contended that “the scope of States’ immunity from suit is demarcated not by the text of the [Eleventh] Amendment alone but by fundamental postulates implicit in the constitutional design.”\(^{125}\)

As a result of the *Seminole* and *Alden* decisions, state employees who believe that their state’s compensation practices violate the requirements of the FLSA may not bring a private action against the state seeking an award of back-pay or liquidated damages in either state or federal court. The *Alden* court rejected the argument that its decision to bar private actions against states would effectively confer on a state the right to disregard valid federal law.\(^{126}\) The Court identified six avenues of judicial review that are available to enforce the states’ obligation to obey validly implemented federal law in light of a series of limits implicit in the constitutional principle of state sovereign immunity.\(^{127}\)

\(^{119}\) See, e.g., *Mills v. Maine*, 118 F.3d 37, 48 (1st Cir. 1997) (deciding that FLSA was enacted pursuant to Congress’s Commerce Clause powers and therefore granting sovereign immunity to State of Maine on basis of *Seminole Tribe*).

\(^{120}\) See, e.g., *Alden v Maine*, 715 A.2d 172, 173 (Me. 1998).

\(^{121}\) Compare *id.* (holding suit barred from state court), with *Jacoby v. Ark. Dep’t of Educ.*, 962 S.W.2d 773, 778 (Ark. 1998) (determining that Eleventh Amendment does not protect states from suit for violations of FLSA in state court).


\(^{124}\) *Id.* at 754.

\(^{125}\) *Id.* at 729.

\(^{126}\) *Id.* at 754–55.

\(^{127}\) *Id.* at 755–57.
First, states may consent to suit. Second, in ratifying the Constitution, the states consented to suits brought by the federal government. Third, Congress may abrogate sovereign immunity when acting pursuant to its Fourteenth Amendment powers. Fourth, sovereign immunity does not extend to lesser entities such as “municipal corporation[s] or other government entit[ies] which are not an arm of the state.” Fifth, the Ex parte Young doctrine permits individuals to seek an injunction against officers of the state to perform a duty which they have neglected or refused to perform, so long as the relief sought is merely prospective injunctive or declaratory relief. Finally, the Court suggested that an action for damages may be brought against state officers in their individual capacities.

IV. PERSONAL-CAPACITY SUITS PROVIDE THE ONLY AVENUE FOR STATE EMPLOYEES TO PROTECT THEIR FLSA RIGHTS

If state employees wish to pursue claims for back wages on their own behalf, the only cause of action available (absent waiver by the state) is an individual-capacity suit against a state officer based on that official’s status as an employer. None of the other enforcement mechanisms suggested by the Alden Court provide a means for state employees to protect their rights under the FLSA on their own behalf, as envisioned by Congress. The Alden Court suggested that individual-capacity suits may be an available option to enforce some statutory rights, relying on Scheuer v. Rhodes and Ford Motor Co. v. Department of Treasury as support for the availability of this cause of action. However, neither Scheuer nor Ford provides a basis for concluding that this option is available under the FLSA.

128. Id. at 755.
129. Id. at 755–56.
130. Id. at 756.
131. Id.
133. See Alden, 527 U.S. at 757.
134. See id.
137. Alden, 527 U.S. at 757.
A. Excepting the Use of Individual-Capacity Suits, the Alden Court’s Proposed Enforcement Options Fail To Provide State Employees with an Enforceable Cause of Action Under the FLSA.

The first five enforcement options suggested by the Alden Court are either inapplicable to the FLSA or fail to provide state employees with the ability to seek financial compensation on their own behalf as envisioned by the FLSA. First, the states’ history of antipathy to the FLSA\textsuperscript{138} suggests that few state employees will be able to rely on their states’ voluntary waiver of sovereign immunity. Indeed, in the short period of time since the Seminole Tribe decision, twenty-two states and Puerto Rico have invoked immunity in FLSA and related actions.\textsuperscript{139} As a political question, state employees may be able to lobby for legislative action to waive reliance on sovereign immunity by their state; however, courts are bound by the political decision of the states as to this matter. Second, reliance on DOL representation is likely to be unavailing due to a lack of resources and is clearly contrary to Congress’s intent for private enforcement of the FLSA.\textsuperscript{140} Third, no court has accepted an argument that the FLSA represents an exercise of Congress’s Fourteenth Amendment powers.\textsuperscript{141} Fourth, the opportunity to bring suit against a lesser entity is of no value to any state employee who does not work for such a lesser entity. Finally, the \textit{Ex parte Young} doctrine is inapplicable to actions to enforce the FLSA because the FLSA permits only the Secretary of Labor to bring actions for injunctive relief.\textsuperscript{142}

\begin{footnotesize}
\begin{enumerate}
\item See, e.g., Mills v. Maine, 118 F.3d 37, 44–49 (1st Cir. 1997).
\item See 29 U.S.C. § 211(a) (1994); see also Lorillard v. Pons, 434 U.S. 575, 581 (1978) (holding that FLSA precludes employee’s access to injunctive relief).
\end{enumerate}
\end{footnotesize}
B. Alden Provides No Guidance as to Whether Individual-Capacity Suits Provide an Enforcement Mechanism Applicable to State Employee FLSA Rights

Because neither Scheuer nor Ford concerned suits brought under the FLSA, the Alden Court’s reliance on these decisions provides no guidance as to whether the FLSA provides a cause of action against individual state officials. Furthermore, the Court’s reliance on Scheuer raises, but does not answer, the question as to whether a state official accused of responsibility for an FLSA violation is entitled to the shield of qualified immunity.

The passage from Ford relied on by the Alden Court is dictum because no individual-capacity action was at issue. The Ford Court determined that Ford had a cause of action against the state. The specific passage in Ford relied on by the Alden Court refers to two cases concerning the recovery of unconstitutional taxes from a state tax collector. However, at the time of those decisions “taxes were collected by a revenue officer whose relation to the state was closer to that of an independent contractor than to that of an employee.” Furthermore, personal-capacity actions against state tax collectors could only proceed if the collector had not yet turned the wrongfully withheld taxes over to the state. Since no individual state official will ever personally hold the funds wrongfully withheld because of underpayment of a state employee’s wages, no analogous liability can exist for an FLSA violation.

The Alden Court’s reliance on Scheuer v. Rhodes provides no guidance either, because the cause of action in Scheuer was provided by § 1983, which establishes a cause of action against state employees in their individual capacities for constitutional violations. While § 1983 provides a cause of action for constitutional torts, it does not provide the basis for a cause of action against a state official accused of violating the

143. See Ford Motor Co. v. Dep’t of Treasury, 323 U.S. 459, 462–63 (1945).
144. See Ford, 323 U.S. at 462 (citing Atcheson, Topeka, & Santa Fe Ry. Co. v. O’Connor, 223 U.S. 280 (1912) & Matthews v. Rodgers, 284 U.S. 521 (1932)).
148. See supra notes 84–87 and accompanying text.
FLSA. Furthermore, the primary issue in Scheuer was whether the parties in question could invoke either absolute or qualified immunity from suit. If anything, this case identifies the difficulties of bringing an individual-capacity suit against a state employee, rather than suggesting that this is a viable cause of action.

V. THE FLSA PROVIDES A CAUSE OF ACTION AGAINST RESPONSIBLE STATE AGENTS AND DOES NOT PERMIT THOSE AGENTS TO INVOKE A QUALIFIED-IMMUNITY DEFENSE

If state employees are unable to prosecute supervisory personnel for FLSA violations, their ability to protect their FLSA rights will be severely compromised. In order to preserve a cause of action for state employees to bring individual-capacity suits against appropriate state supervisory personnel, state employee defendants in FLSA actions should not be entitled to rely on a qualified-immunity defense. Qualified immunity should not shield state officials from liability under the FLSA because there is no basis for applying the qualified-immunity doctrine to FLSA actions based on either accepted principles of statutory construction or on the basis of the common law underpinnings of the doctrine.

A. The History and Language of the FLSA Contemplate Individual Liability for Supervisory State Officers Involved in the Activity Giving Rise to the FLSA Violation

Permitting state employees to sue state agents in their individual capacities is consistent with the policy of holding a broad class of people liable for FLSA violations if they are involved in the decisions leading to a violation. The history of the FLSA indicates that Congress’s intent was to attach individual liability to government employees in the same

151. See supra notes 138–42 and accompanying text.
152. See supra notes 35–39 and accompanying text.
way that the statute had been applied to corporate employees.\textsuperscript{153} When Congress wishes to treat government employees accused of statutory violations differently from corporate employees, it is free to do so.\textsuperscript{154} As noted by a widely cited\textsuperscript{155} district court opinion,\textsuperscript{156} Congress had explicitly stated its intent to limit those who could be sued in federal employment discrimination cases in 42 U.S.C. § 2000e-16(c) two years before the 1974 FLSA amendments.\textsuperscript{157} Congress was aware of how to indicate that it wished government defendants to be treated differently from private-party defendants and yet chose not to do so in the FLSA. This omission suggests that Congress intended that government employees could be sued as individuals under the FLSA.\textsuperscript{158}

Furthermore, Congress has on a number of occasions recognized that special rules may be required when applying the FLSA to states as employers.\textsuperscript{159} Yet Congress chose to adopt existing definitions as to who may be held liable for FLSA violations when expanding protection to state employees.\textsuperscript{160} Congress must be presumed to have been aware of the courts’ expansive interpretation as to who could be held liable as an employer under the FLSA.\textsuperscript{161} Therefore, the decision to adopt the same definition of “employer” when extending protection to state employees

\begin{itemize}
  \item \textsuperscript{153} See H. Rep. No. 93-913, at 28 (1974), \textit{reprinted in} 1974 U.S.C.C.A.N. 2811, 2837 (“It is the intent of the committee that . . . the provisions of the law [be applied to public employees] in such a manner as to assure consistency with the . . . application . . . to other sectors of the economy.”).
  \item \textsuperscript{154} See, \textit{e.g.}, 29 U.S.C. § 207(k), (o), & (p) (1994); 29 C.F.R. § 541.5d (1999) (Special Provisions Applicable to Employees of Public Agencies).
  \item \textsuperscript{157} See \textit{id.}.
  \item \textsuperscript{158} Cf. Central Bank of Denver, N.A. v. First Interstate Bank of Denver, N.A., 511 U.S. 164, 176–77 (1994) (“Congress knew how to impose aiding and abetting liability when it chose to do so. . . . [Therefore] [i]f . . . Congress intended to impose aiding and abetting liability, we presume it would have used the words ‘aid’ and ‘abet’ in the statutory text.”); \textit{In re Haas}, 48 F.3d 1153, 1156–57 (11th Cir. 1995) (“Where Congress knows how to say something but chooses not to, its silence is controlling . . . [even if in different statutes] because Congress is presumed to be aware of pertinent, existing law when it passes legislation.”).
  \item \textsuperscript{159} See 29 U.S.C. §§ 207(k), (o), & (p); \textit{see also} 29 C.F.R. § 541.5d.
  \item \textsuperscript{160} See 29 U.S.C. §203.
  \item \textsuperscript{161} See United States v. Jordan, 915 F.2d 622, 628 (11th Cir. 1990) (“Under accepted rules of statutory construction, it is generally presumed that Congress, in drafting legislation, is aware of well-established judicial construction of . . . existing statutes.”).
\end{itemize}
implies that Congress envisioned the same expansive exposure to liability in the public sector.162

In summary, based on the history and the language of the FLSA, individual state officers who are responsible for violative employment conditions may be held liable as employers within the terms of the FLSA.

B. There Is No Basis in the Statute or Common Law for Granting State Officers a Qualified Immunity for Violations of the FLSA

The U.S. Supreme Court has provided extensive guidance as to how the qualified-immunity doctrine is to be applied to a particular statute: § 1983 of the Civil Rights Act.163 However, whether other statutes should be interpreted to admit this defense is a question that the lower courts have addressed without guidance from the U.S. Supreme Court.164 Lower courts have, for the most part, blindly assumed that the doctrine is available to a state official accused of violating any federal statute.165

162. The courts have consistently taken this view with regard to other extensions of the FLSA. For example the Equal Pay Act (EPA) was enacted as an amendment to the FLSA in 1963. Equal Pay Act of 1963, Pub. L. No. 88-38, § 3, 77 Stat. 56 (codified at 29 U.S.C. §206(d)). The EPA incorporates unchanged the definition of “employer” from § 203(d) of the FLSA. The courts have universally relied on FLSA precedent to determine interpretation of “employer” status under the EPA. See, e.g., Welch v. Laney, 57 F.3d 1004, 1011 (11th Cir. 1995) (relying on Wirtz v. Lone Star Steel Co., 405 F.2d 668 (5th Cir. 1968), which was decided under FLSA, to guide interpretation as to who may be employer under EPA); cf. Auer v. Robbins, 519 U.S. 452, 457 (1997) (noting that salary-basis test used to determine exemptions from FLSA overtime requirement was adopted in 1940 before coverage of public-sector employees, but applying test identically to public-sector employees to determine exemptions from Act.); Bryant v. Delbar Prods., 18 F. Supp. 2d 799, 807 (M.D. Tenn. 1998) (noting near identity of language of FMLA and FLSA definitions of employer to adopt interpretations of employer status from FLSA into FMLA).

163. See supra notes 69–79, 97–108 and accompanying text.

164. Every U.S. Supreme Court decision on qualified immunity has involved either a claim under § 1983 or a Bivens action. The Court recently declined to grant certiorari to consider the application of qualified immunity to the Federal Wiretap Act. See Petition For Writ of Certiorari at i, Blake v. Wright (99-848) (stating question presented as “[s]hould the common law qualified-immunity defense afforded to public officials in claims based upon constitutional rights be extended to claims arising from violations of statutory rights?”), cert. denied, 120 S. Ct. 980 (2000).

165. See Gary Gildin, Dis-Qualified Immunity for Discrimination Against the Disabled, 1999 U. ILL. L. REV. 897, 900 (1999) (“In short, the courts adjudicating damage claims under the Rehabilitation Act, [American with Disabilities Act, and Individuals with Disabilities in Education Act] have blindly cloned the § 1983 qualified immunity defense without considering whether the defense is consonant with Congress’s intent.”); see also Baker v. Stone County, 41 F. Supp. 2d 965, 1003 (W.D. Mo. 1999) (relying on fact that “no Defendant has intentionally acted to violate Plaintiffs’ rights” to grant qualified immunity to various officials in their individual capacities). But see Barfield v. Madison County, 984 F. Supp. 491, 496 n.8 (S.D. Miss. 1997) (rejecting defendant’s
This ignores the fact that the U.S. Supreme Court “look[s] to both the history and to the purposes that underlie government employee immunity” to determine the reach of the qualified-immunity doctrine. To respect both Congress’s intent with regard to the interpretation of a federal statute and the U.S. Supreme Court’s approach to qualified immunity, courts should conduct a preliminary inquiry as to whether the grant of qualified immunity is appropriate for the statute at issue. Applying qualified immunity to the FLSA is inconsistent with both Congress’s intent and with the common law basis for the doctrine.

1. There Is No Basis for Granting Qualified Immunity to Defendants in an FLSA Action Based on the Language or History of the Statute

The courts should not apply personal-immunity defenses developed in the context of constitutional torts to federal statutes without consideration of the language of the statute and congressional intent. The Court of Appeals for the District of Columbia has noted that because both § 1983 and Bivens actions were essentially Court-created actions, it was within the U.S. Supreme Court’s discretion to apply common law immunities to these common law torts. However, when the statute at issue provides its own limited defenses, there is no scope for the courts to “graft common law defenses on top of those that Congress creates.”

---

166. Richardson v. McKnight, 521 U.S. 399, 404 (1997); see also Wyatt v. Cole, 504 U.S. 158, 167–68 (1992) (declining to extend qualified immunity to private persons who conspired with state officials to violate constitutional rights, because “the nexus between private parties and the historic purposes of qualified immunity is simply too attenuated to justify such an extension of our doctrine of immunity”); Tower v. Glover, 467 U.S. 914, 920–33 (1984) (declining to extend qualified immunity to public defenders because no such immunity existed at common law).


168. See, e.g., Berry, 146 F.3d at 1013; McClelland v. McGrath, 31 F. Supp. 2d 616, 619 (N.D. Ill. 1998).

169. See Berry, 146 F.3d at 1013; see also Crawford-El v. Britton, 523 U.S. 574, 611–12 (1998) (Scalia, J., dissenting) (“We find ourselves engaged . . . [in] crafting a sensible scheme of qualified immunities for the statute we have invented—rather than . . . the statute that Congress wrote.”); Lewis & Blumoff, supra note 89, at 759 (“As it stands today, § 1983 is almost entirely a judicial construct.”).

170. Berry, 146 F.3d at 1013.
The FLSA contains both a cause of action[171] and its own enumerated defenses. Congress has included in the FLSA a grant of immunity to actors who act in “good faith . . . reliance on any administrative order, ruling, approval, or interpretation” of an administrative agency.[172] Furthermore, defendants may avoid an award of liquidated damages by establishing that they acted in good faith and on the basis of a reasonable belief that they were not violating the FLSA.[173] It is inappropriate for the courts to import common law doctrines to expand those defenses that Congress, within the terms of a statute, has provided to the cause of action it has created.[174]

Furthermore, in no congressional record of the various amendments to the FLSA extending coverage to state employees in 1966, 1974, and 1985 is there any mention that Congress wished state employers to be able to rely on any special defense not granted to private-sector employees.[175] Instead, the legislative history suggests that it was Congress’s intent to grant state employees the identical protection available to private sector employees.[176]

It is also significant that Congress has recognized the special needs of state employers in delimiting certain exceptions to the terms of the FLSA when applied to public sector employees,[177] yet no such special treatment is accorded to state officers liable in their individual capacities under the FLSA. Furthermore, Congress has, when it so desired, specified statutory defenses available to state employees not available to private-sector individuals.[178] The absence of such language in the FLSA should, therefore, be determinative of Congress’s intent.

171. See 29 U.S.C. § 216(b) (1994) (“An action to recover the liability prescribed . . . may be maintained against any employer . . . by any one or more employees . . ..”).
172. Id. § 259.
173. See id. § 260.
174. See Wyatt v. Cole, 504 U.S. 158, 164 (1992) (“Irrespective of the common law support, we will not recognize an immunity available at common law if § 1983’s history or purpose counsel against applying it in § 1983 actions.”).
175. This arguably reflects the presumption at the time that state employees would be able to enforce their rights directly against the states.
176. See supra note 153.
177. See 29 U.S.C. § 207(k), (o), & (p).
2. There Is No Common Law Basis for Granting Qualified Immunity to Defendants in FLSA Actions

Even if a court were to decide that established common law defenses to tort claims against government officials survive passage of the FLSA, unless Congress explicitly rules otherwise, there is no common law basis for the existence of immunity for a public official accused of violating the FLSA. In evaluating the reach of qualified immunity the U.S. Supreme Court looks to whether there is a specific historical application of immunity granted to the type of official in question. In addition, the Court considers whether the need to protect public officials from suit is of sufficient gravity to outweigh the rights of plaintiffs to have their rights protected. Neither test supports allowing a qualified-immunity defense to an FLSA suit.

There is no historical basis for qualified immunity for state employees who violate the FLSA. In looking to the historical basis for the common law defense, a court must consider what common law defenses were available at the time of the passage of the FLSA. The first extension of the FLSA to cover a limited number of state employees was enacted in 1966. Because this predates both Pierson and Scheuer, none of the qualified-immunity defenses subsequently established under § 1983 were available at the time of the passage of the 1966 amendments. Although the Pierson precedent has, over time, been extensively broadened to give a variety of administrative personnel protection for employment-related decisions, these extensions have not been based on any pre-1966 historical precedent. Numerous commentators have noted that qualified immunity is no longer grounded
in its historical roots, so that it is inconceivable that Congress, in 1966, could have anticipated the establishment of such a broad-based defense by subsequent U.S. Supreme Court decisions.

There is no persuasive policy argument for granting a special defense to state officers responsible for FLSA violations. In establishing the parameters of the qualified-immunity defense, the U.S. Supreme Court has balanced the interests of those who have been injured by official misconduct against the harm to the public that may result if government officials are distracted from the effective performance of their public duties by harassing lawsuits. The Court has identified two distinct threats from which it wishes to protect public employees. One is the threat to their time that would result from being forced to contest actions. The second is the financial threat that may deter those officials from carrying out their duties “with the decisiveness and the judgment required by the public good” and that may also act to discourage qualified workers from seeking employment in the public sector.

There is far less need to protect state officers from the risk of time-consuming court actions resulting from insubstantial FLSA claims than is the case for constitutional tort claims. The contrast between the legal standard required for constitutional violations and for violations of the FLSA is of great relevance in establishing the need for such protection. Determining whether a constitutional right has been violated frequently requires a court either to determine whether an official’s conduct was reasonable under the circumstances or to balance competing interests. As a result, potentially harassing accusations of constitutional violations are likely to proceed to time-consuming court actions due to the


190. See Harlow, 457 U.S. at 814.

191. See supra note 81 and accompanying text.
difficulty of evaluating such standards in pre-trial procedures. The strict liability nature of FLSA violations leads to a far greater likelihood that harassing accusations can be dealt with prior to the commencement of time-consuming litigation. The need for a qualified-immunity defense designed to act as a gatekeeper allowing “a procedural mechanism for disposing of . . . cases at the earliest possible stage of litigation . . . through summary judgment” is absent when the cause of action itself may be disposed of in this way. Hence, in weighing the need to protect state officers from harassment, the threat of insubstantial claims proceeding to the point of presenting a distraction from those employees’ abilities to undertake their public duties is greatly reduced.

The need for protection of state employees from time-consuming, insubstantial lawsuits is also less significant because the circumstances of an FLSA violation differ from those accompanying constitutional violations. Unlike the broad language of the Constitution, the FLSA represents a carefully drafted statute with an extensive history of interpretation by the courts and by the DOL. The long history of application of the FLSA to private-sector employers has ensured that many of the potential ambiguities of the law have been clarified. If an application of the FLSA is in doubt, employment administrators may request clarification from the DOL as to how the FLSA applies to specific factual situations. Further, an experienced administrative officer establishing a state’s employment practices is not subject to the stressful need to make on-the-spot decisions akin to those of the Governor of Ohio facing student riots in Scheuer or the police officers or prison officials who are the most common targets of § 1983 and

192. See supra note 23 and accompanying text.
194. Further, the fear of harassment by particularly litigious individuals is not as significant when the plaintiffs in FLSA actions will themselves be government employees, not members of the public with an ax to grind against the government. See Crawford-el v. Briton, 523 U.S. 574, 575 n.1 (1998) (referring to problems of “litigious prisoners” and “legal troublemakers”).
196. For example, Freeman Wood, the official responsible for setting the Maine probation officers’ pay and employment practices, specifically stated that he had seen DOL opinion letters relating to the issue of pay for probation officers, yet chose not to determine how these applied to the Maine probation officers. See Mills v. Maine, 853 F. Supp. 551, 554–55 (D. Me. 1994) aff’d 118 F.3d 37, 41 (1st Cir. 1997).
197. See supra notes 101–05 and accompanying text.
Bivens actions. The circumstances that have led the courts to give deference to government officials operating in such stressful environments do not apply to state administrative officials responsible for establishing a state’s employment practices. The courts’ willingness to accept that when making spur-of-the-moment decisions in stressful situations a state officer may not know the fine details of judicial interpretation of constitutional rights has no relevance to administrative officers who are responsible for violations of the FLSA.

Furthermore, the threat to a state employee’s pocketbook is overstated in the FLSA context. The widespread existence of state reimbursement statutes greatly alleviates the financial threat to state employees from individual-capacity suits. Even absent such relief from financial penalty, the need for protection of state officials from suit depends on whether the fear of being sued will “dampen the ardor . . . of public officials” so as to cause harm to the public. Absent the potential for harm to the public there is no reason for providing protection to such an official. A law enforcement officer’s timidity in stopping, interrogating, or searching a suspected criminal may result in less effective crime prevention, and ultimately in harm to the public. However, as to a public official operating in an administrative capacity determining the pay practices of state employees, the only likely effect of greater timidity is to cause

198. See Arthur Wallberg, More Than a Defense: Absolute and Qualified Immunities of State Officials Under 42 U.S.C §1983, 69 FLA. B.J. 108, 109 (1995) (noting that due to sheer number of adversarial contacts with citizenry, police are likely most common target of § 1983 actions); see also Scheuer v. Rhodes 416 U.S. 232, 244–45 (1974) (noting that police officers are “that segment of the executive branch . . . that is most frequently and intimately involved in day-to-day contacts with the citizenry, and hence, most frequently exposed to situations which can give rise to claims under § 1983”).

199. See A. Allise Burris, Comment, Qualifying Immunity in Section 1983 & Bivens Actions, 71 TEX. L. REV. 123, 175 (1992) (“The fearlessness and independence required of judges, cited in Pierson and early common law immunity decisions, do not justify the breadth of protection that has ensued.”).

200. See Vázquez, supra note 145, at 1796 n.464 (listing forty-five states with reimbursement statutes).

201. Statutory reimbursement of state employees for liabilities incurred when performing official duties does not imply that a suit against that individual becomes an official-capacity suit potentially subject to sovereign immunity. See Erwin Chemerinsky, Ability to Sue State Governments Narrowed, TRIAL, Dec. 1999, at 72, 74 (citing cases from Sixth, Seventh, and Ninth Circuits).


203. See supra notes 76, 187, and accompanying text.
greater care in ensuring compliance with the law. Congress has provided a defense to employers who rely in good faith on an administrative ruling by the DOL. Therefore, timid state administrators can protect themselves against the threat of financial liability arising from their decisions by requesting a ruling on any doubtful application of the FLSA from the DOL. As a result, at worst, a threat to a state officer’s pocketbook will result in that officer taking a more considered approach to determining pay policies and a possible increase in the number of inquiries fielded by the DOL. This burden to the public does not outweigh the rights of plaintiffs in an FLSA action.

Finally, the fear that the threat of liability may discourage qualified people from seeking public sector employment is irrelevant in the context of the FLSA. Because only public officials may be the target of constitutional tort claims, an individual considering public-sector employment can avoid exposure to constitutional torts by taking a position in the private sector. In contrast, a personnel officer cannot avoid the threat of an individual-capacity FLSA claim by working in the private sector. Hence, this concern provides no justification for providing special protection to public sector employees.

In summary, the threat to effective government that concerned the U.S. Supreme Court when developing the qualified-immunity doctrine in the context of constitutional torts is largely absent from FLSA causes of action. Hence, in balancing the interests in plaintiffs’ abilities to protect their FLSA rights against the threat to the public from exposing government officials to liability for FLSA violations, the balance favors allowing state employees to protect their FLSA rights unburdened by any additional defenses beyond those enumerated in the FLSA itself.

204. See 29 U.S.C. § 259 (1994) (providing defense for employer who has in “good faith . . . rel[i]ed on any administrative regulation, order, ruling, approval, or interpretation” of Wage and Hour Division of DOL).

205. See Harlow, 457 U.S. at 814.

206. See Joseph D. McCann, The Interrelationship of Immunity and the Prima Facie Case in Section 1983 and Bivens Actions, 21 GONZ. L. REV. 117, 117–18 (1985) (noting that § 1983 actions provide remedies against state or municipal officials and that Bivens provides remedies against federal officials, but private parties can be liable only for conspiring with such officials).

207. See supra note 46–49 and accompanying text.
CONCLUSION

As a result of the U.S. Supreme Court’s recent reinterpretation of the reach of state sovereign immunity, state employees are prevented from seeking restitution for wages wrongfully withheld due to having been paid in violation of the FLSA. For these employees to be able to protect their own rights as intended by Congress, the only form of relief available is for the wronged state employees to bring an individual-capacity suit against state agents responsible for establishing the violative pay practices. Individual-capacity suits against appropriate state agents can provide an effective means for state employees to protect their rights, ensuring that the fate that befell John Alden and his fellow probation officers need not be suffered by other state employees. However, the courts must recognize that qualified immunity is an inappropriate doctrine to apply to protect individual state officers from liability for FLSA violations if state employees are to be afforded an opportunity to defend their rights as Congress intended.
ROBERTS v. DUDLEY: AN UNNECESSARY BROADENING OF THE PUBLIC POLICY EXCEPTION TO THE EMPLOYMENT-AT-WILL DOCTRINE IN WASHINGTON

Brian Hersey

Abstract: In Roberts v. Dudley, the Supreme Court of Washington dramatically expanded the previously narrow public policy exception to the employment-at-will doctrine and created a dangerous precedent. The court held that small employers, explicitly exempt from the Washington Law Against Discrimination (WLAD), could be liable at common law for the tort of wrongful discharge in violation of Washington’s public policy against sex discrimination as found in the WLAD. The tort of wrongful discharge in violation of public policy requires a finding of a “clear mandate of public policy.” This Note argues the court should not have found in the WLAD a mandate of public policy sufficiently clear to support the common law tort claim. The court disregarded the plain language of the WLAD, the relevant legislative history, and its own precedent. In dissent, Justice Madsen identified flaws in the majority’s reasoning, but failed to acknowledge that the WLAD’s small-employer exemption cannot block a common law claim based outside of the WLAD. Therefore, although the court’s ultimate conclusion was correct, its use of the WLAD to support an action against an employer expressly exempted by the WLAD threatens other exempted groups and invites the usurpation of legislative power.

In 1996, the Supreme Court of Washington held that Sharon Griffin, a former legal secretary, could not sue her employer under the Washington Law Against Discrimination (WLAD) after he discharged her on the basis of her sex. Ms. Griffin had alleged that her employer subjected her to a daily barrage of lewd comments regarding women in general and herself in particular. Nevertheless, the court found the WLAD did not apply because her employer had fewer than eight employees. Four years later, in Roberts v. Dudley, the Supreme Court of Washington held that Lynne Roberts, a former veterinary-clinic employee, could sue her former employer in tort for wrongful discharge in violation of public policy after he discharged her on the basis of her sex, even though he employed fewer than eight employees. Ms. Roberts argued that the WLAD, a recent Supreme Court of Washington case,  

1. WASH. REV. CODE §§ 49.60.010–.410 (2000).
3. Id. at 62, 922 P.2d at 789.
4. Id. at 61, 922 P.2d at 789.
5. 140 Wash. 2d 58, 993 P.2d 901 (2000).
6. Id. at 77, 993 P.2d at 910.
and Washington’s equal opportunity statute, all articulated a public policy against sex discrimination in employment. The Roberts court held that each of her three proffered reasons independently manifested the sufficiently clear mandate of public policy required for the common law tort action. It ruled that, although plaintiffs may not pursue claims against employers with fewer than eight employees under the WLAD, the WLAD’s underlying policy against sex discrimination could support a common law tort claim.

This Note contends that the Roberts court erred when it found, in a statute explicitly excluding the defendant, a public policy sufficient to support a claim for wrongful discharge in violation of public policy. Part I of this Note introduces the doctrine of employment at will and the common law public policy exception. Part I also addresses the “clear mandate of public policy” necessary for a tort of wrongful discharge in violation of public policy. Part II discusses Washington’s laws against sex discrimination, including the equal opportunity statute, and the history behind the development of the WLAD. It focuses on the limits to the WLAD’s application found in its definitions and interpretive case law. Part III examines both the majority and dissenting opinions in Roberts. Part IV argues that the Roberts majority erred by finding the WLAD evidences a clear mandate of public policy against sex discrimination by an employer with fewer than eight employees. The court’s reasoning threatens other exempted groups as well as the separation of legislative and judicial power. Part IV also critiques the dissent for disregarding clear mandates of public policy found outside the WLAD.

I. EMPLOYMENT AT WILL AND THE PUBLIC POLICY EXCEPTION

For more than a century, an employment contract of indefinite duration has been presumed to be terminable at will by either party with or without cause. However, courts and legislatures have tempered the harshness of this rule. One of the judicial limitations of the doctrine is the

8. Women May Pursue Any Calling Open to Men, 1890 Wash. Laws 519–20, § 1 (codified at WASH. REV. CODE § 49.12.200 (2000)). This Note will refer to this statute as the Washington equal opportunity statute.
9. Roberts, 140 Wash. 2d at 77, 993 P.2d at 911.
10. See id.
“public policy exception,” a common law tort action barring discharges that violate a public policy. Washington courts require plaintiffs to identify a “clear mandate of public policy” in statutes, case law, or either the United States or Washington Constitutions.11 When identified in a statute, the purported public policy must not exceed the scope of that statute.

A. The Doctrine of Employment At Will

Although employment at will has been the American default rule for employer-employee relations for more than 120 years,12 both the judiciary and legislatures have tempered it. Under this doctrine, either party may terminate indefinite-term employment contracts at any time and for any or no reason.13 The parties are free to modify their relationship by written or oral contract.14 Until the 1930s, courts struck down attempts to modify this doctrine on the grounds that interfering with the employer-employee relationship would unconstitutionally impede employers’ and employees’ freedom to contract.15 Congress first began to regulate employer-employee relationships through labor legislation, giving American workers the right to organize unions and bargain collectively.16 Several decades later, Congress limited at-will employment directly by prohibiting discriminatory discharges in the Civil Rights Act of 1964.17 Concurrent with federal legislation, the Washington Legislature has taken steps to address workplace discrimination by enacting anti-discrimination laws.18

12. See id. at 225, 685 P.2d at 1085 (citing HORACE WOOD, MASTER AND SERVANT 134 (2d ed. 1886)). This doctrine has been the rule in Washington since at least 1928. See Davidson v. Mackall-Paine Veneer Co., 149 Wash. 685, 688, 271 P. 878, 879 (1928). Today, an employment contract of indefinite duration is still presumed to be at will. See Roberts, 140 Wash. 2d at 63, 993 P.2d at 904 (citing Roberts v. Atl. Richfield Co., 88 Wash. 2d 887, 894, 568 P.2d 764, 768–69 (1977)).
13. ARTHUR LARSON & LEX K. LARS, EMPLOYMENT DISCRIMINATION § 1.01, at 1-4 (1994).
14. See Thompson, 102 Wash. 2d at 227–28, 685 P.2d at 1086–87 (noting that employers and employees may bargain for any terms and that any intrusion into employment contracts is best left to legislative process); see also RICHARD A. POSNER, SEX AND REASON 253 (1992) (“Rarely is a person made better off by having an option removed.”).
16. LARSON & LARS, supra note 13, § 1.01, at 1-4.
18. See infra Part II.C.
B. Wrongful Discharge in Violation of Public Policy: A Tort Exception to the At-Will Doctrine

State courts nationwide have used tort law to limit the at-will doctrine by holding employers liable for discharging employees on grounds that violate public policy. Washington adopted this exception in Thompson v. St. Regis Paper Co. Since Thompson, the judiciary has refined and clarified what constitutes a clear mandate of public policy sufficient to support this common law tort action.

1. The Development of the Public Policy Exception to Employment At Will

By creating a common law tort for wrongful discharge in violation of public policy, state judiciaries curbed the at-will doctrine’s most troubling aspect: the ability of employers to discharge employees for reasons that run counter to the public welfare.20 In Petermann v. International Brotherhood of Teamsters,21 the Supreme Court of California paved the way by establishing an exception to employment at will when it held that considerations of public policy can limit an employer’s right to discharge.22 The Petermann defendant had discharged the plaintiff for refusing to testify falsely before a legislative committee.23 The court reasoned that allowing an employer to condition continued employment on the commission of a felony would injure the public welfare.24 This cause of action for wrongful discharge in violation of public policy became known as the “public policy exception” to the doctrine of employment at will.25 It prevents employers from discharging employees when the reason for the discharge is inconsistent with a public policy and

22. Id. at 27.
23. Id. at 26.
24. Id. at 27.
the community interests that policy advances. 26 As of 1994, all but six states had adopted some form of this exception. 27

2. Washington’s Adoption of the Public Policy Exception

In 1984, the Supreme Court of Washington adopted the public policy exception in Thompson v. St. Regis Paper Co. 28 Thompson involved a dismissal resulting from an employee’s attempt to institute federally mandated accounting procedures. 29 The court held that discharging an employee for obeying a statutory mandate violated the public policy against bribery and thus gave rise to a cause of action in tort. 30 The court unanimously joined “the growing majority of other jurisdictions” in recognizing a cause of action for wrongful discharge in violation of public policy. 31

By recognizing a tort claim for wrongful discharge in violation of public policy, the Thompson court did not disavow the doctrine of employment at will, but rather balanced the interests of employers and employees. 32 The court recognized employers must maintain the ability to make non-discriminatory personnel decisions to run a business effectively. 33 On the other hand, employers should not condition continued employment on the commission of illegal acts, which damage the public welfare. 34 To maintain this balance the court created a test that places the burden of proof on an employee claiming wrongful discharge. 35 A plaintiff must first demonstrate the existence of a “clear mandate of public policy found in either the letter or purpose of a constitutional, statutory, or regulatory provision or scheme,” 36 or a policy


27. MARK A. ROTHEINSTEIN ET AL., EMPLOYMENT LAW § 9.9 (1994); see also Greeley v. Miami Valley Maint. Contractors, 551 N.E.2d 981, 986 n.3 (Ohio 1990) (noting at least thirty-nine states have adopted public policy exception).


29. Id.

30. Id. at 234, 685 P.2d at 1090.

31. Id. at 232, 685 P.2d at 1089.

32. Id. at 232–33, 685 P.2d at 1089.

33. Id.

34. See id.

35. See id.

36. Dicomes v. State, 113 Wash. 2d 612, 617, 782 P.2d 1002, 1006 (1989). The court noted that the employer’s conduct need not violate the law to constitute a violation of public policy. Id. at 620,
found in a prior judicial decision. Next, the plaintiff must demonstrate why the challenged discharge violates that policy. If the employee establishes a presumption of wrongful discharge by meeting this burden, the employer can rebut the presumption by providing a credible, non-discriminatory explanation for the discharge. Since Thompson, Washington courts have gradually refined the rules for determining when plaintiffs have met their burdens.

C. Judicial Rules for Finding in Statutes a Clear Mandate of Public Policy Sufficient To Support a Claim of Wrongful Discharge

Washington courts have found clear mandates of public policy where an employer discharges an employee for refusing to commit an act contrary to the public welfare, for exercising a right, or for refusing to neglect a public duty. Acts contrary to the public welfare include outright illegal acts and not reporting an employer’s wrongful act. Public duties include responsibilities such as voting and jury duty. Washington courts have also held employers liable for wrongful discharge in violation of public policy where a statute has expressly granted employees the right to engage in concerted action, file worker’s compensation claims, or complain to officials when denied overtime pay.

The Thompson court emphasized the narrowness of the public policy exception. Washington courts have sought to maintain this narrow scope by refusing to extend public policies beyond the explicit terms of

---

782 P.2d at 1008. However, in the whistle-blowing context, the court decided that employee conduct motivated purely by private interest cannot form the basis for a charge of violation of public policy. Id. at 620, 782 P.2d at 1008.

37. Thompson, 102 Wash. 2d at 232, 685 P.2d at 1089.
38. Id.
39. Id. at 232–233, 685 P.2d at 1089.
40. See Gardner v. Loomis Armored, Inc., 128 Wash. 2d 931, 936, 913 P.2d 377, 379 (1996). Generally, these rights and duties arise from statutes or judicial decisions not directly regulating employment relationships. In Thompson, for example, the Foreign Corrupt Practices Act regulated accounting practices rather than employment relationships. 102 Wash. 2d at 234, 685 P.2d at 1090.
41. See Gardner, 128 Wash. 2d at 936, 913 P.2d at 379.
45. See Thompson, 102 Wash. 2d at 232, 685 P.2d at 1089.
the statutes in which courts find these policies. Two cases illustrate plaintiffs’ difficulties in persuading courts to recognize public policies expanded beyond explicit statutory terms. In *Roe v. Quality Transportation Services*, the court of appeals refused to use a statute regulating a specific aspect of employment relations to find a public policy sufficient to support a claim for wrongful discharge based on a different aspect of employment relations not covered by the statute. In *Gardner v. Loomis Armored, Inc.*, the Supreme Court of Washington limited a statute’s policy to the circumstances the statute regulates.

In *Roe*, the court of appeals refused to recognize a broad right to privacy sufficient to constitute the clear mandate of public policy necessary to support a claim for wrongful discharge. The court found no public policy prohibiting employers from requiring drug testing. The employee alleged her dismissal constituted a violation of Washington’s public policy of protecting private employees’ privacy, a policy manifested in a statute prohibiting employers from requiring any current or prospective employer to take a lie detector test. The court refused to find a clear mandate of policy protecting employees’ broad privacy rights in a statute that regulated only lie detector tests. It noted that the Legislature, although aware of the issue of drug testing, had remained silent on the issue. The court reasoned that the judiciary, unable to hold hearings and debates, was ill-suited to balance properly the interests of employees

---

48. Id. at 610, 838 P.2d at 132.
50. Id. at 942–43, 913 P.2d at 383.
52. Id. at 605, 838 P.2d at 129.
53. Id. at 609, 838 P.2d at 131 (citing WASH. REV. CODE § 49.44.120 (2000)).
54. See id. In *Twigg v. Hercules Corp.*, 406 S.E.2d 52 (W. Va. 1990), the Supreme Court of West Virginia held a similar statute barring lie detector tests would also bar drug tests in most circumstances. See 406 S.E.2d at 55. The *Roe* court distinguished the standards for the public policy exception, noting that the West Virginia standard called only for a “substantial” mandate of public policy as opposed to Washington’s “clear” mandate requirement. *Roe*, 67 Wash. App. at 609, 838 P.2d at 131.
and employers with regard to drug testing. 56 The Roe court also rejected the Washington Constitution’s privacy provision as a source of public policy. 57 The court found this provision was a restraint on government action and not actions by private individuals. 58

Similarly, in Gardner, then regarded as the Supreme Court of Washington’s broadest application of the public policy exception, 59 the court remained cautious in inferring a clear mandate of public policy from general statutory language. The defendant employer discharged a guard for abandoning his armored truck, in violation of company policy, to rescue a woman from an assailant. 60 Although the court ultimately found a clear mandate of public policy allowing employees to neglect their job duties to protect human life, 61 it rejected two of the plaintiff’s proposed public policies. First, the plaintiff argued that the Crime Victims, Survivors, and Witnesses Act; 62 a statute granting limited immunity to citizens assisting law enforcement officials; 63 and a statute making it a misdemeanor to refuse an officer’s request to summon aid 64 all mandated a public policy encouraging civilians to assist law enforcement officers. 65 Although the court agreed there was a “limited,
albeit clear, public policy” in favor of civilians aiding police officers, the
court narrowly circumscribed this policy by noting that it did not imply
that civilians should “jump into the middle of every criminal situation.”66
Second, the court rejected the plaintiff’s claim that his discharge violated
the rescue doctrine’s policy of helping those in need.67 The rescue doc-
trine, which is codified in Washington,68 allows an individual injured in
the course of rescuing another to recover against the person rescued
when the rescuer’s own negligence created the situation.69 The court held
the doctrine’s public policy applied only to those who render emergency
care or transportation to persons whose own negligence caused the
situation.70 Although the court recognized that the rescue-doctrine statute
supported a broad public policy in favor of aiding others, it held that the
legisature limited the application of that policy to the circumstances of
the statute.71 Therefore, because the employer did not discharge the
plaintiff for actions specifically protected by these statutes, the court held
the discharge was not wrongful because the plaintiff had failed to show
that the discharge violated a sufficiently clear mandate of public policy.72
Thus, under Thompson, a plaintiff must prove not only a public policy’s
existence, but also that the statute in which the policy is found protects
the particular action for which the plaintiff was discharged.

II. WASHINGTON LAWS AGAINST SEX DISCRIMINATION:
THE EQUAL OPPORTUNITY STATUTE, THE EQUAL
RIGHTS AMENDMENT, AND THE WASHINGTON LAW
AGAINST DISCRIMINATION

Washington has a long-standing, well-established public policy
against sex discrimination. This policy dates back more than one hundred
years and is found in both statutory law and the Washington
Constitution. Today, the Washington Law Against Discrimination
(WLAD)73 deals with the bulk of sex-discrimination claims. Although

66. Id.
67. Id. at 943, 913 P.2d at 383.
69. See Gardner, 128 Wash. 2d at 943, 913 P.2d at 383.
70. See id. at 943–44, 913 P.2d at 383.
71. See id.
72. See id.
73. WASH. REV. CODE §§ 49.60.010–.410 (2000).
the WLAD seeks to eradicate employment discrimination in Washington, the Legislature has limited its scope. The Supreme Court of Washington has recognized the WLAD’s explicit exclusion of employers employing fewer than eight employees (small employers).

A. Washington’s Policy Against Gender-Based Discrimination: The Equal Opportunity Statute and the Equal Rights Amendment

Washington first prohibited sex-based employment discrimination in 1890, when the Washington State Legislature enacted a statute entitled “Women May Pursue Any Calling Open to Men,” which will be subsequently referred to as the equal opportunity statute.74 The equal opportunity statute mandates the doors of employment be equally open to both sexes.75 Located in a chapter of the Washington Revised Code that deals with private industrial regulation, this statute applies to Washington’s private as well as public employers.76

Case law interpreting the equal opportunity statute, albeit sparse, evidences a long-standing intolerance of sex-based discrimination. In 1893, the Supreme Court of Washington held that the equal opportunity statute prohibited a local practice of shutting down, as a “nuisance,” any saloon where women were employed.77 In 1971, the court of appeals noted that a local ordinance prohibiting “massagists” from giving massages to members of the opposite sex did not violate the statute because it regulated all members of a profession without regard to sex.78 Judicial interpretations of the equal opportunity statute are sparse because for much of its existence, social, if not legal, constraints have limited women’s employment opportunities.79 In addition, since women

74. 1890 Wash. Laws 519–20, § 1 (codified at WASH. REV. CODE § 49.12.200). Public office was originally exempt from the requirement of equal opportunity. Id. In 1963, the Legislature eliminated this exception and amended the act to prohibit the exclusion of women from work sites on the basis of sex. See 1963 Wash. Laws 229, § 1 (codified at WASH. REV. CODE § 49.12.200).
76. See id. (citing WASH. REV. CODE §§ 49.12.005–902).
began more aggressively enforcing their equal opportunity rights, most have been able to pursue discrimination claims under the WLAD.

In 1972, the Washington Legislature and the voters of Washington adopted the Equal Rights Amendment (ERA) to the state constitution. The amendment reads: “Equality of rights and responsibility under the law shall not be denied or abridged on account of sex.” Washington courts have acknowledged the ERA absolutely prohibits discrimination on the basis of sex. Thus, Washington’s constitution, the bedrock of Washington law, clearly condemns sex discrimination.

B. The Washington Law Against Discrimination

1. The History, Coverage, and Enforcement of the WLAD

When the Washington State Legislature enacted the WLAD in 1949, it declared that practices of discrimination constitute a threat not just to state residents, but also to “the institutions and foundation of a free democratic state.” The Legislature did not limit the WLAD’s coverage to employment discrimination but included discrimination in credit and insurance transactions in public places. The Legislature also declared the “right to obtain and hold employment without discrimination” to be a civil right. The WLAD created a state agency, the Washington Human Rights Commission (WHRC), with power to eliminate and prevent discrimination in employment and to enforce the civil rights provisions of the state constitution.

The WLAD neither precludes any other cause of action nor repeals any provision of other Washington law; in fact, it expressly preserves pre-existing law. Furthermore, the WLAD contains a clause mandating

80. See id. at 1795–99.
81. WASH. CONST. art. XXXI (1972) (adopted in Novemver 1972 election as Amend. 61).
82. Id. at § 1.
84. 1949 Wash. Laws 183, § 1 (codified as amended at WASH. REV. CODE § 49.60.010 (2000)).
85. WASH. REV. CODE § 49.60.010 (2000).
86. 1949 Wash. Laws 183, § 2 (codified as amended at WASH. REV. CODE § 49.60.030(1)(a)).
88. WASH. REV. CODE § 49.60.020. This provision contrasts with the worker’s compensation statute, which explicitly precludes other actions. See WASH. REV. CODE § 51.04.010 (2000).
that courts liberally construe the law to achieve its purposes. 

For example, in *Marquis v. Spokane*, a case in which a female golf professional alleged her employer discriminated against her on the basis of sex, the Supreme Court of Washington broadly construed the WLAD’s definition of “employee” to include independent contractors. Although the WLAD does not expressly address independent contractors, the court reasoned that the WLAD’s purpose of deterring and eradicating discrimination was of paramount importance and justified construing the statutory term “employee” to include independent contractors.

Since 1949, the WLAD’s coverage has gradually expanded. Originally, the WLAD made it an unfair employment practice to “discharge or bar any person from employment because of such person’s race, creed, color or national origin.” In 1971, the Legislature added “sex” to the list of protected classes. Initially, the WHRC had exclusive power to enforce the WLAD. Yet the only remedies available to the WHRC were “conference, conciliation and persuasion,” cease-and-
desist orders requiring an employer to halt the unfair employment practice, and individual restraining orders. 98 In 1973, the Legislature amended the WLAD to allow an individual employee injured by an employer’s violation of the WLAD to sue for civil damages. 99 The Legislature authorized specific remedies including actual damages, attorneys’ fees, and other remedies authorized by Title VII of the federal Civil Rights Act of 1964. 100

2. The Limits of the WLAD’s Reach

In enacting the WLAD, the Legislature sought to balance the interests of employers, employees, and society. 101 For example, to protect religious freedom, the definition of employer excludes religious organizations. 102 In addition, the Legislature excluded domestic servants and employees who are either parents or children of their employer. 103 These exemptions reflect the Legislature’s unwillingness to interfere in matters as personal as family relationships or who may enter one’s home. 104 The WLAD also protects small employers’ interests by defining “employers” as those who employ eight or more persons. 105

---

100. Id. Until the 1991 Civil Rights Act, the remedies available under Title VII were limited to back pay and reinstatement. See Lissau v. S. Food Serv., Inc., 159 F.3d 177, 180–81 (4th Cir. 1998). The 1991 provision created a sliding scale ranging from a maximum of $50,000 for employers with between 14 and 101 employees and up to $300,000 for employers with more than 500 employees. 42 U.S.C. § 1981a(b)(3) (1994). For a more detailed history of the statutory development of the WLAD, see Darya V. Swingle, Comment, Personal Liability for Sexual Harassment Under the Washington Law Against Discrimination, 74 WASH. L. REV. 483 (1999).


103. WASH. REV. CODE § 49.60.040(4) (2000).


105. WASH. REV. CODE § 49.60.040(3).
The definition of employer has not changed since 1949 and no legislative history explains the intent behind the exclusion of employers with fewer than eight employees. Recently, two legislative attempts to reconsider the small-employer exemption failed. In 1999, the Legislature proposed broadening the definition of employer to include anyone with one or more employees and adding a positive declaration that all employees are entitled to a workplace free from discrimination. This proposal died in committee. Another proposal would have created a task force to study and make recommendations regarding employment discrimination. This proposal passed the Senate but died in the House Judiciary Committee. Thus, the Legislature has thus far guarded the exemption. The administrative regulations for the WLAD suggest this exemption seeks to protect small employers from the burden of defending discrimination suits and to reduce the WHRC’s caseload. Regardless of the policy underlying the WLAD’s small-employer exemption, the Legislature has never eliminated or changed this minimum number.

To help determine the legislative intent behind the small-employer exemption, the Supreme Court of Washington has looked to the history of the California Fair Employment Practices Act (FEPA), which

106. Compare WASH. REV. CODE § 49.60.040(4), with 1949 Wash. Laws 183, § 3(b), at 507 (codified as amended at WASH. REV. CODE § 49.60.040).
110. See 1 Legislative Digest and History of Bills, 56th Leg. Reg. Sess. 51 (Wash. 1999).
111. See S.B. Rep. E.S.B. 5337, 56th Leg. Reg. Sess. 1–2 (Wash. 1999) (reporting that both E.S.B. 5337 and S.B. 5130 were direct responses to court holding that small employers could not be sued under WLAD in Griffin v. Eller, 130 Wash. 2d 58, 922 P.2d 788 (1996)); see also infra Part II.B.3.
113. The WHRC regulations state the exemptions exist “[t]o relieve small businesses of a regulatory burden” and “[i]n the interest of cost effectiveness, to confine public agency enforcement of the law to employers whose practices affect a significant number of persons.” WASH. ADMIN. CODE § 162-16-160(2) (2000). The Supreme Court of Washington has also noted that the guidelines, though persuasive, are of limited value since they were written thirty-three years after the statute’s enactment. See Griffin, 130 Wash. 2d at 69, 922 P.2d at 792.
114. See 1 Legislative Digest and History of Bills, 56th Leg. Reg. Sess. 51; see also Roberts v. Dudley, 140 Wash. 2d 58, 84, 993 P.2d 901, 915 (2000) (Madsen J., dissenting).
115. CAL. GOV’T CODE § 12900 (West 1999).
closely parallels the WLAD. The California State Assembly exempted employers with fewer than five employees from FEPA for four reasons: (1) because employers should have some small measure of the “so-called freedom to discriminate”; (2) because discrimination on a small scale would be hard to detect and police; (3) because employment situations with five or fewer employees make for close personal relationships, and fair-employment law should not attempt to regulate these relationships; and (4) because the Assembly was primarily interested in eliminating large-scale discrimination by significant employers rather than redressing individual instances of discrimination. This reasoning reflects the California State Assembly’s balancing of community interests and public policy in enacting an employment discrimination statute. The Supreme Court of Washington presumed that the Washington Legislature had similar considerations in mind when it included a small-employer exemption in the WLAD.

3. **Griffin v. Eller: The Small-Employer Exemption Survived a Constitutional Challenge and the Court Discussed Its Purpose**

In *Griffin v. Eller*, the small-employer exemption survived a constitutional challenge when the Supreme Court of Washington held that the exemption did not violate the privileges and immunities clause of the Washington Constitution. The court affirmed the dismissal of the plaintiff’s WLAD claims, holding that no statutory cause of action existed where the statute expressly excluded employers who employed fewer than eight people. The court based its decision on the plain language of the statute as well as previous decisions in which small employers and religious organizations had been held exempt from

---

116. See *Griffin*, 130 Wash. 2d at 67 n.1, 922 P.2d at 791 n.1.
118. See *Robinson*, 825 P.2d at 775 (Cal. 1992).
119. See *Griffin*, 130 Wash. 2d at 66–67, 922 P.2d at 791.
120. See *Griffin*, 130 Wash. 2d at 58, 922 P.2d 788 (1996).
121. WASH. CONST. art. 1 § 12.
122. *Griffin*, 130 Wash. 2d at 61, 922 P.2d at 789. In dissent, Justice Talmadge argued that the definition of employer was intended to be applied only to actions initiated by the WHRC, thereby permitting a private action against employers with fewer than eight employees. *Id.* at 72–97, 922 P.2d at 794–806 (Talmadge, J., dissenting).
statutory remedies. In dicta, the Griffin court hypothesized that the Legislature might have sought to conserve state resources or shield small businesses from regulatory burdens and private-litigation expenses. It reasoned that these concerns provided a rational basis for distinguishing employers by size; therefore, the WLAD did not violate the privileges and immunities clause.

III. THE COLLISION OF THE WLAD AND THE COMMON LAW TORT OF WRONGFUL DISCHARGE IN VIOLATION OF PUBLIC POLICY IN ROBERTS v. DUDLEY

In Roberts v. Dudley, the Supreme Court of Washington confronted the question of whether a small employer, exempted from the WLAD, could be liable in tort for wrongful discharge in violation of a public policy against sex discrimination. The majority held that a discharge on the basis of sex violated Washington’s public policy against sex discrimination as articulated in the equal opportunity statute, Marquis v. City of Spokane, and the WLAD. In a vigorous dissent, Justice Madsen argued the WLAD’s small-employer exemption muted the clarity of the public policy against discrimination where the defendant had fewer than eight employees; thus, no “clear mandate of public policy” existed in Washington law.

A. Facts and Procedural History

In 1993, Dr. Eric Dudley discharged Lynne Roberts from North End Veterinary Clinic in Tacoma following a four-month maternity leave.
He stated he discharged her due to an “economic slowdown.” One year later, Dr. Dudley advertised for Ms. Roberts’s former position; when she applied, he refused to rehire her. Ms. Roberts, alleging her pregnancy led to her discharge, brought a common law claim against Dr. Dudley for wrongful discharge in violation of Washington’s public policy against sex discrimination.

Dr. Dudley moved for summary judgment, arguing no cause of action existed under the WLAD or any other state law because he had fewer than eight employees. He argued that Griffin’s discussion of possible legislative purposes for the WLAD’s small-employer exemption constituted a judicial acknowledgement of a public policy protecting small employers from discrimination claims. The trial court agreed and granted summary judgment. The court of appeals reversed. It found that sex discrimination, even by small employers like Dudley, violated a clear mandate of public policy. The court held the WLAD, which explicitly preserves all other causes of action, did not preempt a common law cause of action for wrongful discharge in violation of public policy. The court declared that the ERA, the WLAD, and Marquis v. City of Spokane each provided clear mandates of Washington’s public policy against sex discrimination. Accordingly, for the purpose of a common law tort claim, the court held that the small-employer exemption had no effect. The Supreme Court of Washington affirmed the appeals court’s decision.
B. The Supreme Court of Washington’s Holding in Roberts v. Dudley

The Roberts majority held that Washington has a public policy against sex discrimination, mandated in statutes and judicial precedent, that applies to all employers regardless of the number of employees.148 With little analysis, the majority held that both Marquis v. City of Spokane149 and the equal opportunity statute150 established clear mandates of public policy sufficient to support an action in tort for wrongful discharge.151 In addition, the court found that the WLAD independently constituted a clear mandate of public policy that applied to all employers, regardless of size.152 Notwithstanding contrary dicta in Griffin153 and the explicit language of the small-employer exemption,154 the court refused to find a public policy shielding employers with fewer than eight employees from common law claims in the WLAD’s definition section.155

The Roberts majority held that Marquis, which summarized the purpose of the WLAD,156 constituted a clear judicial mandate of public policy prohibiting discrimination by employers irrespective of the number of employees.157 The Marquis court found a clear mandate of public policy against sex discrimination in the legislative purpose of deterring and eradicating discrimination,158 WLAD’s liberal construction provision,159 and the ERA.160 Thus, the Roberts majority found in Marquis a clear mandate of public policy against discrimination.161 In addition to Marquis, the Roberts majority found the equal opportunity statute, based on its plain language and context, constituted a clear

148. See id. at 77, 993 P.2d at 911.
149. 130 Wash. 2d 97, 922 P.2d 43 (1996).
151. Roberts, 140 Wash. 2d at 66–68, 993 P.2d at 906–07.
152. Id. at 68–73, 993 P.2d at 907–09.
154. See WASH. REV. CODE § 49.60.040(3).
157. Roberts, 140 Wash. 2d at 66, 993 P.2d at 906.
158. Id. at 108, 922 P.2d at 49; see also Marquis, 130 Wash. 2d at 109, 922 P.2d at 49.
159. WASH. REV. CODE § 49.60.020; see also supra note 104 and accompanying text.
160. Marquis, 130 Wash. 2d at 109, 922 P.2d at 49.
161. Roberts, 140 Wash. 2d at 66, 993 P.2d at 906; see also supra notes 90–93 and accompanying text.
mandate of public policy against sex discrimination sufficient to support a common law tort claim for wrongful discharge.\(^{162}\)

The court then reasoned that the WLAD expressed an independent, clear mandate of Washington’s public policy against sex discrimination in employment.\(^{163}\) It rejected the defendant’s argument that it was illogical to allow a common law claim when the public policy is found in a statute specifically exempting the defendant employer.\(^{164}\) The majority held the small-employer exemption served merely to limit the statutory remedies.\(^{165}\) Based on the language of the purpose\(^{166}\) and civil rights sections of the statute,\(^{167}\) the majority found the WLAD’s public policy to be broader than those remedies.\(^{168}\) Therefore, it held that even employers who cannot be held liable under the WLAD can be held liable for violating its policy in a common law tort action for wrongful discharge.\(^{169}\)

To support its conclusion, the court cited *Bennett v. Hardy*,\(^{170}\) a case holding that the WLAD’s definition of employer did not apply outside the context of the statute.\(^{171}\) In *Bennett*, two sisters sued their former employer, who had never employed eight or more employees at any one time.\(^{172}\) The sisters alleged two counts of wrongful discharge in violation of public policy: one based on a public policy against age discrimination and one on a public policy against retaliatory discharge.\(^{173}\) The age discrimination policy derived not from the WLAD, but from a statute that, although affording no remedy, made terminating an employee between the ages of forty and seventy because of age an unfair employment practice.\(^{174}\) The retaliatory-discharge claim derived from

\(^{162}\) *Roberts*, 140 Wash. 2d at 67–68, 993 P.2d at 906. The court, citing *Marbury v. Madison*, 5 U.S. (1 Cranch) 137 (1803), noted that where a statute creates a right of recovery, the judiciary is obligated to devise a remedy. *Roberts*, 140 Wash. 2d at 68, 993 P.2d at 907.

\(^{163}\) *Roberts*, 140 Wash. 2d at 68–71, 993 P.2d at 907–08.

\(^{164}\) Id. at 68–69, 993 P.2d at 906.

\(^{165}\) Id. at 70, 993 P.2d at 908.

\(^{166}\) WASH. REV. CODE § 49.60.010 (2000).

\(^{167}\) Id. § 49.60.030.

\(^{168}\) *Roberts*, 140 Wash. 2d at 70, 993 P.2d at 908.

\(^{169}\) Id. at 77, 993 P.2d at 911.


\(^{171}\) Id. at 929, 784 P.2d at 1266.

\(^{172}\) Id. at 917, 784 P.2d at 1259.

\(^{173}\) Id. at 916, 784 P.2d at 1259.

\(^{174}\) Id. at 917, 784 P.2d at 1260; see also WASH. REV. CODE § 49.44.90 (2000).
Dicomes v. State, which recognized a public policy protecting employees discharged in retaliation for reporting employer misconduct. The Bennett court held the sisters’ claims were actionable and that the WLAD’s definition section “does not operate to restrict ‘employer’ as used in the separate chapter, 44.90.” In its discussion of the public policy against retaliatory discharge, the court noted the Legislature had adopted this policy in the WLAD. Analogizing from Bennett, the Roberts majority reasoned that the WLAD’s definition of employer had no impact on a common law tort action brought independently from the WLAD itself.

The Roberts majority sought to distinguish Griffin because the defendant relied heavily on Griffin’s language to argue the WLAD’s public policy protected small employers from discriminatory discharge claims. The Roberts majority held that the Griffin court had not announced a public policy shielding small employers from discrimination claims, but had merely speculated in dicta that the WLAD’s eight-employee floor might exist to shield small employers from regulatory burdens or litigation expenses. The Griffin majority reasoned that the issue was whether a WLAD action could be brought against an employer with fewer than eight employees. Therefore, it concluded Griffin did not control whether a plaintiff could maintain an action against an employer with fewer than eight employees for wrongful discharge in violation of public policy when that policy was found in the WLAD.

C. The Dissent

Justice Madsen, joined by Justice Guy, dissented. She argued that the court should not allow a common law action for wrongful discharge

175. 113 Wash. 2d 612, 782 P.2d 1002 (1989).
176. Bennett, 113 Wash. 2d at 924, 782 P.2d at 1263 (citing Dicomes, 113 Wash. 2d at 618–19, 782 P.2d at 1004).
177. Id. at 926, 784 P.2d at 1264.
178. Id.
180. Id. at 74, 993 P.2d at 910.
181. Id. at 75, 993 P.2d at 910.
182. Id. at 74, 993 P.2d at 910.
183. Id. at 75, 993 P.2d at 910.
184. Id. at 80–88, 993 P.2d at 912–17 (Madsen, J., dissenting).
based on a mandate of public policy found in a statute explicitly exempting the defendant employer.\textsuperscript{185} Justice Madsen’s analysis relied on Thompson v. St. Regis Paper Co.,\textsuperscript{186} which first established Washington’s standard for wrongful discharge in violation of public policy.\textsuperscript{187} The Thompson court, she reasoned, intended the public policy exception to be applied “narrowly and cautiously.”\textsuperscript{188} Justice Madsen contended that the majority failed to find a mandate sufficiently clear to meet the Thompson standard because any public policy against discrimination by an employer with eight or fewer employees was “muted” by the WLAD’s small-employer exemption.\textsuperscript{189}

Justice Madsen singled out the majority’s use of the WLAD as particularly “disingenuous.”\textsuperscript{190} Justice Madsen maintained that the majority had overstepped its judicial boundaries when it ignored the WLAD’s small-employer exemption and found a clear mandate of public policy against discrimination applicable to all employers regardless of size.\textsuperscript{191} She argued the Legislature, not the judiciary, should remedy any injustice resulting from the WLAD’s small-employer exemption.\textsuperscript{192} Justice Madsen supported her argument by noting that the Legislature had reaffirmed the policy of protecting small employers from discrimination claims by twice failing to pass proposals to eliminate or consider eliminating the small-employer exemption.\textsuperscript{193}

The dissent also attacked several other aspects of the majority’s reasoning, particularly its “strange” assertion that “a definitions section could not be a source of public policy.”\textsuperscript{194} Applying principles of statutory interpretation, Justice Madsen argued that the majority had

\begin{flushleft}
\textsuperscript{185} Id. at 80, 993 P.2d at 913 (Madsen, J., dissenting).
\textsuperscript{186} 102 Wash. 2d 219, 685 P.2d 1081 (1984).
\textsuperscript{187} See Roberts, 140 Wash. 2d 80, 993 P.2d at 913 (Madsen, J., dissenting); see also supra notes 28–31 and accompanying text.
\textsuperscript{188} Id. at 81, 993 P.2d at 913 (Madsen, J., dissenting).
\textsuperscript{189} Id. (Madsen, J., dissenting).
\textsuperscript{190} Id. (Madsen, J., dissenting).
\textsuperscript{191} Id. (Madsen, J., dissenting).
\textsuperscript{192} Id. (Madsen, J., dissenting).
\textsuperscript{193} Id. at 85, 993 P.2d at 915 (Madsen, J., dissenting). The majority responded that such failures do not signify a retraction of a policy against discrimination. Id. at 69 n.9, 993 P.2d at 907 n.9. The court further noted: “[I]f this demonstrates any legislative intent at all, it simply indicates that if the bill had been enacted, the legislature would have removed the small employer ‘exemption’ to the provisions of RCW 49.60.” Id.
\textsuperscript{194} Id. at 81, 993 P.2d at 913 (Madsen, J., dissenting).
\end{flushleft}
rendered the definitions section of the statute meaningless, a result courts should generally avoid. She argued that the WLAD, as interpreted in *Griffin*, reflected a balancing of employers’ and employees’ interests. By ignoring this balance, she argued that the majority rejected the Legislature’s judgment in favor of its own.

Justice Madsen further supported her opinion with case law from California, where the Supreme Court of California refused to recognize a public policy based on Fair Employment Practices Act (FEHA) which, like the WLAD, explicitly exempted the defendant employer. In *Jennings v. Marralle*, the Supreme Court of California refused to allow a tort claim for age discrimination in violation of a public policy when that policy was found exclusively in the California FEHA and where the employer had fewer than FEHA’s five-employee statutory minimum. Justice Madsen concluded that the court should have followed *Jennings*’s reasoning and held the WLAD could not constitute a clear mandate of public policy against discrimination when the statute specifically exempted the defendant.

In addition to her critique of the majority’s use of the WLAD, Justice Madsen argued the equal opportunity statute did not constitute a clear mandate of public policy against discrimination.

---

195. *Id.* at 82, 993 P.2d at 914 (Madsen, J., dissenting) (citing Childers v. Childers, 89 Wash. 2d 592, 596–97, 575 P.2d 201, 205 (1978)). An exception to this rule exists when rendering a section meaningless is necessary to avoid constitutional infirmities. *Id.* (Madsen, J., dissenting).
196. *Id.* at 83, 993 P.2d at 914 (Madsen, J., dissenting).
197. *See id.* at 80, 993 P.2d at 912 (Madsen, J., dissenting).
199. *Cal. Gov’t Code § 12926(d)* (West 1992); *see also supra* notes 115–118 and accompanying text.
201. 876 P.2d 1074 (Cal. 1994).
202. *Roberts*, 140 Wash. 2d at 84, 993 P.2d at 915 (Madsen, J., dissenting) (citing *Jennings*, 876 P.2d at 1076). Justice Madsen also noted another California court's holding that “when the constitutional provision or statute articulating a public policy also includes certain substantive limitations in scope or remedy, these limitations also circumscribe the common law wrongful discharge cause of action.” *Id.* (Madsen, J., dissenting) (quoting *City of Moorpark*, 959 P.2d at 762).
203. *Roberts*, 140 Wash. 2d at 84, 993 P.2d at 915 (Madsen, J., dissenting).
Roberts v. Dudley

policy against discrimination by private employers. The case law had never applied this law in the context of private employers. Therefore, she concluded, the equal opportunity statute applied only to government action and could not constitute a clear mandate of public policy against private employers.

IV. WASHINGTON’S PUBLIC POLICY AGAINST DISCRIMINATION BY EMPLOYERS WITH FEWER THAN EIGHT EMPLOYEES CANNOT BE FOUND IN THE WLAD

Although Washington public policy, as manifested in statutes and the Washington Constitution, condemns sex discrimination by all employers, the Roberts majority’s use of the WLAD marks a sharp and ill-advised departure from the narrow public policy exception to the employment-at-will doctrine introduced by Thompson. The Roberts majority erred in holding that the WLAD represents a clear mandate of public policy against sex discrimination when the defendant employer has fewer than eight employees. In doing so, the majority dramatically and unnecessarily broadened the public policy exception and misapplied its own precedent. The holding raises serious separation-of-powers concerns and creates the possibility of discrimination lawsuits against other exempt groups, including non-profit religious organizations. The Roberts majority correctly held the equal opportunity statute represented a clear mandate of public policy against sex discrimination. Washington’s ERA embodies another mandate of that public policy.

A. The Majority Erred in Finding that the WLAD Contains a Sufficiently Clear Mandate of Public Policy To Support a Common Law Wrongful Discharge Tort Action

Any public policy against sex discrimination by small employers found in the WLAD lacks the requisite clarity to support a tort action for wrongful discharge. Thompson explicitly requires a clear mandate of

204. Id. at 87–88, 993 P.2d at 916 (Madsen, J., dissenting).
205. Id. at 87, 993 P.2d at 916 (Madsen, J., dissenting).
206. Id. (Madsen, J., dissenting).
207. See supra notes 47–71 and accompanying text.
208. See Roberts, 140 Wash. 2d at 67–68, 993 P.2d at 906–07.
209. See id. at 77–78, 993 P.2d at 911–12 (Alexander, J., concurring).
public policy.\textsuperscript{210} Not only is the public policy found in the WLAD unclear with regard to small employers, it arguably indicates a public policy of not subjecting small employers to discrimination suits. The Roberts majority’s reasoning disregards the clear language and legislative intent of the statute, the implications of its own holding in Griffin, and other precedent where courts refused to apply the broad policy derived from statutes to a situation beyond the statute’s breadth.

1. The Plain Language of the Small-Employer Exemption and Relevant Legislative History Indicate the Legislature Did Not Intend To Regulate Small Employers Through the WLAD

The Legislature’s deliberate exemption of small employers from the WLAD clouds the statute’s mandate of public policy with regard to small employers and prevents it from providing the clear mandate required by Thompson. The plain language of the WLAD unequivocally exempts small employers from its coverage.\textsuperscript{211} It is unclear whether the Legislature intended to permit small employers to discriminate, chose simply not to address the problems of small employers in the WLAD, or, as the Roberts majority held, merely intended to limit the remedies of the statute and alleviate the WHRC’s regulatory burden.\textsuperscript{212} This lack of clarity, however, is exactly what precludes the WLAD from representing a clear mandate of public policy with respect to small employers. Thus, one need not reach Justice Madsen’s conclusion that the small-employer exemption actually mandates a public policy of shielding small employers from discrimination suits.\textsuperscript{213} Like religious groups and employment relationships within families, small employers were purposefully excluded from the statute. Whether that purpose is obvious, as in the case of nonprofit religious organizations,\textsuperscript{214} or left unclear, as in the case of the small employer,\textsuperscript{215} courts should not presume the legis-

\textsuperscript{211} WASH. REV. CODE § 49.60.040(3) (2000).
\textsuperscript{212} See Roberts, 140 Wash. 2d at 70, 993 P.2d at 908 (“By this section the legislature narrows the statutory remedies but does not narrow the public policy which is broader than the remedy provided.”).
\textsuperscript{213} See Roberts, 140 Wash. 2d at 93, 993 P.2d at 916 (Madsen, J., dissenting).
\textsuperscript{214} See supra note 102 and accompanying text.
\textsuperscript{215} See supra notes 106–119 and accompanying text.
lature’s inclusion of the exemption was an oversight.\textsuperscript{216} Therefore, the court should not have ignored the small-employer exemption when determining the public policy articulated in the WLAD.

Although limited, the available legislative history buttresses this conclusion. The small-employer exemption has been a part of the WLAD since its original enactment.\textsuperscript{217} Despite numerous revisions of the statute, the exemption has survived intact.\textsuperscript{218} The Legislature rejected two recent attempts to reconsider the exemption.\textsuperscript{219} As Justice Madsen noted, the Roberts majority’s assertion that the Legislature’s consideration of these bills supports the position that the Legislature intended the small-employer exemption to limit only statutory remedies is absurd.\textsuperscript{220} Given that both proposals were introduced in direct response to Griffin’s holding that the WLAD provides no cause of action against small employers,\textsuperscript{221} the Legislature’s failure to alter the definition of employer can only strengthen the dissent’s suggestion that the exemption represents a public policy protecting small employers.\textsuperscript{222} At the very least, the Legislature’s failure to provide a remedy to employees of small businesses contradicts any claim of a “clear mandate” to prevent discrimination by small employers. Had the Legislature felt that employees of small employers should be provided a remedy, it would not have passed up the opportunity to change or eliminate the small-employer exemption.

\textsuperscript{216} See, e.g., Childers v. Childers, 89 Wash. 2d 592, 596–97, 575 P.2d 201, 205 (1978) (noting that statutes should not be construed to render provisions meaningless except where necessary to avoid constitutional infirmity or reconcile conflicting statutes); Oak Harbor Sch. Dist. v. Oak Harbor Educ. Ass’n, 86 Wash. 2d 497, 500, 545 P.2d 1197, 1199 (1976) (citing Kelleher v. Ephrata Sch. Dist. No. 165, 56 Wash. 2d 866, 873, 355 P.2d 989, 993 (1960)) (presuming legislature does not engage in “vain and useless acts” and that some significant purpose or object is implicit in every legislative act).

\textsuperscript{217} See supra note 106 and accompanying text.

\textsuperscript{218} See id.

\textsuperscript{219} See supra notes 108–112 and accompanying text.

\textsuperscript{220} See Roberts v. Dudley, 140 Wash. 2d 58, 84, 993 P.2d 901, 914 (2000) (Madsen, J., dissenting).

\textsuperscript{221} See id. at 70, 993 P.2d at 908.

\textsuperscript{222} See supra note 193.
2. The Supreme Court of Washington’s Holding in Griffin Further Undermines the Majority’s Claim of a Clear Mandate of Public Policy

Despite the Roberts majority’s efforts to distinguish its holding from the holding of Griffin v. Eller, the reasoning of Griffin casts further doubt on the clarity of the mandate of public policy found in WLAD. The Roberts majority correctly rebutted Justice Madsen’s assertion that Griffin constitutes recognition of a public policy wholly protecting small employers from discrimination suits. However, the majority erred in holding that Griffin stands solely for the limitation of the WLAD’s remedies. In Griffin, the court explicitly acknowledged that the WLAD did not apply to small employers and, in the context of determining the constitutionality of the small-employer exemption, speculated that the Legislature may have sought to protect small employers from the burdens of excessive litigation. Although this speculation in dicta does not represent judicial recognition of a broad policy protecting small employers, it raises, at a minimum, the possibility that the Legislature intended to free small employers from a duty not to discriminate. The statements in Griffin, when combined with the express exemption itself, should have prevented the majority from finding in WLAD a clear mandate of public policy sufficient to satisfy the Thompson standard.

3. The Reasoning of Roberts Violates the Principle of Limiting the Public Policy of a Statute to the Circumstances the Statute Seeks To Regulate

The majority’s holding represents an about-face from Roe v. Quality Transportation Services and Gardner v. Loomis Armored, Inc., previous judicial refusals to read broad public policies into statutes that

---

223. 130 Wash. 2d 58, 922 P.2d 788 (1996); see also supra notes 180–183 and accompanying text.
224. See Roberts, 140 Wash. 2d at 86, 993 P.2d at 916.
225. See id. at 81, 993 P.2d at 913 (Madsen, J., dissenting).
226. See id. at 70–71, 993 P.2d at 908.
227. See supra notes 122–125 and accompanying text.
228. See supra note 126 and accompanying text.
regulate narrow sets of circumstances. 231 Unlike Roe, where the court refused to apply a policy derived from a statute to circumstances the Legislature had chosen not to address, 232 the Roberts majority interpreted the Legislature’s silence as to whether small employers could discriminate as an invitation to provide its own remedy. 233 In Gardner, the court held the public policies of assisting police officers and encouraging rescues, mandated by the statutes plaintiff cited, were limited to the statute’s specific circumstances. 234 In contrast, the Roberts majority found the policy of the WLAD to be broader than the circumstances the statute was intended to regulate, that is, employment relations at businesses with eight or more employees. 235

4. The Roberts Dissent Correctly Criticized the Majority for Finding a Clear Mandate of Public Policy by Reading One Section of the WLAD Outside Its Context

As the dissent noted, the majority’s use of the WLAD’s purpose section violated the principle that a court should ascertain a statute’s legislative purpose by construing the whole statute in order to avoid unintended consequences. 236 By reading the WLAD’s purpose section without considering the constraints found elsewhere in the statute, the court attempted to justify its conclusion that small employers are liable for violating the policy of a statute that expressly excludes them. 237 In doing so, the Roberts majority stripped the WLAD of its internal limitations, most notably the statutory definition of employer. From the remnants of the dismantled statute, the court inferred a broad policy against discrimination. It then concluded that because the discharge violated the WLAD’s broad purpose, the policy could encompass facts falling outside the statute’s scope. The majority’s removal of the purpose section from its context obscures the fact that the small-employer exemption clouds the policy of the WLAD with regard to small employers.

231. See supra notes 47–72 and accompanying text.
235. See Roberts, 140 Wash. 2d at 70, 993 P.2d at 908.
237. See supra notes 163–169 and accompanying text.
B. Bennett Does Not Support the Majority’s Finding of a Clear Mandate of Public Policy Against Discrimination by Small Employers

Bennett v. Hardy does not support the majority’s decision to disregard the WLAD’s definition section when determining the boundaries of the WLAD’s public policy. The court correctly cited Bennett for the proposition that the definition section of the WLAD does not bar a common law action against an employer with fewer than eight employees. However, the plaintiff in Bennett never contended the WLAD represented a clear mandate of public policy against either age or retaliatory discrimination. The Bennett court simply held that the WLAD’s definition of employer could not limit the public policy found outside of the WLAD. This holding differs dramatically from the Roberts court’s holding that the definition section of the WLAD does not limit the policy found in the WLAD itself.

Contrary to the majority’s assertion, Bennett does not support the argument that the WLAD can form a basis for public policy where the employer is excluded from the WLAD’s coverage. Having already established, without citing the WLAD, that a public policy against retaliatory discharge existed, the Bennett court noted in dicta that the Legislature had explicitly chosen to adopt this policy into the WLAD. By taking this minor premise out of context, the Roberts majority wrongly supported its unprecedented finding that a clear mandate of public policy can be found in a statute expressly exempting a potential target of that policy.

238. 113 Wash. 2d 912, 784 P.2d 1258 (1990).
239. Roberts, 140 Wash. 2d at 72, 993 P.2d at 908.
240. See supra notes 177–179 and accompanying text.
241. Bennett, 113 Wash. 2d at 926, 784 P.2d at 1264.
243. See supra note 179 and accompanying text.
244. See supra note 178 and accompanying text.
C. The Roberts Majority’s Use of the WLAD Represents a New Application of the Tort of Wrongful Discharge in Violation of Public Policy and the Doctrine of Separation of Powers

The Roberts majority, in a rush to accomplish the equitable goal of holding all employers liable for sex discrimination, opened a Pandora’s box of legal and policy problems. In addition to threatening religious organizations and family businesses with greater tort liability, the court’s holding considerably broadens the public policy exception to the employment-at-will doctrine. The result is a decision that not only undermines the legislative purpose of the WLAD, but greatly expands the role of the court as a policymaking body, a role it should not play. The holding also introduces tort causes of action for all other discriminatory actions prohibited by the WLAD.

The Roberts court’s holding raises serious separation-of-powers issues by using the public policy exception to fill a gap deliberately left in the statute by the Legislature. The Roe court recognized this problem and refused to find a policy protecting employees from drug testing because the Legislature, while aware of the issue, chose not to regulate it.245 In the case of the WLAD, the Legislature had not merely passively ignored an issue, as it did with employee drug testing, but actively created a specific exemption for certain employers. This is distinct from the situation found in Thompson, where a legislative act inadvertently affected employment relations while regulating an unrelated area of law.246 The WLAD both directly regulates employment relationships and provides remedies for violations of the rights it creates.247 In the WLAD, as Justice Madsen noted, the Legislature reached a balance between the rights of employees and employers, a balance reflecting the Thompson standard.248 The Roberts decision, therefore, represents a new application of the tort of wrongful discharge in violation of public policy: filling statutory gaps where the court determines the statutory remedy does not adequately reflect the Legislature’s intent.249

---

247. Supra Part I.C.
249. E.g., id. at 78–79, 993 P.2d at 912 (Talmadge, J., concurring).
The implications of *Roberts* illustrate why courts should construe public policies narrowly and limit the policies’ scope to the scope of the statutes from which they are drawn.250 After *Roberts*, other exempt groups, particularly religious organizations251 and persons employing spouses, children, or parents,252 are vulnerable to suits for wrongful discharge in violation of the WLAD’s public policy. Washington courts must now balance a common law action based on the WLAD against the policies insulating family members or religious groups from discrimination claims. The exemption for nonprofit religious and sectarian organizations, however, represents a legislative desire to uphold separation of church and state, a fundamental tenet of American democracy.253 Had Roberts accused a fundamentalist church of discharging her based on her conversion to another religion, the court could not have so blithely disregarded the policy set forth in the definition section. Nevertheless, under *Roberts*, a religious employer who discriminates, although protected from the WLAD’s statutory remedies, may be liable in common law tort for wrongful discharge in violation of the WLAD’s policy against religious discrimination.254

Furthermore, although the majority held that *Roberts* supports tort claims only for wrongful discharges and not for other manifestations of employment discrimination, the court’s reasoning is inconsistent. In discussing the equal opportunity statute,255 the court reasoned that where the Legislature created a right without a remedy, the judiciary must create a remedy.256 The WLAD expressly creates a broad right to be free of discrimination in employment extending beyond discharges to failure to hire, failure to promote, and wrongful demotions.257 If small employers are not exempt from this policy and the judiciary has a duty to create a remedy where the Legislature has failed to do so, the *Roberts*

250. *Id.* at 79, 993 P.2d at 912 (Talmadge, J., concurring).
251. See WASH. REV. CODE § 49.60.040(3) (2000).
252. See id. § 49.60.040(4).
253. See Farnam v. CRISTA Ministries, 116 Wash. 2d 659, 676–677, 807 P.2d 830, 838 (1991) (regarding Christian nursing home exempt from WLAD-based cause of action arising from discharge for religion-based objection to removing feeding tubes from patients); see also U.S. CONST. amend I.
254. See WASH. REV. CODE § 49.60.010.
255. *Id.* § 49.12.200.
257. See WASH. REV. CODE § 49.60.180.
court’s judicial remediation means that a job applicant of a small employer could also sue in tort for discriminatory failure to hire. 258


Justice Madsen’s rejection of both the equal opportunity statute and the ERA, on the grounds that they do not apply to private employers, was unfounded. The plain language of these statutes strongly evinces a public policy against treating men and women differently. As the Roberts majority noted, the language of the equal opportunity statute as well as its location in a chapter regulating the industrial practices of private corporations indicates it is applicable to private and public employers. No case law or legislative history indicates private employers are beyond its reach.

Justice Madsen, although correct in her concern about the majority’s use of the WLAD, erred in arguing the small-employer exemption mutes Washington’s public policy against sex discrimination. As the Bennett court found, the WLAD’s explicit provision that it not repeal any other anti-discrimination laws 259 prohibits the application of the small-employer definition beyond the WLAD. 260 Therefore, a wrongful-discharge claim based on a clear mandate of public policy drawn from a statute other than the WLAD would be enforceable against any employer, provided there was not a similar provision exempting the employer from the scope of that statute. 261 The Supreme Court of California employed this line of reasoning in Jennings v. Marralle 262 where it held that, had the plaintiff identified a public policy against age discrimination outside FEHA, then FEHA’s small-employer exception would not block a common law wrongful discharge claim. 263

A legally sound middle ground exists between the majority’s distortion of the WLAD’s policy and the dissent’s refusal to recognize a

259. WASH. REV. CODE § 49.60.020.
260. See supra notes 170–179 and accompanying text.
261. See supra notes 201–202 and accompanying text.
262. 876 P.2d 1074 (Cal. 1994).
263. See id. at 1076.
broad policy against sex discrimination in Washington. The court should have reached the same result on much narrower grounds by allowing an action for wrongful discharge in violation of public policy based on the public policy against sex discrimination in the equal opportunity statute. The Legislature, through the equal opportunity statute, created a right to be free from sex discrimination in employment but trusted the courts to remedy employers’ violations of that policy.264 Because case law is so sparse,265 the Roberts dissent’s objection that cases interpreting the equal opportunity statute have all been applied only to public employers is unpersuasive.266 Based on Bennett, the court could have addressed the WLAD only to note that its small-employer exemption cannot block a common law action based on a separate statute. Unlike the WLAD, the equal opportunity statute contains no limiting provision exempting any employers—public or private, large or small—from the duty to allow women the same employment opportunities given to men.267 Therefore, under the equal opportunity statute, even a small employer such as Dudley has a duty not to discriminate on the basis of sex. A discharge on that basis is tortious and violates the public policy of Washington.268 Rather than reading the WLAD to imply a broader public policy than its remedies allow, thereby jeopardizing other exempted groups, the court should have used the equal opportunity statute to achieve the same result.

Although the judiciary prefers not to decide cases on constitutional grounds,269 the Washington Constitution’s ERA provides a second mandate of public policy against sex discrimination in employment. Again, the clear language of the amendment ensures women receive equal treatment. In Guard v. Jackson,270 the Supreme Court of Washington explicitly recognized that the ERA unequivocally condemns sex discrimination.271 Thus, the ERA also creates a mandate of public policy against sex discrimination sufficient to meet Thompson’s clarity requirement for a tort of wrongful discharge in violation of public policy. Either the equal opportunity statute or the ERA would have been

264. Wash. Rev. Code § 49.12.200; see also supra notes 74–78 and accompanying text.
265. Supra notes 74–79 and accompanying text.
267. Supra note 76 and accompanying text.
268. Roberts, 140 Wash. 2d at 67, 993 P.2d at 906.
269. Id. at 62 n.2, 993 P.2d at 904 n.2.
271. Id. at 330–31, 921 P.2d at 547.
sufficient grounds for upholding an exception to employment at will in *Roberts*.

V. CONCLUSION

Given Thompson’s “clear mandate” and the plain language of the small-employer exemption, the *Roberts* majority’s justifications for determining that the WLAD is a sufficiently clear mandate of public policy to support a tort of wrongful discharge are not convincing. Although the WLAD may indicate a general policy against employment discrimination, the exemption for small employers, held constitutionally sound and surviving years of statutory revisions, should have precluded the court from finding a clear mandate of public policy in the WLAD against sex discrimination by small employers. Instead, *Roberts* gives courts free rein to create remedies and public policies where they see fit, by dissecting statutes without the support of legislative history or case law. It also unnecessarily exposes other exempted groups to tort liability. Ironically, the court could have avoided these problems by finding that the WLAD’s definitions do not insulate small employers from common law tort claims based on other laws’ public policies against discrimination. Instead, it chose to set a precedent that broadened the holding of *Thompson* to give Washington courts the ability to construe public policy too broadly and thereby threaten the balance of powers on which the government is founded.
THE REASONABLE GIRL: A NEW REASONABLENESS STANDARD TO DETERMINE SEXUAL HARASSMENT IN SCHOOLS

Carrie L. Hoon

Abstract: The U.S. Supreme Court held in Davis v. Monroe County Board of Education that schools may be liable under Title IX of the 1972 Education Amendments for student-to-student hostile-environment sexual harassment. Although the Court required that conduct be severe, pervasive, and objectively offensive to qualify as sexual harassment under the statute, it did not establish an objective reasonableness standard to evaluate allegedly harassing conduct. In the context of Title VII employment-discrimination jurisprudence, some courts apply a reasonable-woman standard to determine what conduct is objectively hostile or abusive such that it constitutes actionable hostile-environment sexual harassment in the workplace. This Comment argues that a reasonable-girl standard, which is an amalgamation of reasonable-woman precedent and the reasonable child from tort law, is consistent with previous U.S. Supreme Court interpretations of Title IX. This Comment further contends that courts should adopt the reasonable-girl standard because it will further girls' equal educational opportunities, thereby serving the goal of Title IX.

In February of 2000, a thirteen-year-old girl in Federal Way, Washington, filed a complaint against Kilo Junior High School for failing to prevent a classmate from repeatedly accosting her in the school’s hallways and reaching under her clothing to grope her.1 The school had responded to the girl’s complaint of harassment by having her confront her alleged perpetrator, whom she feared, and eventually suggesting that she leave the school district, which she did.2 The school never punished the perpetrator.3 The experience caused the girl’s grades to suffer and left her feeling as though she had no friends.4

This thirteen-year-old girl is not alone. Despite studies chronicling the prevalence of sexual harassment in schools and its adverse effects on girls, including reduced academic performance and damaged emotional well-being,5 the U.S. Supreme Court has not yet established an objective standard that schools and courts can use to determine what types of conduct rise to the level of sexual harassment. Consequently, Kilo Junior

2. Id.
3. Id.
4. Id.
5. See infra Part I.A.
High School’s administrators had no guidance to determine whether the
conduct the thirteen-year-old girl complained of amounted to sexual
harassment for which the school could become liable.

In some circuits, the courts of appeals have been one step ahead of the
U.S. Supreme Court in adopting a reasonableness standard to aid in
determining what conduct constitutes hostile-environment sexual harass-
ment. In the workplace context, these courts have adopted a gender-
specific reasonable-woman standard, rather than a gender-neutral
reasonable-person standard, in determining what is objectively hostile or
abusive. Under Title IX, a statute that prohibits sex discrimination in
schools, the U.S. Supreme Court has determined that sexual harassment
is a form of sex discrimination and thus violates the statute. In Davis v.
Monroe County Board of Education, the U.S. Supreme Court held that
schools may be held liable for student-to-student hostile-environment
sexual harassment. However, by focusing the inquiry on whether a
school has taken appropriate steps to remedy the harassment rather than
on the type of conduct that amounts to sexual harassment in schools, the
Davis Court neglected to establish a reasonableness standard to
determine whether allegedly harassing conduct is objectively offensive
and thus amounts to actionable hostile-environment sexual harassment.

This Comment examines the standards of reasonableness by which
student-to-student sexual harassment should be evaluated when girls in
junior high and high school are the victims and boys are the

9. Id. at 650. There is great disagreement regarding what conduct constitutes hostile-environment
sexual harassment. This Comment adopts the U.S. Supreme Court’s definition of sexual harassment,
meaning conduct such as comments, gestures, or unwelcome touching that is so severe, pervasive,
and objectively hostile or abusive that it limits the victim’s access to equal opportunity. See, e.g.,
U.S. 17, 21 (1993). Different courts of appeals have found hostile-environment sexual harassment to
stem from different conduct. Compare Torres v. Pisano, 116 F.3d 625, 632–33 (2d Cir. 1997)
(holding that calling woman “dumb cunt” and telling her to “stay home, go on welfare, and collect
food stamps like the rest of the ‘spics’” constituted hostile environment), and Ellison, 924 F.2d at
878 (holding that continuing to send love letters despite request to stop constituted hostile
environment), with DeAngelis v. El Paso Mun. Police Officers Ass’n, 51 F.3d 591, 596 (5th Cir.
1995) (holding that derogatory comments about women in general and plaintiff in particular
published in monthly workplace newsletter could not amount to hostile-environment sexual
harassment), and Rabidue v. Osceola Ref. Co., 805 F.2d 611, 622 (6th Cir. 1986) (holding that
pornographic comments and pictures that sexually objectified plaintiff did not create hostile
environment).
10. See Davis, 526 U.S. at 650.
Sexual Harassment in Schools

Part I of this Comment provides an overview of sexual harassment in schools. Part II explores previous judicial interpretations of Title IX of the Education Amendments of 1972. Part III discusses competing reasonableness standards that courts use to determine what conduct is objectively offensive—the reasonable-person standard, the reasonable-woman standard, and the reasonable-person-in-the-plaintiff’s-position standard—and the U.S. Supreme Court’s position regarding these standards. Part IV argues that courts should adopt a reasonable-girl standard when determining the objective offensiveness of harassing conduct under Title IX because this standard both comports with previous U.S. Supreme Court interpretations of the statute and best serves the purposes of the statute. This Comment concludes that the reasonable-girl standard would put girls on an even playing field with boys with regard to educational opportunities.

I. SEXUAL HARASSMENT IN SCHOOLS

Statistics show that sexual harassment in schools is prevalent and has many negative effects on girls. In *Davis*, the U.S. Supreme Court addressed this problem by finding that schools that allow girls to continue to be harassed by their peers may be liable in private actions brought under Title IX. *Davis* extended case law that had found that individuals may bring private actions against schools when they are deliberately indifferent to known sexual harassment to include school liability for student-to-student sexual harassment.

A. The Prevalence and Effects of Sexual Harassment in Schools

Empirical evidence collected throughout the past decade demonstrates that sexual harassment in schools is prevalent and that girls are common victims of sexual violence. Half of the high school girls surveyed in one study reported having been sexually harassed in school. 12 A study of junior high school students found that 17% reported having been

11. The “reasonable-girl standard” is one standard that may be used to determine whether alleged harassment is objectively offensive such that it may amount to sexual harassment. Although the reasonable-girl standard may be applicable in same-sex harassment cases, or in cases where boys are the victims and girls are the perpetrators, these situations are beyond the scope of this Comment. The gender issues at play in those two instances could affect the type of standard that would be appropriate.

sexually coerced by a teenager, 19% reported feeling pressure from their friends to have intercourse, 7% reported having been sexually coerced by an adult, and 6% reported having sexually coerced someone else.\textsuperscript{13} Another study found that 7% of respondents, primarily women between the ages eighteen and twenty-two, had experienced at least one episode of involuntary sexual intercourse during childhood or adolescence, while less than one-half of these experiences occurred before age fourteen.\textsuperscript{14} The majority of these unwanted sexual experiences happened to girls between the ages of thirteen and sixteen.\textsuperscript{15}

Sexual harassment in schools can be devastating to girls both academically and emotionally. Empirical evidence focused on high school students demonstrates that sexual harassment can lead to absenteeism, decreased quality of schoolwork, skipping or dropping classes, lower grades, loss of friends, tardiness, and truancy, all of which may result in ineligibility for specific colleges or merit scholarships and loss of recommendations for awards, colleges, or jobs.\textsuperscript{16} Furthermore, early adolescents who were less successful academically were more likely to have been coerced sexually.\textsuperscript{17}

B. Davis v. Monroe County Board of Education

Extended School Liability for Sexual Harassment to Include Liability for Student-To-Student Sexual Harassment

In \textit{Davis}, the U.S. Supreme Court held that a school could be liable under Title IX for allowing known student-to-student harassing conduct to occur.\textsuperscript{18} The plaintiff in \textit{Davis} was a fifth-grade girl. Another student had touched and rubbed up against her breasts and genital area, and told her vulgar statements such as “I want to get in bed with you” and “I want to feel your boobs.”\textsuperscript{19} The girl’s mother complained to the school principal, but the school took no disciplinary action; meanwhile the harassment continued for months, stopping only when the harasser was

\begin{footnotes}
\footnote{15. Pamela I. Erickson et al., \textit{Unwanted Sexual Experiences Among Middle and High School Youth}, 12 J. ADOLESCENT HEALTH 319, 321 (1991).}
\footnote{16. Fineran et al., \textit{supra} note 12, at 55.}
\footnote{17. Jordan et al., \textit{supra} note 13, at 290.}
\footnote{18. See \textit{Davis v. Monroe County Bd. of Educ.}, 526 U.S. 629, 633 (1999).}
\footnote{19. \textit{Id.}}
\end{footnotes}
Sexual Harassment in Schools

charged with sexual battery.\textsuperscript{20} The Court concluded that the school had created a hostile environment for the girl by allowing the harassment to continue for months and found that the school could be held liable under Title IX for its deliberate indifference to the harassment.\textsuperscript{21}

\textit{Davis} extended prior cases holding schools liable for inaction in the face of known harassment. In \textit{Franklin v. Gwinnett County Public Schools},\textsuperscript{22} the plaintiff was a student who had been subjected by a teacher to continued sexual harassment that included sexually oriented conversations, forced kissing, and coercive intercourse.\textsuperscript{23} The school district allegedly became aware of the teacher’s misconduct and investigated but took no remedial action until the victim filed the lawsuit.\textsuperscript{24} The U.S. Supreme Court found that under Title IX, the statute that prohibits sex discrimination in schools, sexual harassment may amount to sex discrimination and held the school liable for the teacher’s actions.\textsuperscript{25} The plaintiff in \textit{Gebser v. Lago Vista Independent School District}\textsuperscript{26} was an eighth-grade student when her teacher began making sexually suggestive comments to her.\textsuperscript{27} The school responded to complaints of these comments by telling the teacher to be careful about his classroom comments.\textsuperscript{28} Nevertheless, the teacher initiated sexual contact with the plaintiff, kissed and fondled her, and later engaged in sexual intercourse with her, often during class time.\textsuperscript{29} The Court concluded that schools may only be held liable for taking inadequate remedial action when they have been informed of the harassment.\textsuperscript{30}

II. JUDICIAL INTERPRETATION OF TITLE IX IN LIGHT OF LEGISLATIVE HISTORY, TITLE VI, AND TITLE VII

The U.S. Supreme Court has examined Title IX sexual harassment in schools from three perspectives. First, the fact that Title IX is a contract

\begin{itemize}
  \item \textit{Id.} at 634.
  \item See \textit{id.} at 653–54.
  \item 503 U.S. 60 (1992).
  \item \textit{Id.} at 63.
  \item \textit{Id.} at 64.
  \item See \textit{id.} at 75.
  \item 524 U.S. 274 (1998).
  \item \textit{Id.} at 277–78.
  \item \textit{Id.} at 278.
  \item \textit{Id.}
  \item See \textit{id.} at 290.
\end{itemize}
between schools and the federal government is the foundation on which all U.S. Supreme Court interpretations of the statute are built. Second, the Court has analogized Title IX to Title VI of the 1964 Civil Rights Act, interpreting Title IX as a contractual agreement that not only explicitly calls for administrative remedies, but also implicitly allows private causes of action when schools act with deliberate indifference to the rights of students by taking clearly unreasonable remedial actions. Lastly, the Court has relied on Title VII of the 1964 Civil Rights Act when interpreting sex discrimination under Title IX to include hostile-environment sexual harassment when a school has actual notice of the harassment.

A. Title IX Is a Binding Contractual Agreement that Prohibits Schools from Discriminating on the Basis of Sex

Title IX of the 1972 Education Amendments is a contract between schools and the federal government that conditions federal funding on a promise to end sex discrimination in schools. This statute prohibits schools from excluding, on the basis of sex, students from the participation in or benefits of any federally funded educational activity. Implementing regulations provide that a school district receiving federal funding shall not “otherwise limit any person in the enjoyment of any right, privilege, advantage, or opportunity.” Courts have also used Title IX to remedy broad-based discrimination in admissions and athletic programs.

Title IX’s contractual nature controls when and how schools may become liable in private actions. Enacted pursuant to Congress’s

33. See Davis, 526 U.S. at 648.
34. 42 U.S.C. § 2000e.
35. See Davis, 526 U.S. at 648.
40. See, e.g., Favia v. Ind. Univ. of Penn., 7 F.3d 332, 343-44 (3d Cir. 1993) (holding that university receiving federal funding must make same amount of money available to women’s and men’s athletic programs); Berkelman v. S.F. Unified Sch. Dist., 501 F.2d 1264, 1269-70 (9th Cir. 1974) (holding that school district may not apply higher admission standards to girls than to boys).
Sexual Harassment in Schools

spending powers, Title IX makes the initial and continued provision of federal funding for educational programs and activities contingent upon the schools’ promise not to discriminate on the basis of sex. The Court has concluded that because Title IX is a contract, rather than a broad-based remedial measure such as Title VII, liability may not arise under it unless educational institutions have been notified that they may become liable. Contractual statutes such as Title IX, enacted pursuant to the Spending Clause, prohibit only intentional discrimination. For a school to intentionally violate Title IX when it is not the actual harasser, it must have actual knowledge of the harassment, know what standard to use to determine whether the harassment rises to the level of discrimination, and ignore harassment that amounts to a violation of the statute such that it allows discrimination against the victim. In the administrative-remedy context, this means that the Department of Education (DOE) must notify a school of its violation of the statute and offer the school an opportunity to remedy the situation before remedies are imposed upon it. In the private-action context, this means that schools may be liable for damages only when they had actual notice of discrimination, had an opportunity to remedy the discrimination, and were deliberately indifferent to known discrimination.

Only deliberate indifference will support monetary damages in private Title IX actions. To make schools liable for their own inaction, the Court in Gebser held schools liable for allowing discrimination to

42. See id. at 286.
43. See id. at 290.
44. U.S. Const., art. I, § 8, cl. 1.
46. When school administrators investigate and resolve claims of sexual harassment of students, they are guided by the Department of Education (DOE) Office of Civil Rights Guidance on Sexual Harassment, which has been cited with approval by the U.S. Supreme Court, applies to students at every level of education, and is intended to inform officials of the standards and procedures that schools are expected to follow under Title IX. Sexual Harassment Guidance: Harassment of Students by School Employees, Other Students, or Third Parties, 62 Fed. Reg. 12,034, 12,034 (March 13, 1997); see also Davis v. Monroe County Bd. of Educ., 526 U.S. 629, 651 (1999). Although implementation of the standard proposed by this Comment would require a change in the standard at some regulatory level before a school could become liable under the new standard, details regarding implementation are beyond the scope of this Comment.
47. See Gebser, 524 U.S. at 287.
48. See id. at 288.
49. See Davis, 526 U.S. at 643–44.
50. Gebser, 524 U.S. at 290. Monetary damages include both punitive and compensatory damages. See id. at 285–86.
continue despite actual notice of the discrimination. The Court determined that agency liability of any kind, like that arising under Title VII when a party has constructive knowledge and thus should have known that discrimination was taking place, is untenable under Title IX given the statute’s contractual nature and the notice requirement that flows from it. Only when schools’ intentional discrimination through deliberate indifference subjects victims of discrimination to further harassment is the threshold of action met. For example, in Davis the school became liable by refusing to take any remedial action in response to known acts of harassment such as the persistent unwanted groping and fondling of a fifth-grade girl.

The “clearly unreasonable” standard governs the determination of deliberate indifference. Courts will deem school administrators’ actions or lack thereof deliberately indifferent only if their responses to harassment that is severe, pervasive, and objectively offensive are clearly unreasonable in light of the known circumstances. While the U.S. Supreme Court has not clarified its definition of this standard in the Title IX context, it has explained why it uses the standard and how it expects courts to utilize it. In her majority opinion in Davis, Justice O’Connor stated that the clearly unreasonable standard would allow administrators considerable discretion to determine what remedial response was appropriate, because the judiciary should “refrain from second guessing the disciplinary decisions made by school administrators.” She emphasized that this standard is not one of “mere reasonableness” and that courts should be able to determine for themselves on motions for directed verdict whether the school’s response was clearly unreasonable. Finally, she stated that this standard would allow sufficient flexibility for schools of different types to address harassment

---

51. See id. at 285.
52. See id. at 287–88.
53. Id. at 290.
54. See Davis, 526 U.S. at 653–54.
55. Id. at 648–49.
56. Id. at 648 (“School administrators will continue to enjoy the flexibility they require so long as funding recipients are deemed ‘deliberately indifferent’ to acts of student-on-student harassment only where the recipient’s response to the harassment or lack thereof is clearly unreasonable in light of the known circumstances.”).
57. See id. at 648–49.
58. Id.
59. Id. at 649.
Sexual Harassment in Schools

commensurate with the level of control they have over their students.60 One method of showing that the school’s response was clearly unreasonable under Title IX is to demonstrate, as the Davis plaintiff did, that by failing to respond in any way to her complaints of sexual harassment the school board in effect subjected the student to continued discrimination.61 The Court has not yet decided a case in which the plaintiffs alleged that a school responded, but did so in a clearly unreasonable manner.

B. Liability Arises Similarly Under Both Title IX and Title VI

Because Congress patterned Title IX after Title VI and intended both statutes to end discrimination in schools,62 courts have looked to Title VI liability standards to determine when liability arises under Title IX.63 Title VI forbids the use of federal funds in programs that discriminate on racial grounds.64 Title VI has both an administrative remedy, which authorizes the DOE to investigate and force educational institutions to comply with the statute or risk the loss of federal funding, and a private remedy, which makes schools liable to victimized students for allowing race discrimination to continue.65 A school becomes liable under Title VI when there is a racially hostile environment of which the district has notice but fails adequately to remedy.66

In addition to the potential loss of federal funding, educational institutions also face liability for monetary damages under Title IX when sex discrimination occurs in their programs and activities. In Cannon v. University of Chicago,67 the U.S. Supreme Court held that under Title IX victims of such sex discrimination could bring a private cause of action against educational institutions.68 The Court’s decision rested on four grounds. First, the Court noted that the legislative history of Title IX indicates congressional approval of Title VI’s statutory enforcement

60. Id.
61. Id. at 646–47, 654.
65. See Monteiro v. Tempe Union High Sch. Dist., 158 F.3d 1022, 1033 (9th Cir. 1998).
66. Id.
68. See id. at 705–08.
procedures, which include a private remedy. Second, because the Court had already interpreted Title VI as allowing private causes of action, the Court observed that the similarities between Titles VI and IX supported interpreting the latter as giving rise to private remedies. Third, the Court noted that the administrative procedure for cutting off federal financial support for institutions was an insufficient remedy for victims of sex discrimination. Finally, the Court reasoned that a private remedy would enhance the efficiency of the enforcement mechanism of an exclusively administrative scheme because the burden of proof for a single plaintiff—that a violation of the statute occurred—is much lower than the “pervasive discrimination” required for agency action.

C. Courts Have Looked to Title VII Jurisprudence To Define Sex Discrimination Under Title IX

Courts often look to Title VII case law to determine what behaviors Congress meant to proscribe when it stated that “sex discrimination” was a violation of Title IX. In the Title IX context, courts rely on Title VII precedent holding that sexual harassment is a form of sex discrimination, defining sexual harassment as conduct that is severe, pervasive, and objectively offensive. Further, the Court cited Title VII precedent when finding that Title IX protects subordinates and students from harassment by superiors and peers.

While courts look to Title VI to determine standards of liability under Title IX, they also look to Title VII to determine what conduct amounts to sex discrimination under Title IX. Title VII prohibits sex discrimination in the employment context. Courts have interpreted this statute as prohibiting both quid pro quo sexual harassment and hostile-

---

69. Id. at 706 n.40; see also 118 Cong. Rec. 5807 (1972); 118 Cong. Rec. 5806–07 (1972); 117 Cong. Rec. 30,408 (1971).
70. See Cannon, 441 U.S. at 704–05.
71. Id.
72. Id. at 705.
74. See Davis, 526 U.S. at 651; Franklin, 503 U.S. at 75.
76. Quid pro quo sexual harassment under Title VII involves a supervisor offering an employee tangible goods such as job benefits in exchange for sexual favors. See, e.g., Henson v. City of Dundee, 682 F.2d 897, 908–09 (11th Cir. 1982) (finding that employer may not require employee to exchange sexual favors for job benefits). Quid pro quo sexual harassment is the easier form of sexual
environment sexual harassment when the harassment interferes with the victim’s working conditions such that she no longer has equal opportunity to work. In *Meritor Savings Bank v. Vinson*, the U.S. Supreme Court first recognized an action under Title VII for hostile-environment sexual harassment, defining it as conduct by the employer that is sufficiently severe or pervasive to alter the victim’s conditions of employment. Thus, only that sexual harassment that alters the conditions of employment is actionable under Title VII, whereas harassment that does not rise to the level of interfering with the conditions of employment is not actionable. The Court justified this recognition of hostile-environment sexual harassment on the same grounds that it has used to prevent racial harassment, stating that requiring a man or woman to “run a gauntlet of sexual abuse in return for the privilege of being allowed to work can . . . be as demeaning and disconcerting as the harshest of racial epithets.”

The U.S. Supreme Court and some federal circuit courts have looked to Title VII case law to determine whether sexual harassment has occurred under Title IX. In *Davis*, the Court defined hostile-environment sexual harassment as conduct that is severe, pervasive, and objectively offensive such that it undermines the victim’s educational experience and frustrates the purpose of Title IX: to afford everyone the benefit of equal access to education. By adopting the *Meritor* definition of sex discrimination for Title IX purposes, the *Davis* Court equated sex

77. See generally supra note 9.
80. Id. at 67. Although the Court later eliminated the distinction between quid pro quo and hostile-environment sexual harassment in the employment context, see *Burlington Indus., Inc. v. Ellerth*, 524 U.S. 742, 754 (1998), *Faragher v. City of Boca Raton*, 524 U.S. 775, 802 (1998), this Comment does not rely on or reference a distinction between the two types of sexual harassment.
81. See *Meritor*, 477 U.S. at 67.
82. Id. (quoting Henson v. City of Dundee, 682 F.2d 897, 902 (11th Cir. 1982)).
84. See supra notes 79–82.
discrimination with sexual harassment.\footnote{85 See Davis v. Monroe County Bd. of Educ., 526 U.S. 629, 650–51 (1999) (citing Meritor, 477 U.S. 57; Oncale v. Sundowner Offshore Serv., Inc., 523 U.S. 75 (1998)).} The Davis Court concluded that a hostile environment can be created by verbal as well as physical harassment.\footnote{86 See Davis, 526 U.S. at 650–51.} In addition, the Ninth Circuit has explicitly concluded that “Title VII standards apply to hostile environment claims under Title IX,”\footnote{87 Oona, 143 F.3d at 477.} using Title VII precedent to interpret sex discrimination under Title IX.

The U.S. Supreme Court and the Ninth Circuit have also cited Title VII case law as precedent for protecting students from harassment by a superior. In Franklin, the Court established that a school district can be liable for damages in cases involving a teacher’s sexual harassment of a student.\footnote{88 Franklin, 503 U.S. at 75.} The Court reasoned that the teacher-student relationship in the educational context is analogous to the supervisor-subordinate relationship in the employment discrimination context.\footnote{89 Id.} The Court relied on Meritor as persuasive authority for imposing liability on the school district.\footnote{90 Id.} By explaining how school liability arises, the Franklin Court also clarified its reference to Meritor in Gebser, where it held that a school district is not liable for damages in a teacher-student sexual harassment case “unless an official who at a minimum has authority to address the alleged discrimination and to institute corrective measures . . . has actual knowledge of discrimination . . . and fails adequately to respond.”\footnote{91 Gebser v. Lago Vista Indep. Sch. Dist., 524 U.S. 274, 290 (1998).} In Oona, R.-S.- v. McCaffrey,\footnote{92 143 F.3d 473 (9th Cir. 1998).} the Ninth Circuit held that school districts may be held liable for teacher-student harassment when they fail to remedy a known hostile environment by failing to supervise the teacher.\footnote{93 Id. at 477–78.} In all of these cases, the courts ultimately found that schools may be held liable for failing to remedy sexual harassment.\footnote{94 See Gebser, 524 U.S. at 478; Franklin, 503 U.S. at 63–64, 76; Oona, 143 F.3d at 477–78.} Moreover, the U.S. Supreme Court and the Ninth Circuit have found support in Title VII precedent when concluding that Title IX, like Title VII, also protects victims from peer hostile-environment sexual
Sexual Harassment in Schools

harassment. The Davis Court again cited Meritor for the proposition that school officials’ potential liability for failure to remedy known peer hostile-environment sexual harassment under Title IX is triggered only when a student’s conduct is so “severe, pervasive, and objectively offensive” as to deprive the victim of the educational opportunities provided by the school. In Oona, the Ninth Circuit cited Ellison v. Brady, a Title VII hostile-environment sexual harassment case, as authority for finding that school districts are liable under Title IX for not remedying peer hostile-environment sexual harassment in schools. Thus, both the U.S. Supreme Court and the Ninth Circuit look to Title VII precedent when determining what conduct amounts to peer hostile-environment sexual harassment.

III. COURTS DO NOT AGREE ON A SINGLE REASONABLENESS STANDARD TO DETERMINE WHETHER A VICTIM EXPERIENCED SEXUAL HARASSMENT

Because the Davis Court did not adopt a specific reasonableness standard to be used under Title IX to determine whether conduct is objectively offensive, the historical debate among courts and legal commentators regarding which reasonableness standard best aids in the objective evaluation of harassing conduct continues in the Title IX context. Some courts use the traditional reasonable-person standard to determine whether harassing conduct was objectively offensive. In reaction to what courts and commentators have identified as the failings of the reasonable-person standard, some courts apply a reasonable-woman standard. In Harris v. Forklift Systems, Inc. and Oncale v. Sundowner Offshore Services, Inc., the U.S. Supreme Court could have

95. See Davis v. Monroe County Bd. of Educ., 526 U.S. 629, 650 (1999) (“Having previously determined that ‘sexual harassment’ is ‘discrimination’ in the school context under Title IX, we are constrained to conclude that student-on-student sexual harassment . . . can likewise rise to the level of discrimination actionable under the statute.”).
96. 924 F.2d 872 (9th Cir. 1991).
97. Oona, 143 F.3d at 477 (“We [have] expressly recognized that hostile environments include peer harassment in Ellison v. Brady [924 F.2d 872 (9th Cir. 1991)]. . . . Accordingly, we hold that the defendants are not entitled to immunity for their failure to take steps to remedy the hostile environment created by the male students in Oona’s class.”).
98. See, e.g., DeAngelis v. El Paso Mun. Police Officers Ass’n, 51 F.3d 591, 594 (5th Cir. 1995).
squarely resolved the question of which standard applies. Instead, the decisions in those cases caused more rather than less confusion.

A.  The Reasonable-Person Standard

The evolution of the reasonable-person standard, which can be traced to the reasonable-man standard, originated in English common law as an objective standard to which courts compared litigants’ actions. When using this standard, courts occasionally consider certain defining characteristics of the person whose conduct is being evaluated. For example, courts have fashioned standards for a contextualized reasonable man, and in cases utilizing a reasonable-child standard when they find that a child would not have interpreted facts in the same way as an adult. Although courts have resisted altering the reasonable-man standard in order to create incentives for people to be careful, they have sometimes raised the standard by requiring defendants to consider the unique perspective of the person being harmed. Recently, courts have adopted gender-neutral language, changing the reasonable-man standard to that of the reasonable person, and have applied the standard to men and women alike.

101. See, e.g., Vaughan v. Menlove, 132 Eng. Rep. 490, 491 (1837). This Comment recognizes the difference between judging the reasonableness of a defendant’s conduct, as occurs in torts cases, and judging the reasonableness of a plaintiff’s response to some conduct, as occurs in discrimination cases. Instead, however, this Comment considers whether the conduct of the person who is being judged is objectively offensive.


103. Roth, 13 Wash. at 545, 43 P. at 647.

104. See RESTATEMENT (SECOND) OF TORTS §§ 283B(4), 895I(a) (1965); see also Johnson v. Lambotte, 363 P.2d 165, 166 (Colo. 1961); Ellis v. Fixico, 50 P.2d 162, 164 (Okla. 1935) (holding mentally ill people to same standard of care as mentally fit people).

105. See Scott v. Bradford, 606 P.2d 554, 558 (Okla. 1979) (holding that, in informed-consent case, doctor must anticipate what reasonable patient would want to be informed of); see also Nickerson v. Hodges, 84 So. 37, 39 (La. 1920) (holding that defendant should have anticipated that plaintiff would experience emotional distress because of his actions).


Courts have also utilized the reasonable-person standard as the generic standard by which women’s interpretations of discrimination are judged to determine whether their reaction was reasonable. For example, in DeAngelis v. El Paso Municipal Police Officers Association, the Fifth Circuit found, under the reasonable-person standard, that when an anonymous co-worker wrote harassing columns in a work newsletter that harassed a female colleague by questioning her competence, her looks, and her commitment to being a police officer because she was a woman, this conduct would not be offensive to a reasonable person. The court in Rabidue v. Osceola Refining Co. used the reasonable-person standard to compare purportedly harassing conduct to what the court held to be a societal norm that accepted the objectification of women through pornography and the media in general.

Some courts have criticized a purely objective reasonable-person standard, particularly in the context of sex discrimination case law. This critique has focused on the reasonable-person standard’s impact on female victims of male harassment because women are largely the victims of sex discrimination. In Steiner v. Showboat Operating Co., the Ninth Circuit found that there is a perception gap between men and women on the subject of sexual harassment, and that “words from a man to a man are differently received than words from a man to a woman.” The Ninth Circuit has also criticized the reasonable-person standard as assessing reasonableness from the harasser’s perspective, rather than the victim’s, thereby downplaying the adverse effect of offensive conduct.

A prominent commentator has also critiqued the reasonable-person standard, claiming that it reifies a sexist community norm to which the facts of the cases are compared. Professor Nancy Ehrenreich has

108. 51 F.3d 591 (5th Cir. 1995).
109. See id. at 594–96.
110. 805 F.2d 611 (6th Cir. 1986).
111. See id. at 620–21.
112. See, e.g., Torres v. Pisano, 116 F.3d 625, 632 (2d Cir. 1997); Burns v. McGregor Elec. Indus., Inc., 989 F.2d 959, 965 (8th Cir. 1993); Ellison v. Brady, 924 F.2d 872, 878 (9th Cir. 1991).
113. See Ellison, 924 F.2d at 879 (citing Note, Sexual Harassment Claims of Abusive Work Environment Under Title VII, 97 HARV. L. REV. 1449, 1459 (1984)).
114. 25 F.3d 1459 (9th Cir. 1994).
115. Id. at 1464.
116. Ellison, 924 F.2d at 878.
argued that a community norm masks women’s lack of power over what norm is adopted and what aspects of life fit into that norm. She contends that the reasonable-person standard allows judges to clothe their decisions in an aura of neutrality while making decisions based on subjective ideals. Professor Ehrenreich argues that the Rabidue court based its conclusion on the false premise that societal norms are egalitarian and nondiscriminatory. Further, cloaking its decision in such norms enabled the Rabidue court to ground its rationale in society’s view of acceptable conduct, thus avoiding a politically charged decision about what kinds of conduct constitute harassment.

B. The Reasonable-Woman Standard

Courts view harassment from a gender-specific perspective when they use the reasonable-woman standard. Although men and women might agree that some blatant conduct constitutes harassment, difficult cases will produce some disagreement. Given this difference in perspective and given statistics indicating that women are more often the victims of sexual violence, several circuits have adopted the reasonable-woman standard to determine when conduct rises to the level of hostile-environment sexual harassment in the workplace.

In Ellison v. Brady, the Ninth Circuit followed the Third, First, and Sixth Circuits when it held that the determination of what conduct is

118. See id. (noting that courts “implicitly assume that sexual discrimination is merely deviant behavior by individuals, rather than a structural problem inherent in American ideology and institutions” by focusing on whether harassing conduct was generally acceptable, “rather than . . . [on] whether it perpetuated conditions of inequality”).

119. See id. at 1189–90 (“[R]elativism translates into neutrality, a refusal to ground judicial decisions on personal preferences for particular perspectives or political judgments about the importance of certain group interests.”).

120. See id. at 1205.

121. See id.

122. E.g., Burns v. McGregor Elec. Indus., Inc., 989 F.2d 959, 965 (8th Cir. 1993); Ellison v. Brady, 924 F.2d 872, 880–81 (9th Cir. 1991).


124. Ellison, 924 F.2d at 879.

125. Torres v. Pisano, 116 F.3d 625, 632 (2d Cir. 1997); Burns, 989 F.2d at 965; Ellison, 924 F.2d at 880–81; Andrews v. City of Phila., 895 F.2d 1469, 1485 (3d Cir. 1990); Lipsett v. Univ. of P.R., 864 F.2d 881, 906 (1st Cir. 1988); Yates v. Avco Corp., 819 F.2d 630, 637 (6th Cir. 1987).

126. 924 F.2d 872 (9th Cir. 1991).
offensive should take into consideration only conduct that a reasonable woman would consider offensive for Title VII purposes. The Ellison court articulated three grounds supporting the reasonable-woman standard. First, it reasoned that this standard “ensures that courts will not sustain ingrained notions of reasonable behavior [protecting the] offender.” Second, the court held that the reasonable-woman standard better prevents the perpetuation of stereotypes that may be ingrained in popular culture and thus better serves the purpose of Title VII. Finally, because Title VII is not based on the perpetrator’s fault, but on employer action, the court determined that the reasonableness standard used need not focus on the motivation of the harassing action, but could focus, as does the reasonable-woman standard, on the harassment’s effect on women.

Even though the reasonable-woman standard has resulted in more findings of hostile-environment sexual harassment than has the reasonable-person standard, criticism of the standard abounds. The Fifth Circuit has found that the reasonable-woman standard is a preferential one. Further, one critic has argued that the reasonable-woman standard does not empower men to prevent possibly harassing conduct because there is no way men can discern what a reasonable woman would find offensive. The reasonable-person standard, the critic claims, may better allow men to prevent their own harassing conduct because it incorporates their viewpoint, thereby enabling them to recognize the conduct as harassing. This critic argues that use of the

---

127. See Andrews, 895 F.2d at 1485; Lipsett, 864 F.2d at 906; Yates, 819 F.2d at 637.
128. Ellison, 924 F.2d at 880; see also Brennan v. Metro. Opera Ass’n Inc., 192 F.3d 310, 320–21 (2d Cir. 1999) (Newman, J., dissenting in part) (comparing sex discrimination to race discrimination and stating that because most white people might not know that some remarks are offensive to most black people, one can see why “the perspective of the reasonable ‘person’ might turn out to be the very stereotypical views that Title VII is designed to outlaw”).
129. Ellison, 924 F.2d at 880–81 (citing Lipsett, 864 F.2d at 898 (quoting Rabidue v. Osceola Ref. Co., 805 F.2d 611, 626 (6th Cir. 1986) (Keith, J., dissenting))).
130. Id. at 881.
131. Id. at 880.
132. Childers, supra note 76, at 894 n.133.
134. DeAngelis, 51 F.3d at 593 (concluding that reasonable-woman standard would “create incentives for employers to bend over backwards in women’s favor for fear of lawsuits”).
135. See Bittner, supra note 76, at 132.
136. Id.
reasonable-woman standard conveys the message that sexual harassment is not a serious issue, important to both men and women, but that courts will remedy it only to placate oversensitive women. Those who criticize the reasonable-woman standard share the belief that the reasonable-person standard is a non-preferential standard that aims to enforce equality rather than preference.

C. Harris v. Forklift Systems, Inc. Does Not Resolve the Circuit Split Regarding the Appropriate Reasonableness Standard

The U.S. Supreme Court first broached the issue of reasonableness in the workplace discrimination context in its Harris decision in 1993. However, the Court granted certiorari in Harris to determine only whether proof of psychological harm was necessary to prove subjective offensiveness and thus did not directly address the standard to prove objective offensiveness. The Court stated in dicta that the standard used to determine whether conduct was abusive or hostile must be objective based on what the reasonable person would find hostile or abusive given the totality of the circumstances. Although the Court’s reference to all of the circumstances suggested a subjective component to the standard, it did not address the use of the reasonable-woman standard, even though the parties briefed and argued this issue. The Court failed explicitly to include or exclude gender as one of the circumstances important to determining objective offensiveness. It thus failed to elaborate on the extent to which courts can take gender into account, even though it was aware of the circuit split on this issue.

After Harris, one circuit changed from the reasonable woman to the reasonable-person standard, another refused to adopt the reasonable-woman standard, and others applied a reasonable-person-in-the-

137. See id.
140. Id. at 20.
141. Id. at 21–23.
142. See Brief for Petitioner at 34–35, Harris, 510 U.S. 17 (No. 92-1168); Transcript at 12, Harris, 510 U.S. 17 (No. 92-1168).
143. See Brief for Petitioner at 35, Harris, 510 U.S. 17 (No. 92-1168).
144. See id.
145. See Richardson v. N.Y. State Dep’t of Corr. Serv., 51 F.3d 591, 593 (5th Cir. 1995).
plaintiff’s-position standard rather than a reasonable-woman standard.\textsuperscript{147} The Eighth and Fifth Circuits have elected to apply the reasonable-person standard because they believe \textit{Harris} requires it.\textsuperscript{148} According to these circuits, when the Court stated that the standard should be based on the reasonable person, it foreclosed any other reasonableness standards.\textsuperscript{149} Other circuits have applied the reasonable-person-in-the-plaintiff’s-position standard by reasoning that \textit{Harris} referred to this standard when it stated that the objective element of proof may be met by utilizing a reasonable-person standard.\textsuperscript{150}

While the Ninth Circuit has followed \textit{Harris} in applying an objective standard when deciding Title VII cases, it has used the reasonable-woman standard as its objective standard.\textsuperscript{151} Since \textit{Harris}, district courts in the Ninth Circuit have used jury instructions that refer to the reasonable person in the plaintiff’s position, but the court of appeals has interpreted this standard to include a reasonable woman analysis.\textsuperscript{152} In \textit{Steiner v. Showboat Operating Co.},\textsuperscript{153} the court concluded that \textit{Ellison v. Brady} directs it to apply a reasonable-woman standard in the employment discrimination context.\textsuperscript{154} In \textit{Fuller v. City of Oakland},\textsuperscript{155} the Ninth Circuit approved a reasonable-person-with-the-same-fundamental-characteristics standard, but then analyzed the facts of the case under the reasonable-woman standard, comparing the facts of the case to what the

\textsuperscript{147} See Brown v. Hot, Sexy, & Safer Prods., Inc., 68 F.3d 525, 540 (1st Cir. 1995); West v. Phila. Elec. Co., 45 F.3d 744, 753 (3d Cir. 1995).

\textsuperscript{148} Gillming v. Simmons Indus., 91 F.3d 1168, 1172 (8th Cir. 1996); DeAngelis, 51 F.3d at 594.

\textsuperscript{149} See Gillming, 91 F.3d at 1172; DeAngelis, 51 F.3d at 594.

\textsuperscript{150} Brown, 68 F.3d at 540 (“The court must consider not only the actual effect of the harassment on the plaintiff, but also the effect such conduct would have on a reasonable person in the plaintiff’s position.”); West, 45 F.3d at 753 (concluding that discrimination must have “detrimentally affected a reasonable person of the same protected class in that position.”).

\textsuperscript{151} Steiner v. Showboat Operating Co., 25 F.3d 1459, 1464 (9th Cir. 1994) (finding that it must consider what is offensive and hostile to reasonable woman).

\textsuperscript{152} Monteiro v. Tempe Union High Sch. Dist., 158 F.3d 1022, 1033 (9th Cir. 1998); see also Fuller v. City of Oakland, 47 F.3d 1522, 1527 (9th Cir. 1995); Steiner, 25 F.3d at 1464. The Ninth Circuit is not alone in continuing to apply a more contextualized standard than the reasonable-person standard espoused in \textit{Harris}. E.g., King v. Frazier, 77 F.3d 1361, 1363 (Fed. Cir. 1996) (noting that objective inquiry “require[s] that sexual harassment be judged from the perspective of the one being harassed”); Brown, 68 F.3d at 540; West, 45 F.3d at 753; Dey v. Colt Constr. & Dev. Co., 28 F.3d 1454, 1466 (7th Cir. 1994) (“We thus consider not only the actual effect of the harasser’s conduct on his victim, but also the effect similar conduct would have had on a reasonable person in the plaintiff’s position.”).

\textsuperscript{153} 25 F.3d 1459 (9th Cir. 1994).

\textsuperscript{154} Id. at 1464 (citing Ellison v. Brady, 924 F.2d 872, 878 (9th Cir. 1991)).

\textsuperscript{155} 47 F.3d 1522 (9th Cir. 1995).
reasonable woman would feel.\textsuperscript{156} In a later case, the Ninth Circuit deemed a jury instruction that referred to a reasonable person with the same fundamental characteristics as the plaintiff, rather than to a reasonable woman, not to constitute an abuse of discretion.\textsuperscript{157} The court reasoned that this instruction incorporated the standards set out in both the U.S. Supreme Court’s decision in \textit{Harris} and its own in \textit{Ellison}.\textsuperscript{158} Thus, when discussing whether conduct is objectively hostile under Title VII, the Ninth Circuit uses conflicting terminology in upholding the use of the reasonable-person-with-the-same-fundamental-characteristics standard while referring to \textit{Ellison} and the reasonable-woman standard.

When using the reasonable-person-in-the-plaintiff’s-position standard, courts take the plaintiff’s particular characteristics into account.\textsuperscript{159} In \textit{Oncale v. Sundowner Offshore Services, Inc.},\textsuperscript{160} the U.S. Supreme Court stated that the objective severity of allegedly harassing conduct should account for the social context in which conduct occurs and is experienced.\textsuperscript{161} Although the Court cited \textit{Harris} for its standard,\textsuperscript{162} it did not use the exact terminology used in \textit{Harris}. Instead, \textit{Oncale} referred to the “reasonable person in the plaintiff’s position,”\textsuperscript{163} and it did not state whether the reasonable-person-in-the-plaintiff’s-position standard is a new standard altogether or simply a new name for the reasonable-person standard. Circuit courts have used a similar standard but referred to the reasonable person of the same protected class,\textsuperscript{164} or the reasonable person with the same fundamental characteristics.\textsuperscript{165} Under these standards, courts consider the totality of the circumstances when determining the objective severity of alleged harassment.\textsuperscript{166}

While the \textit{Harris} Court could have addressed the issue of the appropriate reasonableness standard in sexual harassment cases, it did not speak to the issue directly and different standards have abounded in

\begin{itemize}
  \item \textsuperscript{156} \textit{Id.} at 1528 (“[T]he fact that she couldn’t escape would lead a \textit{reasonable woman} to feel her working environment had been altered.”) (emphasis added).
  \item \textsuperscript{157} \textit{Crowe v. Wiltel Communications Sys.}, 103 F.3d 897, 900 (9th Cir. 1996).
  \item \textsuperscript{158} \textit{Id.}
  \item \textsuperscript{159} \textit{See Brown v. Hot, Sexy, & Safer Prods., Inc.}, 68 F.3d 525, 540 (1st Cir. 1995); \textit{West v. Phila. Elec. Co.}, 45 F.3d 744, 753 (3d Cir. 1995).
  \item \textsuperscript{160} 523 U.S. 75 (1998).
  \item \textsuperscript{161} \textit{Id.} at 80–82.
  \item \textsuperscript{162} \textit{Id.} at 81.
  \item \textsuperscript{163} \textit{Id.}
  \item \textsuperscript{164} \textit{West}, 45 F.3d at 753.
  \item \textsuperscript{165} \textit{Fuller v. City of Oakland}, 47 F.3d 1522, 1527 (9th Cir. 1995).
  \item \textsuperscript{166} \textit{See id.} at 1527; \textit{West}, 45 F.3d at 756.
\end{itemize}
the circuit courts since then. Some circuits agree with the Fifth Circuit that *Harris* directs them to use a pure reasonable-person standard, rather than a preferential reasonable-woman standard. Others subscribe to the reasonable-person-in-the-plaintiff’s-position standard, which considers the totality of the circumstances. Others, including the Ninth Circuit, have stated that they follow *Harris*, but interpret the case to allow for an objective reasonable-woman standard.

IV. COURTS SHOULD ADOPT A REASONABLE-GIRL STANDARD FOR TITLE IX PURPOSES

Courts should use a reasonable-girl standard to determine whether conduct is objectively offensive such that it creates a hostile environment in a school. A reasonable-girl standard is supported by legal precedent and is consistent with U.S. Supreme Court interpretations of Title IX. Further, it best serves the purposes of Title IX by furthering girls’ equal educational opportunities, and is therefore the best standard to determine whether harassment rises to the level of actionable sexual harassment.

A. Precedent Using a Reasonable-Woman Standard Under Title VII Supports Adopting the Reasonable-Girl Standard Under Title IX

The reasonable-girl standard flows from an analogy to tort law and Title VII jurisprudence. When evaluating a girl’s reaction to harassment in schools, the reasonable-girl standard would take a victim’s sex and age into consideration when determining what conduct is objectively offensive under Title IX. This perspective would create incentives for boys and school administrators to learn what girls consider offensive and to remedy that conduct. Because a “gender-neutral” reasonable-person standard tends to ignore the experiences of females, the reasonable-girl standard is needed to address this issue and allow girls to participate in school activities on an equal footing with boys. Further, the reasonable child standard in tort law\(^{167}\) reflects that a child’s experience and knowledge base are sufficiently unique that evaluation of children’s conduct requires a standard other than the generic reasonable-person standard. Because the need for a separate reasonableness standard reduces as a child matures, the reasonable-girl standard would only apply to students who are in primary or secondary education. This standard

\(^{167}\) See supra note 104 and accompanying text.
would remain objective because it would not account for the oversensitive or idiosyncratic girl who might find conduct offensive when a reasonable girl would not. Thus, the reasonable-girl standard would require that administrators consider whether the reasonable girl would find the harassment offensive, not whether the administrators themselves would find it offensive. Where a reasonable girl would find the conduct to be severe, pervasive, and objectively offensive, schools would be required to remedy the harassment.

Because both Title IX and Title VII cases conclude that sexual harassment constitutes sex discrimination, a reasonableness standard used under Title VII should be applicable under Title IX. Title VII cases using the reasonable-woman standard provide persuasive authority for the adoption of a reasonable-girl standard in the Title IX context because the U.S. Supreme Court and courts of appeals have used Title VII case law to define many other key Title IX terms and to recognize particular types of claims. For instance, the U.S. Supreme Court looked to Title VII case law to recognize peer-to-peer hostile-environment sexual harassment claims under Title IX. Further, the U.S. Supreme Court’s holding that objective offensiveness should be determined by looking at the totality of the circumstances does not reject the reasonable-woman standard. Looking to the reasonable woman of Title VII and the Ninth Circuit’s reasonable person with the plaintiff’s fundamental characteristics, courts should adopt the reasonable-girl standard in the Title IX context.

B. The Reasonable-Girl Standard Is Consistent with U.S. Supreme Court Interpretations of Title IX

The reasonable-girl standard is consistent with the U.S. Supreme Court’s comparison of Title IX to Title VI and its interpretation of Title IX as an implied contract. This standard would only focus on the existence of actionable sexual harassment without altering the standards

169. See supra Part ILC.
172. Fuller v. City of Oakland, 47 F.3d 1522, 1527 (9th Cir. 1995).
173. See supra Parts II.A.–B.
Sexual Harassment in Schools

of school liability. Title IX’s contractual nature demands that schools be notified of possible liability, and Title IX imposes liability only where the schools were deliberately indifferent to the discrimination.\textsuperscript{174} However, Title IX’s contractual nature explains only how schools may become liable for harassment; to determine whether conduct amounts to harassment, the Court has looked to Title VII.\textsuperscript{175} Although the reasonable-girl standard would require schools to remedy more harassing conduct than a gender-neutral standard might, the process by which a school becomes liable would not change. Schools would still have to be notified before they could become liable for sex discrimination. The only change would be in the standard to determine whether the allegedly harassing conduct could be reasonably considered offensive. Where the reasonable-person standard applies a purportedly gender-neutral vision of reasonableness, the reasonable-girl standard would simply apply a gender-specific vision of reasonableness.\textsuperscript{176}

The reasonable-girl standard would not impose additional liability on schools. The key inquiry in a sexual harassment in school claim is whether the harassment rose to the level of actionable discrimination under the statute rather than the appropriateness of the school’s response. Even under the reasonable-girl standard, courts would continue to apply the “clearly unreasonable” standard to administrators’ responses.\textsuperscript{177} The reasonable-girl standard would require the school to apply a new standard to determine whether conduct was objectively offensive, and thus rose to the level of actionable sexual harassment. But once the school has made this decision, essentially any response would satisfy the clearly unreasonable standard, which is highly deferential to schools’ determinations of what response is appropriate.\textsuperscript{178} This deference would remain regardless of what standard was used to determine whether the conduct amounted to actionable harassment in the first place. Therefore, while the reasonable-girl standard does alter the threshold of action, this standard is unlikely to impose additional liability on schools because the clearly unreasonable standard is so deferential to the school’s determination of what response is appropriate.

\textsuperscript{175} See supra Part II.A.–C.
\textsuperscript{176} See supra Part III.B.
\textsuperscript{177} See supra notes 55–60 and accompanying text.
\textsuperscript{178} See Davis v. Monroe County Bd. of Educ., 526 U.S. 629, 650 (1999).
The reasonable-girl standard also would not impose unpredictable liability. The Court’s focus in Title IX cases is on the school’s reaction to discrimination.\textsuperscript{179} Title IX does not require a reasonableness standard that affords the harasser the opportunity to know that he is harassing, as is required in a fault-based tort case, because Title IX relies on the school’s deliberate indifference to harassment rather than on a harasser’s fault to determine liability for peer-to-peer hostile-environment sexual harassment.\textsuperscript{180} Under the reasonable-girl standard, the harasser would not become liable for something he could not have predicted because he would not be the defendant.\textsuperscript{181} A gender-specific reasonable-girl standard would still require that schools be notified of harassment and that schools’ remedial actions be evaluated by the courts, as is required by Title IX.\textsuperscript{182} Therefore, under the reasonable-girl standard schools would not be exposed to liability by conduct that they could not have identified as sexual harassment because it is the school’s reaction to the harassment that is reviewed by courts, not their fault for actually committing the harassment themselves.

C. The Reasonable-Girl Standard Best Fulfills the Purposes of Title IX

The reasonable-girl standard would best fulfill the purposes of Title IX by furthering girls’ equal educational opportunities. Neither the reasonable-person nor the reasonable-person-in-the-plaintiff’s-position standard successfully protect girls’ equal educational opportunities because they do not address harassment that may be prevalent and that girls would consider offensive. The reasonable-girl standard would achieve this goal by focusing on girls’ unique experiences and forcing schools to address girls’ complaints, thus finding offensive, severe and pervasive harassment that so interferes with girls’ educational opportunities that it becomes actionable under Title IX.

\textsuperscript{179} See id. at 651.
\textsuperscript{181} See Ellison v. Brady, 924 F.2d 872, 880 (9th Cir. 1991).
\textsuperscript{182} See Davis, 526 U.S. at 646–47.
1. The Reasonable-Person and Reasonable-Person-in-the- Plaintiff’s-Position Standards Do Not Afford Girls Equal Educational Opportunities

The reasonable-person standard, which uses a community norm to determine objective offensiveness, is inadequate because it considers offensive only harassment that is outside the community norm. Title IX does not merely exempt sex discrimination that is prevalent; it is aimed at eliminating all sex discrimination that limits girls’ equal educational opportunities. Given the statistics indicating the prevalence of sexual harassment in junior high and high schools, the reasonable-person standard is unable to eliminate much of this harassment precisely because it is prevalent and therefore within the status quo. Schools need only respond to harassment that is so severe and pervasive that it interferes with girls’ educational opportunities. In contrast, the reasonable-person standard does not reach harassment that is severe and pervasive when it is prevalent in the school. The community norm is not determinative of whether the harassing conduct interferes with girls’ educational opportunities because conduct may interfere with a girl’s ability to take part in class regardless of whether it is prevalent in the school. Instead, it is the girl’s perception of harassing conduct that determines whether potentially harassing conduct interferes with her educational opportunities and thus frustrates the purpose of Title IX. By measuring this perception against an objective standard, the reasonable-girl standard would not take into account oversensitive or idiosyncratic reactions to conduct.

Not only does the reasonable-person standard allow even severe and pervasive harassment to continue by focusing on what society considers offensive rather than on what girls would consider offensive, it reinforces an often sexist community norm. The reasonable-person standard fails to address root causes because it defines sexual harassment by relying on a sexist community norm and concludes that only conduct that falls outside that norm is objectively offensive. For example, when the Rabidue court applied the reasonable-person standard, it found that potentially harassing conduct was consistent with a societal norm that includes

---

183. See supra Part III.A.
184. Davis, 526 U.S. at 650.
185. See supra Part III.A.
186. See Ehrenreich, supra note 117, at 1205.
pornography and the objectification of women in the media.\textsuperscript{187} The defendant’s conduct was not objectively hostile or abusive precisely \textit{because} this conduct was prevalent in society.\textsuperscript{188} The \textit{Rabidue} court failed to question whether the societal norm it used was a sexist one. Because sexual harassment, like objectification of women in the media, is prevalent in schools,\textsuperscript{189} much of it is considered objectively normal rather than objectively offensive. The reasonable-person standard thereby reinforces sexist community norms rather than remedying sex discrimination.

By focusing on the community norm, the reasonable-person standard ignores the unique experiences of girls that inform their interpretations of potentially harassing conduct. Sexual harassment, sexual coercion, and involuntary sexual intercourse are not uncommon among teenage girls.\textsuperscript{190} Women and girls experience sexual violence more often than males, and men are the perpetrators more than ninety percent of the time.\textsuperscript{191} Because of this unique perspective, girls are likely to interpret potentially harassing conduct by boys differently than boys or men might.\textsuperscript{192} The reasonable-person standard is blind to women’s and girls’ different interpretations of harassing conduct because it ignores prevalent sexual violence against women.\textsuperscript{193} Because societal norms trump girls’ perspectives under the reasonable-person standard, girls will continue to experience severe and pervasive harassment under this standard.

By allowing a baseline of sexual harassment, the reasonable-person standard does not protect girls’ access to equal educational opportunities. This standard’s failure to prevent the emotional distress and reduced academic success\textsuperscript{194} leads to unequal educational opportunities. This in turn allows girls to continue to feel the effects of harassment.\textsuperscript{195} The effects that the thirteen-year-old girl in Federal Way, Washington, experienced included being forced to move to a new school, where she was removed from her friends and school activities, causing her to

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{187} Rabidue v. Osceola Ref. Co., 805 F.2d 611, 620–21 (6th Cir. 1986).
\item \textsuperscript{188} \textit{See id.}
\item \textsuperscript{189} \textit{See supra} Part I.A.
\item \textsuperscript{190} \textit{See supra} Part I.A.
\item \textsuperscript{181} Fineran et al., \textit{supra} note 12, at 56.
\item \textsuperscript{192} \textit{Cf.} Ellison v. Brady, 924 F.2d 872, 879 (9th Cir. 1991) (explaining perception gap in employment-discrimination context).
\item \textsuperscript{193} \textit{See id.}
\item \textsuperscript{194} \textit{See supra} Part I.A.
\item \textsuperscript{195} \textit{See supra} Part I.A.
\end{itemize}
\end{footnotesize}
Sexual Harassment in Schools

become emotionally and academically devastated.196 The fallout from harassment—severe emotional distress, fear, and anxiety that causes absenteeism, tardiness, and reduced work productivity—limits girls’ educational opportunities.197 When girls’ grades are reduced because of on-going harassment, they are not eligible for tangible educational opportunities such as awards, jobs, colleges, and scholarships.198 The cumulative effects of sexual harassment that the reasonable-person standard allows to continue amount to the type of non-physical but effective reduction of educational benefits for girls that the Davis Court expressly held were prohibited by Title IX.199

Although it was mentioned in Oncale, the reasonable-person-in-the-plaintiff’s-position standard200 is ambiguous and does not provide courts and school administrators with sufficient guidance. Circuit courts are split over how to interpret the Oncale Court’s use of this standard.201 This standard does not clarify for courts whether they are to consider only the circumstances that the plaintiff experienced or those circumstances plus the plaintiff’s gender.202 The reasonable-person-in-the-plaintiff’s-position standard takes the totality of the circumstances into account,203 and this may include gender to the extent that a reasonable girl would be the appropriate reasonable person in the plaintiff’s position when the victim is not an adult and is female. However, because the Court has not expressly stated that a gender-based interpretation is appropriate, schools are still left with little guidance. Because the U.S. Supreme Court has not resolved the circuit split regarding the use of the reasonable person or reasonable-woman standard,204 the term “totality of the circumstances” may be interpreted to support the use of a gender-specific or a gender-neutral reasonable-person standard depending on the inclination of a circuit court.

196. See Bartley & Pemberton-Butler, supra note 1.
197. See supra Part I.A.
198. Fineran et al., supra note 12, at 55.
201. See supra notes 164-66 and accompanying text.
204. See supra Part III.B.
2. The Reasonable-Girl Standard Would Protect Girls’ Equal Educational Opportunities

The reasonable-girl standard would offer more protection from the reduced educational benefits caused by sexual harassment, and thus better promote Title IX’s goal of equal educational opportunities. The U.S. Supreme Court stated in *Davis* that all severe, pervasive, and objectively offensive harassment that reduces or limits girls’ access to educational opportunities violates Title IX. By penalizing severe, pervasive harassment that the reasonable girl would find offensive and focusing on girls’ unique perspectives, the reasonable-girl standard would eliminate more sexual harassment, which would in turn reduce the negative effects of the harassment. Under the reasonable-girl standard, the thirteen-year-old girl in Federal Way would not be forced to switch schools to escape being groped and fondled in her school’s hallways because the threshold for action would be lower and administrators would be compelled to take action to remedy harassment. The end result would be more educational opportunities for girls, allowing them to maintain their access to equal educational opportunities.

The reasonable-girl standard would increase administrator action and discourage further harassment in schools by encouraging administrators to view victims’ complaints as credible. Because of shared experiences, a male administrator may identify with a male perpetrator more readily than with a female victim. The administrator must choose between opposing views of harassment: one that views the conduct as harmless amusement and another that sees the conduct as offensive or as a prelude to violence. The reasonable-girl standard would require administrators to credit girls’ perspectives of the harassing conduct, take a second look at complaints of sexual harassment, and work harder at eliminating something that disrupts girls’ benefits from education. Once perpetrators know that complaints of sexual harassment will be considered credible, and that remedial action will be taken, they are less likely to harass

---

206. *Schools would be informed of the reasonable-girl standard through the DOE Guidance on Sexual Harassment,* *supra* note 46, which courts refer to when determining school liability for harassment under Title IX. *E.g.,* id. at 651.
208. *See id.* at 886.
Sexual Harassment in Schools

someone again. Thus, sexual harassment that interferes with girls’ educational opportunities will decrease when courts use the reasonable-girl standard to evaluate complaints of sexual harassment.

V. CONCLUSION

The Davis Court left open the question of which reasonableness standard to apply in Title IX hostile-environment sexual harassment cases. In doing so, the Court exposed schools to unpredictable liability and left girls, common victims of sexual harassment in schools, unprotected. Some circuit courts have already formulated the reasonable-woman standard, a reasonableness standard that is meant to protect women in the workplace from harassment that women consider hostile.

A reasonable-girl standard comports with the U.S. Supreme Court’s interpretation of Title IX as a contract and with the liability scheme that the Court created for the Title IX context. The reasonable-girl standard would not alter the notification requirement or the deliberate-indifference liability scheme that results from Title IX’s contractual nature. Further, schools would still become liable under the reasonable-girl standard only when schools’ remedial responses to harassing conduct were clearly unreasonable. Finally, this standard fits well with a statute such as Title IX that does not inquire into the perpetrator’s fault, but rather demands that schools take appropriate remedial measures.

In order to protect schoolgirls, courts should modify the reasonable-woman standard for the Title IX context by creating a reasonable-girl standard and applying it when analyzing peer-to-peer sexual harassment in school cases. Further, because the reasonable-girl standard would recognize girls’ unique perspective on sexual harassment, it is best capable of reducing the rate of sexual harassment that has reached epidemic proportions. The reasonable-girl standard would view all severe and pervasive harassment that reasonable girls would perceive to be offensive as objectively offensive, thereby addressing all sexual harassment that interferes with girls’ equal educational opportunities. This standard is equipped to place girls back on an even playing field with boys by holding schools liable when they allow girls to be subjected to continued harassment.

209. Harassers often believe that harassment is acceptable at school, that everyone does it, and that they will be able to get away with it. Jordan et al., supra note 13, at 296. Therefore, if school administrators make harassment less common in schools, and they ingrain in their students that harassment is unacceptable at school, harassers will not continue harassing other students.
PROTECTING THE TAX-EXEMPT STATUS OF HOUSING DEVELOPERS PARTICIPATING IN LOW-INCOME HOUSING TAX CREDIT PARTNERSHIPS

Marni Hussong

Abstract: The Low-Income Housing Tax Credit (LIHTC) is an important source of federal funding for developers of affordable housing for low-income persons. Although for-profit and nonprofit developers compete for credits, the federal government reserves ten percent of the credits for nonprofit, tax-exempt developers. Exempt developers often sell the credits to for-profit investors, forming a partnership through which the exempt organization develops the housing and the investors receive tax benefits in exchange for capital contributions. The partnership formation, however, may jeopardize the tax-exempt status of the nonprofit organizations and result in the partnership losing the LIHTC. To maintain exempt status, the Internal Revenue Code requires that organizations be organized and operated to promote a charitable purpose and that no net earnings inure to private individuals. A combination of binding and non-binding authority provides confusing guidelines for exempt organizations seeking to protect their exempt status. This Comment examines the federal requirements for the award of LIHTC and traces the development and application of a two-prong test used by the Internal Revenue Service to determine whether partnership structures jeopardize exempt status. This Comment argues that exempt developers in LIHTC partnerships need binding authority that details the level of control of partnership activities the exempt organization must retain, provides an exception for certain partnership guarantees by exempt organizations that are standard within the development industry, and allows investors to receive private benefits to a greater degree without jeopardizing the organization’s exempt status.

The Low-Income Housing Tax Credit (LIHTC) provides federal tax incentives to encourage private investors to contribute funding for developing housing for low-income households. Since its inception in 1986, the successful program has become the primary source of funding for low-income housing development. Nearly $300 million in credits are available annually for both for-profit and tax-exempt developers of housing for tenants with below average incomes. Although both for-

profit and exempt developers compete for credits, the program maintains a preference for exempt developers by reserving ten percent of the total credits for tax-exempt organizations.6

Tax-exempt, low-income housing developers generally do not owe taxes and cannot directly use credits.7 Exempt developers, therefore, sell the credits to for-profit investors who use the credits to reduce their tax liabilities.8 In addition to the credits, the for-profit investors gain an ownership interest in the project.9 In general, the project is structured as a limited partnership in which the exempt organization serves as the general partner10 and retains a one percent interest, while the investors serve as limited partners,11 obtaining a ninety-nine percent interest in the partnership profits, losses, deductions, and credits.12 As general partner, the exempt organization assumes the partnership liabilities. To qualify for exempt status, the exempt organization must maintain control of the day-to-day activities of the partnership to demonstrate that it is furthering its exempt purpose.13

The resulting partnership may jeopardize the tax-exempt status of the nonprofit developer, which in turn will make the partnership ineligible to use the LIHTC reserved for exempt organizations. The Internal Revenue Code (Code) and corresponding Treasury regulations detail the requirements for federal tax exemption.14 An organization may receive and retain exempt status upon a showing that a charitable purpose was the primary motivation for forming the organization and that such purpose is furthered by the daily operations of the organization.15 To

7. MICHAEL I. SANDERS, JOINT VENTURES INVOLVING TAX-EXEMPT ORGANIZATIONS § 12.2(c), at 417 (2d ed. 2000).
8. Id. The credits attract investors who purchase them for a reduced price. For example, in 1995, credits sold for forty to sixty cents on the dollar. Kawecki & Friedlander, supra note 4, at Part II(2)(B).
9. SANDERS, supra note 7, § 1.5, at 7.
12. SANDERS, supra note 7, § 12.2(c), at 418.
13. Id. § 1.7, at 10.
15. 26 C.F.R. § 1.501(b)–(c).
determine whether the partnership between the for-profit investors and exempt developer violates the requirements of the Code for retaining exempt status, the Internal Revenue Service (IRS) developed a two-pronged “strict scrutiny” test. The test focuses on the organization’s charitable purpose and control of the venture, as well as any resulting private benefits.

An examination of federal requirements for exempt status, General Counsel Memorandums, Private Letter Rulings, Revenue Rulings, and limited case law suggests that many LIHTC partnership agreements actually place the exempt status of organizations at risk. In 1995 the Ninth Circuit, in Housing Pioneers, Inc. v. Commissioner, revoked the tax-exempt status of an organization engaged in an LIHTC partnership based on the exempt organization’s lack of control over partnership activities. In the following two years, the IRS issued three non-binding Private Letter Rulings stating that guarantees committing an exempt organization’s assets to protect the assets of the for-profit investor would produce an impermissible private benefit resulting in a loss of exempt status.

Currently, many uncertainties exist about the degree of control the exempt developer must retain over the partnership, the types of guarantees the exempt developer can provide to the partnership, and the types of benefits that can be gained by for-profit investors without

---

17. Id.
18. A General Counsel Memorandum (GCM) provides advice on a specific substantive or procedural tax issue. Internal Revenue Manual (37) 121 § 1(b) (1983). GCMs are for research purposes for the Internal Revenue Service (IRS) and are not statements of the IRS’s position on issues. Id. § (1)(c).
19. A Private Letter Ruling (PLR) is a written statement to a taxpayer interpreting and applying tax law to a set of facts. Rev. Proc. 2000-1, 2000-1 I.R.B. 11. Taxpayers may rely on PLRs responding to their specific inquiries, but may not rely on another taxpayer’s PLR. Id.
20. A Revenue Ruling is the IRS’s conclusion on how a law is applied to specific facts. Id. at 11. Revenue Rulings are not binding precedent on the Tax Court but may acquire the force of law if the ruling details accepted administrative practices. Am. Campaign Acad. v. Comm’r, 92 T.C. 1053, 1070 (1989). Taxpayers may rely on Revenue Rulings when the facts of the taxpayer’s situation are similar. Id.
21. 58 F.3d 401 (9th Cir. 1995).
22. Id. at 404.
24. Due to the lack of assets of exempt organizations, investors often seek indemnities and guarantees obligating the exempt organization to use available assets to make additional capital contributions or cover unforeseen costs. Klein & Rubin, supra note 2, at 92.
sacrificing the developer’s exempt status. The confusion for organizations in LIHTC partnerships seeking to protect their exempt status is due to a combination of binding and non-binding authority, each applying one or more of several tests to a variety of types of partnerships.

The IRS must give clear guidance to exempt organizations in LIHTC partnerships. Those partnerships that received credits shortly after the program was established in 1986 will realize the end of the minimum fifteen-year compliance period in 2001. If the IRS determines that the exempt organizations have jeopardized their status, the partnerships will be ineligible for the LIHTC and ultimately the development of housing for low-income persons will decline.

Part I of this Comment examines the federal requirements both for the award of LIHTC for financing housing for low-income persons and for exempt status under § 501(c)(3) of the Code based on the creation and commitment to a charitable purpose. Part II details the development of the two-part “strict scrutiny” test for determining whether the exempt status of an organization in a partnership is at risk. Part III explains that the current law surrounding exempt-status determinations of organizations in partnerships does not offer clear guidance to nonprofit developers in LIHTC partnerships. Part IV argues that the IRS should issue a Revenue Procedure as binding authority providing guidance to exempt developers using the LIHTC. In particular, the IRS should address control over daily management activities and ultimate voting control, guarantees by the developer that are standard in the industry but may place charitable assets at risk, and the degree to which private benefit may inure without adversely affecting the organization’s exempt status.

I. TO QUALIFY FOR THE LIHTC EXEMPT DEVELOPERS MUST SATISFY ELIGIBILITY CRITERIA AND EXEMPT-STATUS REQUIREMENTS

To qualify for the ten percent LIHTC set-aside, exempt developers must comply with the Internal Revenue Code requirements governing

25. See SANDERS, supra note 7, § 12.2(a), at 416.
eligibility for receipt of the LIHTC and exemption from federal income tax. Exempt developers of housing for low-income persons are eligible for the LIHTC if they commit to serving low-income households\textsuperscript{28} for a minimum of fifteen years,\textsuperscript{29} retain an ownership interest in the partnership, and actively participate in the management and operation of the project.\textsuperscript{30} An organization developing housing for low-income persons is eligible for exemption from federal income taxation upon proof that the organization was created to and continues to further a charitable purpose without providing a private benefit to any individual or entity.\textsuperscript{31}

\textbf{A. Qualification Requirements for Exempt Organizations Applying for LIHTC}

The federal government allocates the LIHTC to designated state agencies to award credits to developers of housing for low-income families.\textsuperscript{32} Because they generally owe no taxes and cannot directly use the tax credits,\textsuperscript{33} tax-exempt recipients may sell the credits to private investors who reduce their tax liabilities over a ten-year period.\textsuperscript{34} This transaction generally results in the formation of a partnership between the exempt “developer” and the for-profit “investor.”\textsuperscript{35} To be awarded and maintain the credits, a recipient must comply with various Code requirements. First, the Code requires that projects receiving credits restrict unit rental to households with below-average incomes and limit rental prices to less than thirty percent of the household’s monthly income.\textsuperscript{36} Second, developers must maintain these rent restrictions for a minimum compliance period of fifteen years.\textsuperscript{37} The recipient must also

\begin{itemize}
  \item \textsuperscript{28} 26 U.S.C. § 42(g).
  \item \textsuperscript{29} Id. § 42(i)(1).
  \item \textsuperscript{30} Id. § 42(h)(5)(B).
  \item \textsuperscript{31} See id. § 501(a)–(c)(3).
  \item \textsuperscript{32} See SANDERS, supra note 7, § 12.2(d), at 423.
  \item \textsuperscript{33} Id. § 12.2(c), at 417.
  \item \textsuperscript{34} See 26 U.S.C. § 42(i)(1).
  \item \textsuperscript{35} See supra notes 9–12 and accompanying text. This Comment refers to a nonprofit partner as a “developer” and a for-profit partner as an “investor.”
  \item \textsuperscript{36} 26 U.S.C. § 42(g). For income limits in each metropolitan area see generally U.S. DEPARTMENT OF HOUSING AND URBAN DEVELOPMENT, FY 2000 HUD INCOME LIMITS: BRIEFING MATERIAL (2000). HUD defines “low-income” as below eighty percent of the area’s median income. Id. at 1.
  \item \textsuperscript{37} 26 U.S.C. § 42(i)(1).
\end{itemize}
agree to an extended low-income commitment providing further protections to low-income residents including an additional fifteen-year obligation to maintain the restricted rents and a prohibition on the refusal to rent to tenants receiving federally subsidized housing assistance. Finally, the nonprofit organization must “materially participate in the development and operation of the project throughout the compliance period.” An exempt organization will generally satisfy the “material participation” requirement if it serves as the general partner of the partnership owning the project.

Although the LIHTC qualification criteria are the same for both for-profit and nonprofit developers, the Code states a preference for allocating credits to exempt organizations. In particular, the Code requires that states award at least ten percent of the total credits to projects sponsored by “qualified nonprofit organizations.” The Code provides a three-part definition for “qualified nonprofit organizations.” First, the organization must have § 501(c)(3) status. Second, the organization may not be “affiliated with” or “controlled by” a for-profit organization. Finally, one of the exempt purposes of the organization must be “fostering” low-income housing.

B. Exempt Developers Awarded the LIHTC Must Demonstrate that They Were Created for and Continue To Further Charitable Purposes

Organizations meeting the requirements of the Code and the corresponding Treasury Department Regulations (Regulations) may apply for exemption from federal income tax. Section 501 of the Code provides for exemption of certain organizations that were created to and continue to further charitable purposes. The regulations provide a
definition for “charitable” purposes and detail methods for determining whether an organization is organized and operated for such purposes.49

1. The Code Establishes the Eligibility Requirements for Exempt Status

Organizations receive exemption from federal income taxation under § 501(a) of the Code if they meet the requirements detailed in § 501(c).50 Developers of housing for low-income persons apply for exemption as charitable organizations under § 501(c)(3).51 To qualify, a developer must be “organized and operated exclusively for . . . charitable . . . purposes . . . [and] no part of the net earnings of [the developer may inure] to the benefit of any private shareholder or individual.”52 If the organization profits from activities not related to its charitable purpose, it will be taxed on that portion of its income53 and may lose its exempt status.54

2. Treasury Regulations Define Key Terms Relating to Tax-Exempt Status Determinations

The Regulations provide binding authority for interpreting the requirements of § 501 of the Code. First, the Regulations exempt developers of housing for low-income persons if they have a “charitable” purpose under the Code.55 For example, relief of the “poor and distressed or of the underprivileged” is a charitable purpose.56 Another charitable purpose is the promotion of social welfare by organizations designed to “i) lessen neighborhood tensions; ii) to eliminate prejudice and discrimination; . . . [and] iv) to combat community deterioration . . . ”57

51. Id. § 501(c)(3).
52. Id.
53. Id. § 501(b).
54. Tax-exempt organizations will lose exempt status if more than an insubstantial part of their activities are not charitable. See generally 26 C.F.R. § 1.501(c)(3).
55. See id. § 1.501(c)(3)–1(d)(1)(b).
56. Id. § 1.501(c)(3)–1(d)(2).
57. Id.
In addition to having a charitable purpose, a developer seeking exempt status must satisfy the organizational test detailed in the Regulations. Under this test, an organization’s articles of organization must establish a charitable purpose, obligate the organization to further an exempt purpose as its primary activity, and dedicate the assets to a charitable purpose upon dissolution of the entity. In particular, the articles must limit the purposes of the organization to one or more exempt purposes and prohibit the organization from engaging in activities that do not further an exempt purpose unless such activities are insubstantial. Even if the members orally commit or the actual activities of the organization operate to further one or more exempt purpose, a developer will fail the organizational test if the articles do not obligate the organization to further its exempt purpose.

Finally, an organization seeking exempt status must also operate exclusively for one or more exempt purposes. The organization will not meet this test if more than an insubstantial part of its activities does not further an exempt purpose. If an organization operates to further an exempt purpose but substantially engages in another activity with a non-exempt purpose, it will fail the operational test and be ineligible for exempt status. Another factor used in determining if the operational test is met is whether an organization’s “net earnings inure in whole or in part to the benefit of private shareholders or individuals.”

Nonprofit organizations seeking to use the LIHTC to develop housing for low-income households must comply with federal requirements for

58. Id. § 1.501(c)(3)–1(a).
59. The term “articles of organization” refers to any instrument written to create the organization. Id. § 1.501(c)(3)–1(b)(2).
60. Id. § 1.501(c)(3)–1(b)(1)(i)(a).
61. Id. § 1.501(c)(3)–1(b). In addition, tax-exempt organizations may not be organized to influence legislation or participate in political campaigns. Id. § 1.501(c)(3)–1(b)(3).
62. Id. § 1.501(c)(3)–1(b)(1)(i)(a) to (b). “Neither the Internal Revenue Code, the regulations, nor the case law provide a general definition of ‘insubstantial’ for purposes of section 501(c)(3).” Living Faith, Inc. v. Comm’r, 950 F.2d 365, 371 (7th Cir. 1991).
63. 26 C.F.R. § 1.501(c)(3)–1(b)(1)(iv).
64. Id. § 1.501(c)(3)–1(a)(1).
65. Id. § 1.501(c)(3)–1(c)(1); see also Better Bus. Bureau of Wash. D.C., Inc. v. United States, 326 U.S. 279, 283 (1945) (“The presence of a single [non-exempt] purpose, if substantial in nature, will destroy the exemption regardless of the number or importance of truly [exempt] purposes.”).
66. 26 C.F.R. § 1.501(c)(3)–1(c)(2).
67. Id. § 1.501(a)–1(c).
the award of LIHTC and for exempt status under the Code. To receive the credits, an organization must commit to restricting rents for thirty years\textsuperscript{68} and to participating in the operation of the project.\textsuperscript{69} The organization must also maintain its § 501(c)(3) status.\textsuperscript{70} The Code requires that the organization have a charitable purpose and that it be organized and operated further that purpose.\textsuperscript{71}

II. IRS AND JUDICIAL DETERMINATIONS OF EXEMPT STATUS OF NONPROFIT ORGANIZATIONS IN PARTNERSHIPS

Historically, the ability of an organization to receive and maintain exempt status hinged on whether the organization was organized and operated to further an exempt purpose under § 501(c)(3) of the Code. Of particular importance to housing developers is that § 501(c)(3) of the Code permits developing affordable housing to be a charitable purpose. As partnerships between exempt organizations and for-profit partners became more common, a series of IRS rulings and court decisions formed a two-part test to determine whether partnerships jeopardize an organization’s exempt status. The IRS’s application of this two-part test to various exempt organizations created an assortment of binding and non-binding authority for developers in LIHTC partnerships seeking to protect their exempt status. In addition, the courts have developed a separate “commerciality doctrine” that prohibits exempt organizations from competing with for-profit entities. Developers must satisfy both the two-part test and the commerciality doctrine to maintain exempt status.

A. The IRS Initially Established that the Development of Housing for Low-Income Persons Can Be a Charitable Purpose

Prior to 1968, the IRS provided little guidance as to whether housing developers were eligible for exempt status. During the late 1960s and 1970s two Revenue Rulings and a U.S. Tax Court decision offered guidance to determine whether providing housing to low-income persons

\textsuperscript{69} Id. § 42(h)(5)(B).
\textsuperscript{70} Id. § 42(h)(5)(C)(i).
\textsuperscript{71} 26 C.F.R. §1.501(c)(3)–(a)(1).
qualifies as a charitable purpose under the Code.\textsuperscript{72} When providing examples of charitable purposes, the non-binding Revenue Rulings relied on the definition of “charitable” within the Regulations. The Tax Court also required organizations to satisfy the organizational and operational tests from the Regulations.\textsuperscript{73}

In 1968, the IRS issued Revenue Ruling 68-17 and declared that housing developers for low-income persons may qualify as § 501(c)(3) organizations.\textsuperscript{74} The IRS considered the exempt status of a nonprofit organization that instituted a housing program for low-income persons and provided information about the program to other organizations.\textsuperscript{75} The organization acquired, rehabilitated, and sold or leased deteriorating residential buildings in distressed neighborhoods.\textsuperscript{76} The Ruling indicated that the purpose and activities of the organization are charitable to the extent community deterioration is reduced by providing housing to low-income persons.\textsuperscript{77}

Two years later, the IRS released Revenue Ruling 70-585,\textsuperscript{78} providing more specific analysis of types of housing organizations qualifying for exempt status. Similar to Revenue Ruling 68-17, this Ruling relied on the definition of “charitable” in the Regulations. In particular, the IRS found numerous purposes exempt under § 501(c)(3) including (1) relieving the poor by providing housing to those who could otherwise not afford it, (2) eliminating prejudice and discrimination by developing mixed-income buildings, and (3) combating community deterioration by rehabilitating residential buildings.\textsuperscript{79}

Several years later the Tax Court further clarified the activities related to housing development that are considered charitable under the Code. In \textit{B.S.W. Group, Inc. v. Commissioner},\textsuperscript{80} the court applied the organizational\textsuperscript{81} and operational\textsuperscript{82} tests of the Regulations to revoke the tax-

---

\textsuperscript{73} \textit{B.S.W. Group}, 70 T.C. at 356.
\textsuperscript{74} Rev. Rul. 68-17, 1968-1 C.B. 247.
\textsuperscript{75} \textit{Id}.
\textsuperscript{76} \textit{Id}.
\textsuperscript{77} \textit{Id}.
\textsuperscript{79} \textit{Id}.
\textsuperscript{80} 70 T.C. 352 (1978).
\textsuperscript{81} \textit{See supra} notes 58–63 and accompanying text.
\textsuperscript{82} \textit{See supra} notes 64–67 and accompanying text.
exempt status of an organization providing consulting services for a fee to nonprofit organizations engaged in rural development.\(^{83}\) Although B.S.W. Group met the organizational requirements of the Code, the Tax Court held that the organization failed the operational test by not operating exclusively for charitable purposes under § 501(c)(3) because its primary purpose was operating a commercial for-profit business.\(^{84}\) Thus, to qualify as an exempt organization under § 501(c)(3), a low-income housing developer must satisfy the charitable-purpose requirement, the organizational test, and the operational test.

**B. Development of a Two-Prong Test To Determine Whether Organizations Partnering with For-Profit Investors Qualify for Exempt Status**

Even if an organization qualifies as an exempt § 501(c)(3) organization under the above tests, the organization might lose its exempt status if it partners with for-profit investors. The IRS initially took a firm position that partnerships between for-profit and nonprofit entities would result in the loss of tax-exempt status.\(^{85}\) The Ninth Circuit rejected this position in *Plumstead Theatre Society, Inc. v. Commissioner*.\(^{86}\) The IRS then released a memorandum containing a two-prong test to determine whether or not a charitable purpose is fostered by the partnership.\(^{87}\)

In 1975, the IRS established a per se rule against joint ventures between for-profit investors and nonprofit organizations in General Counsel Memorandum 36,293.\(^{88}\) This Memorandum addressed a partnership formed to provide housing to low- and middle-income households.\(^{89}\) The IRS stated that a limited partnership with an exempt organization as general partner was “legally incompatible with operating exclusively” for a charitable purpose.\(^{90}\) The IRS based this decision on three factors. First, the partnership allowed sharing of profits between

---

83. *B.S.W. Group*, 70 T.C. at 352.
84. *Id.* at 360.
86. 675 F.2d 244 (9th Cir. 1982), aff’d 74 T.C. 1324 (1980).
89. *Id.*
90. *Id.*
private investors and charitable organizations. Second, the nonprofit entity serving as general partner promoted the financial interests of the for-profit partners. Finally, the equity of the private investors was protected while the charitable assets of the general partner were at risk.

In 1982, the Ninth Circuit in *Plumstead Theatre Society, Inc. v. Commissioner* affirmed the Tax Court’s decision to reject the per se rule that exempt organizations lose their § 501(c)(3) status when they partner with for-profit investors. *Plumstead Theatre*, a § 501(c)(3) organization that promoted the performing arts, formed a limited partnership with several investors and took on the role of the general partner to produce a single play. The limited partners provided the necessary capital in exchange for a 63.5% share in the profits from the production. The Ninth Circuit upheld the Tax Court’s decision that the exempt status of the theatre was not jeopardized by the partnership. The Tax Court’s decision was based on four factors. First, the transaction was at arm’s length and for a reasonable price. Second, the organization was not obligated to return the limited partner’s capital. Third, the partnership had no interest in any other plays, eliminating the possibility of a profit motive. Finally, the limited partners had no control over the operation or management of the partnership.

The following year, the IRS released General Counsel Memorandum 39,005, which articulated a two-part test for determining whether the formation of a partnership between for-profit investors and an exempt organization requires revoking an organization’s tax-exempt status.

---

91. *Id.*
92. *Id.*
93. *Id.*
94. 675 F.2d 244, 245 (9th Cir. 1982), aff’g 74 T.C. 1324 (1980).
96. *Plumstead*, 675 F.2d at 244.
97. *Id.* at 245.
98. *Plumstead*, 74 T.C. at 1328.
100. *Plumstead*, 74 T.C. at 1333.
101. *Id.* at 1333–34.
102. *Id.* at 1334.
104. *Plumstead*, 74 T.C. at 1334.
Memorandum 39,005 responded to a request for a ruling on the § 501(c)(3) status of a nonprofit general partner in a limited partnership formed to construct and to operate housing for the elderly and handicapped. The IRS determined that the organization’s exempt status was not jeopardized because under the partnership agreement the organization acted exclusively to further its exempt purpose. The IRS set out a two-pronged “strict scrutiny” test. Under this test, a partnership must (1) further a tax-exempt purpose and (2) be structured to ensure that the exempt organization will act exclusively to further exempt purposes without benefiting the nonexempt partners.

C. Recent Application and Refinement of the Two-Prong Test

Although not always referring to the two-prong test, the IRS has continued to refine the several factors within the test to determine whether an organization will lose its exempt status by participating in an LIHTC partnership. The first prong, testing whether the partnership furthers an exempt purpose (the charitable-purpose prong), was clarified by a 1996 Revenue Procedure with guidelines for developers seeking exempt status for developing housing. The second prong, (the private-benefit prong) requires that (1) the nonprofit entity control the partnership to enable it to further its charitable purpose, (2) the partnership does not obligate the nonprofit to place its charitable assets at risk, and (3) the partnership does not create a private benefit for the for-profit investors. The private-benefit prong is generally the crucial part of the test and has been applied in several cases, General Counsel Memorandums, and Private Letter Rulings.

---

106. Id.
107. Id.
108. Id.
109. Id.
111. Id.
1. Housing Developers Can Satisfy the Charitable-Purpose Prong in Two Ways

Revenue Procedure 96-32 defines a safe harbor for organizations to satisfy the charitable-purpose prong by providing relief to the poor through developing housing for low-income persons.\(^\text{115}\) The Procedure also describes the conditions necessary for an organization not falling within the safe harbor to secure exempt status under the Code.\(^\text{116}\) Both situations in the Procedure focus on whether the housing development qualifies as charitable by providing relief to the poor and distressed.\(^\text{117}\)

To fall within the safe harbor, the organization must reserve at least seventy-five percent of the apartments in each project for low-income tenants.\(^\text{118}\) In addition, low-income persons must actually occupy the units.\(^\text{119}\) The organization may rent the remaining units to families that do not qualify as low-income.\(^\text{120}\) The IRS determined that this mix of income levels “assist[s] in the social and economic integration” necessary to aid in the purpose of relieving “the poor and distressed.”\(^\text{121}\)

To determine whether an organization that does not meet the safe-harbor guidelines can obtain charitable status, the IRS examines whether the organization relieves the poor and distressed.\(^\text{122}\) In making this determination, the IRS examines all of the facts and circumstances surrounding the project.\(^\text{123}\) Factors considered include the percentage of residents with certain income levels, the affordability of the units, the level of community-based board membership, the provision of additional social services, an affiliation with a § 501(c)(3) organization, and the existence of affordability covenants or restrictions.\(^\text{124}\) Therefore,

---

115. Rev. Proc. 96-32, 1996-1 C.B. 717. The safe harbor simplified the process of determining whether housing is a charitable purpose by focusing on relief to the poor and distressed. See id.

116. Id.

117. See id.

118. Id.

119. Id.

120. Id.

121. Id. In addition to relief of the poor and distressed, housing developers that meet any of the charitable purposes listed in § 501(c)(3) may be exempt. Id. The purposes within the Revenue Procedure include, but are not limited to, (1) “[c]ombatting community deterioration,” (2) “[l]essening the burdens of government,” (3) “[e]limination of discrimination and prejudice,” (4) “[l]essening neighborhood tensions,” and (5) “[r]elief of the elderly or physically handicapped.” Id.

122. Id.

123. Id.

124. Id.
organizations that do not satisfy the safe harbor must demonstrate a commitment to providing affordable housing to low-income persons.

2. **Housing Developers Must Meet Three Overlapping Conditions To Satisfy the Private-Benefit Prong**

In addition to satisfying the charitable-purpose prong, nonprofit organizations seeking to maintain their exempt status must satisfy the second prong: furthering a charitable purpose without conferring a private benefit. The three factors involved in the private-benefit prong are the organization’s control over partnership activities, risk to the organization’s assets, and private benefit to the for-profit investors. A chronological analysis of the application of the second prong reveals the degree to which the factors are intertwined.

a. **Early Applications of the Private-Benefit Prong Focused on Incidental Benefits**

During the period after *Plumstead*, the IRS and Tax Court focused on whether an organization’s activities created an impermissible private benefit or merely an incidental benefit. In 1985, the IRS issued General Counsel Memorandum 39,444 in response to a request for a determination of the exempt status of an organization entering into a partnership created to purchase and lease an office building. The memorandum stated that certain incidental private benefits would not jeopardize an exempt organization’s status if these benefits did not affect the organization’s ability to further its exempt purpose. The IRS determined that the threshold question is whether the partnership furthers an exempt purpose. In making this decision, the IRS first evaluated whether the organization’s exempt purpose was furthered by the partnership’s activities. The IRS then analyzed whether the partnership agreement satisfied the control factor of the private-benefit prong by requiring the exempt organization to “act exclusively in furtherance of its exempt goals.”

---

126. See id.
127. See id.
128. Id.
129. Id.
private-benefit factor of the prong, requiring that benefits received by for-profit partners be incidental to the public purposes.

In particular, Memorandum 39,444 considered the composition of the partnership’s board. The IRS feared that the nonprofit general partner would be swayed by the limited partners and that the charitable purpose of the organization would not be protected. To combat this vulnerability, the IRS recommended establishing an independent committee to monitor the board. The committee would decrease the control of the limited partners, who would no longer have a strong presence on the board. The IRS found that minimizing the control of the limited partners would eliminate the potential for the limited partners to create private benefits. If the limited partners are not creating benefits, the benefits are more likely to be viewed as incidental.

Four years after Memorandum 39,444, the Tax Court, in *American Campaign Academy v. Commissioner*, provided further guidance in the application of the private-benefit prong. American Campaign Academy was organized to train students for positions in political campaigns. The IRS discovered that the academy’s graduates all obtained positions with Republican candidates and organizations. Consequently, it determined that the school benefited the private interest of the Republican Party. The Tax Court stated that qualifying for exempt status does not depend solely on an organization’s purpose or its original statement of purpose. Rather, the IRS “look[s] beyond the four corners of the organization’s charter to discover ‘the actual objects motivating the organization and the subsequent conduct of the organization.’”

130. Any private benefit must be both qualitatively and quantitatively “incidental” to the overall public benefit of the activity if the organization is to remain exempt. See Gen. Couns. Mem. 39,862 (Nov. 21, 1991). To be qualitatively incidental, the public benefit must be one that could not be achieved without benefiting private individuals. Id. To be quantitatively incidental, a benefit must be insubstantial in relation to the public benefit. Id.


132. Id.

133. Id.

134. Id.

135. Id.


137. Id. at 1053.

138. Id.

139. Id. at 1053–54.

140. Id. at 1064.

141. Id. (quoting *Taxation With Representation v. United States*, 585 F.2d 1219, 1222 (4th Cir. 1978)).
The court analyzed the private-benefit factor by determining if any of the net earnings were distributed to the benefit of private shareholders or individuals. First, the Tax Court held that if more than an insubstantial part of an organization’s activities do not further an exempt purpose, the organization will fail the operational test under § 501(c)(3). In addition, if the organization benefits designated individuals, founders, shareholders, or family members, the organization does not operate exclusively for exempt purposes. On the other hand, infrequent economic benefits that are incidental to the organization’s charitable purpose are acceptable private benefits. This decision is relevant to LIHTC partnerships because it provides further evidence of the IRS’s position that incidental benefits arising out of partnerships will not jeopardize the exempt status of a nonprofit general partner.

b. Housing Pioneers: The LIHTC Partnership that Failed the Private-Benefit Prong

In Housing Pioneers v. Commissioner, the Ninth Circuit restricted the ability of organizations to maintain their exempt status while participating in LIHTC partnerships with for-profit investors. The court held that Housing Pioneers, a nonprofit organization created to provide innovative, affordable housing and participating in an LIHTC partnership, did not qualify for exempt status. In 1989, it partnered with several for-profit limited partners to develop housing that could be rented at reduced rates due to the entity’s exempt status. The Tax Court held that Housing Pioneers did not qualify for exemption because it had a non-exempt purpose that was “substantial in nature” and because private investors benefited from the partnership.

On appeal to the Ninth Circuit, Housing Pioneers argued that as a matter of law the LIHTC limits the requirements of § 501(c)(3) because

---

142. See at 1065.
143. Id. (noting that when organization’s net earnings benefit private shareholder or individual, it creates non-exempt private purpose).
144. See id. at 1066.
145. 58 F.3d 401 (9th Cir. 1995).
146. Id.
147. Id. at 402.
148. Id. (“This non-exempt purpose was to provide the benefit of both the California § 214 exemption and the federal § 42 credit to partnerships that were not exclusively charitable.”).
149. Id.
it provides tax credits to for-profit companies and automatically creates a private benefit for investors.150 The Ninth Circuit did not rule on this issue.151 Instead, the court held that the nonprofit failed to “materially participate” in the development and operation of the project through activities that were “regular, continuous, and substantial” and therefore did not qualify as a qualified nonprofit organization within the LIHTC.152 Because Housing Pioneers did not qualify as a nonprofit under the LIHTC, the court focused solely on Housing Pioneers’ qualification for § 501(c)(3) status.153 The court upheld the Tax Court’s decision that disqualified Housing Pioneers as a § 501(c)(3) organization because its activities included a substantial non-exempt purpose: private investors realized a benefit from the ability to reduce rents because of the tax exemption without having to depend on the partnership’s assets to cover expenses.154 Although the court did not specifically refer to the strict-scrutiny test, this holding provides an additional example of the types of private benefits that jeopardize exempt status.

c. The IRS Has Found Guarantees To Be Impermissible Under the Private-Benefit Prong

Throughout the 1990s, the IRS examined guarantees by exempt organizations that placed charitable assets at risk to protect the assets of investors. The low value of an exempt organization’s assets—generally unimproved land or dilapidated housing—often causes investors to seek indemnities and guarantees to protect their investment.155 Investors in housing for low-income persons want sufficient security before committing capital to the project.156 They have the money and often dictate and draft the terms of the partnership.157 Frequently, investors offer their investment on a take-it-or-leave-it basis.158 Ultimately,

150. See id. at 403.
151. Id. at 404.
152. Id. at 403–04.
153. Id. at 404.
154. Id. at 402–03.
155. Klein & Rubin, supra note 2, at 92; see also SANDERS, supra note 7, §12.2(c)(iii), at 421.
158. Id.
agreements that obligate exempt organizations to place charitable assets at risk are impermissible because they benefit for-profit investors.

The IRS analysis of impermissible guarantees in a Memorandum and three Private Letter Rulings provide non-binding examples of the IRS’s application of the private-benefit prong to exempt organizations that partner with for-profit investors. The rulings detail several types of guarantees and indemnifications and whether they place the exempt status of organizations at risk. One of the rulings also articulates a new standard for the third factor of the private-benefit prong: that partnerships must not “overly benefit” private, for-profit investors.

In 1991, the IRS issued a General Counsel Memorandum that illustrates the Service’s disapproval of loss reserves from a nonprofit’s assets. In responding to an inquiry from an exempt hospital wishing to form a joint venture, the IRS determined that the venture created a private benefit that was “too great . . . to be considered incidental to the charitable purposes.” One reason cited was that the partnership agreement required the hospital to establish loss reserves financed with hospital funds. The IRS viewed the loss reserve as a risk to the exempt hospital’s charitable assets that jeopardized its exempt status under the second factor of the private-benefit prong.

In 1997, the IRS issued Private Letter Ruling 97-31-038 concerning the participation of a § 501(c)(3) organization in a partnership to provide mixed-income housing. The ruling illustrates that when nonprofit general partners give environmental indemnifications, credit-adjustment guarantees, or return-of-capital guarantees to for-profit limited partners, the nonprofit organization jeopardizes its exempt status.

The nonprofit general partner seeking the Private Letter Ruling provided environmental indemnifications to investors. Under an initial environmental indemnification, the nonprofit assumed all liability for

---

161. A loss reserve is an account providing money to the partnership if it does not have enough revenue to cover costs. Kawecki & Friedlander, supra note 4, Part II(3)(A). Loss reserves are established in advance of any losses and can be from partnership assets or the general partner’s assets. Gen. Couns. Mem. 39,862.
163. Id.
164. See id.
any future environmental costs. The partnership later amended the indemnification to limit liability to gross negligence or willful misconduct by the general partner. Furthermore, environmental assessments had previously been made on each of the partnership’s projects. The IRS ruled that under the amended indemnification the organization’s exempt status was not at risk.

The nonprofit general partner also included credit-adjustment guarantees in the partnership agreement. The partnership amended the provisions so that the exempt organization was not required to cover any credit reductions. Credit-adjustment guarantees are commonly used to assure investors that if the project fails to qualify for any or all of the estimated LIHTC, the projected returns to the investor will still be realized. The amended agreement provided that additional payments by the exempt organization would be regarded as a capital contribution. As a result, the IRS found that the organization’s exempt status was not at risk because the increased capital contribution would be used to carry out the charitable purpose and would be returned to the organization upon dissolution.

Finally, the exempt general partner provided a guarantee to return the investors’ capital if the project was not completed on time. The IRS focused on the fact that three of the projects were already in operation and remaining completion dates were within the organization’s control because it served as the developer of the project. Consequently, the IRS determined that the guarantee did not “overly benefit” investors and did not place the exempt status of the general partner at risk.

Of primary importance was the IRS’s use of “overly benefit” as a standard for prohibited private benefits. Prior to this ruling, the IRS

166. Id.
167. Id.
168. Id.
169. See id.
170. See id. Tax credits are allocated based on estimated expenses that may change as the project proceeds. See SANDERS, supra note 7, § 12.2(h), at 433.
171. See SANDERS, supra note 7, § 17.6(b)(ii), at 565; Michael Sanders, Hot Issues Affecting Partnerships and Joint Ventures Involving Nonprofits, 20 EXEMPT ORG. TAX REV. 93, 100 (1998).
173. Id.
174. Id.
175. Id.
176. Id.
177. See id.
had used a standard of whether the private benefit to limited partners was quantitatively and qualitatively incidental to public benefits from the partnership.\textsuperscript{178} The “overly benefit” standard signified a relaxation of the incidental benefit standard.

In another non-binding ruling, Private Letter Ruling 97-36-039, the IRS specifically addressed the participation of an exempt organization acting as general partner in a limited partnership using the LIHTC. The exempt organization had a 0.15\% general-partner interest and a for-profit corporation held a 0.85\% general-partner interest.\textsuperscript{179} Another for-profit investor owned the remaining ninety-nine percent interest.\textsuperscript{180}

The original partnership agreement was problematic because it obligated the general partners to return capital to the for-profit investors upon an allocation differential in the projected credits, a credit shortfall, or a credit recapture.\textsuperscript{181} The partnership also approved a pledge and security agreement, in which only the nonprofit agreed to pledge its entire interest upon default.\textsuperscript{182} Therefore, the exempt general partner was obligated to surrender its partnership interest if the two general partners could not return the capital of the for-profit investors if there was a reduction in the amount of LIHTC received by the partnership. The parties eventually terminated the pledge and security agreement, which benefited the investors and the for-profit general partner, because it would be exercised if the for-profit general partner did not satisfy its guarantees to return capital to the investors.\textsuperscript{183} Consequently the IRS ruled in favor of maintaining the exempt status of the organization because the nonprofit did not place its assets at risk and did not cause an impermissible private benefit.\textsuperscript{184} This decision offers further evidence that the IRS will deny exempt status to nonprofit general partners who provide standard guarantees such as loss reserves, environmental indemnifications, credit-adjustment guarantees, or return-of-capital guarantees to for-profit limited partners on the basis that the agreements confer an impermissible private benefit.

\textsuperscript{178} Sanders & Cobb, \textit{supra} note 95, at 216. Although no other rulings use this standard, the IRS referred to the “overly benefit” standard in 1995 continuing professional education materials for exempt organizations. Kawecki & Friedlander, \textit{supra} note 4, Part II(2)(B).

\textsuperscript{179} Priv. Ltr. Rul. 97-36-039 (June 9, 1997).

\textsuperscript{180} Id.

\textsuperscript{181} Id.

\textsuperscript{182} Id.

\textsuperscript{183} Id.

\textsuperscript{184} See id.
d. The Private-Benefit Prong Is Met when the Exempt Partner Maintains Control over the Joint Venture Such that It Furthers the Organization’s Exempt Purpose

In 1998, the IRS issued Revenue Ruling 98-15, which determined the extent of control by the tax-exempt partner over the partnership’s activities needed to satisfy the control factor of the private-benefit prong. The ruling focused on whether a nonprofit hospital participating in a joint venture with a for-profit entity continued to qualify for an exemption under § 501(c)(3) of the Code. The IRS ruled that an organization may participate in a partnership without jeopardizing its exempt status if “participation in the partnership furthers a charitable purpose, and the partnership arrangement permits the exempt organization to act exclusively in furtherance of exempt purposes and only incidentally for the benefit of the for-profit partners.”

The ruling detailed two fact patterns. In the first situation, the articles of organization stated that the governing board would consist of three members appointed by the exempt organization and two by the for-profit partners. The document also called for a majority vote (at least three members) to approve “certain major decisions.” Furthermore, the governing documents required that the partnership be operated to further charitable purposes and stipulated that this purpose would override any duty to operate for financial benefit. The second situation differed because it called for six board members, equally split between the exempt organization and the for-profit partners. It also required a majority vote on major decisions, although the list of decisions was shorter than in the first situation. Finally, there was no statement of putting the charitable purpose before the financial interests of the partners.

Applying the two-part test, the IRS stated that the control requirement of the private-benefit prong is met when the exempt partner maintains control over the joint venture such that it furthers the organization’s exempt purpose.

---

186. Id.
187. Id. at 9.
188. Id. at 6.
189. Id.
190. Id.
191. Id. at 6–7.
192. Id.
193. Id. at 7.
sufficient control over the partnership’s activities to demonstrate that the partnership furthers exempt purposes. The IRS limited the determination of satisfying the control factor to whether the organization maintains voting control. The ruling dictated that a fifty-fifty split in the board membership agreement would not allow the exempt organization to make decisions without securing a vote from a member of the for-profit partner. The IRS determined that the inability to control the partnership by initiating programs without a vote from the for-profit member limited the exempt hospital’s power to operate “exclusively for exempt purposes.” Thus, a nonprofit must maintain fifty-one percent voting control to satisfy the control requirement and retain exempt status.

Nonprofit developers in LIHTC partnerships must satisfy the two-prong strict-scrutiny test to maintain exempt status. The IRS issued a safe harbor for developers of housing for low-income persons to satisfy the charitable-purpose prong. The private-benefit prong, on the other hand, consists of three overlapping factors: the ability to maintain control, the risk to charitable assets, and the degree of private benefit. These factors have been the source of numerous General Counsel Memorandums, Private Letter Rulings, and Revenue Rulings.

D. The Commerciality Doctrine Prohibits Exempt Organizations from Participating in Activities Considered Commercial

Apart from IRS rulings, the courts have developed a separate body of law, the commerciality doctrine, which has become important in determining an organization’s exempt status. The commerciality doctrine reflects the belief that it is unfair for exempt organizations to compete with organizations in the for-profit sector. Under the doctrine, a tax-

194. Id. at 8.
195. Id. at 9.
196. Id.
197. See id. at 9.
exempt organization is engaged in a non-exempt activity when that activity is undertaken in a commercial manner.\textsuperscript{201}

The first articulation of the commerciality doctrine came in \textit{Better Business Bureau v. United States}.\textsuperscript{202} The U.S. Supreme Court held that the organization had a “commercial hue” and that its “activities [were] largely animated by this commercial purpose.”\textsuperscript{203} The clearest explanation of the doctrine is found in \textit{Living Faith, Inc. v. Commissioner}.\textsuperscript{204} The Seventh Circuit identified factors indicative of commercial activities: (1) selling goods and services to the public, (2) direct competition with a for-profit, (3) prices common in retail, (4) advertising, (5) promotional materials, (6) hours of operation similar to for-profit, (7) salaried employees rather than volunteers, and (8) lack of charitable contributions.\textsuperscript{205} In both cases the courts affirmed the denial of exempt status based on the commercial aspects of each organization’s activities.\textsuperscript{206}

\section*{III. ONLY PIECMEAL GUIDANCE IS AVAILABLE TO DEVELOPERS WISHING TO MAINTAIN EXEMPT STATUS AND PARTICIPATE IN LIHTC PARTNERSHIPS}

Although the IRS and federal courts have provided guidance to various organizations seeking to maintain exempt status, the lack of comprehensive authority provides only piecemeal guidance for exempt developers wishing to partner with for-profit investors in LIHTC partnerships. The clearest requirement is that all exempt organizations must be organized and operated to further a charitable purpose under § 501(c)(3) of the Code.\textsuperscript{207} In addition, the Treasury regulations guide organizations seeking to obtain or maintain exempt status by providing both a detailed definition of charitable and tests to satisfy the organizational and operational requirement.\textsuperscript{208}

\begin{itemize}
  \item \textsuperscript{201} \textit{Id.} at 629. An act is commercial “if it has a direct counterpart in, or is conducted in the same manner,” as an act of a for-profit organization. \textit{Id.} at 629–30.
  \item \textsuperscript{202} 326 U.S. 279 (1945).
  \item \textsuperscript{203} \textit{Id.} at 283–84.
  \item \textsuperscript{204} 950 F.2d 365 (7th Cir. 1991).
  \item \textsuperscript{205} \textit{Id.} at 372–75; \textit{see also} \textit{Hopkins}, supra note 200, at 640.
  \item \textsuperscript{206} \textit{See Better Business}, 326 U.S. at 286; \textit{Living Faith}, 950 F.2d at 376–77.
  \item \textsuperscript{207} \textit{See} 26 U.S.C. § 501(c)(3) (1994).
  \item \textsuperscript{208} \textit{See} 26 C.F.R. § 1.501(c)(3) (2000).
\end{itemize}
requirements, there is only confusion for the developer hoping to partner with a for-profit investor.

The inconsistency of the IRS and the courts in the method used to determine whether an organization’s exempt status has been jeopardized creates a confusing environment for exempt organizations seeking to protect their status while participating in LIHTC partnerships. Furthermore, many of the decisions do not focus on the particularities of the housing industry. Private Letter Ruling 97-31-038 provides the most complete analysis of the types of guarantees that the IRS will view to be impermissible private benefits.209 Yet, the ruling is of limited value because the risk to the general partner’s exempt status was minimized due to the completion of three of the five projects prior to the ruling.210 This ruling also introduced the standard of “overly benefit,” a more relaxed standard than the incidental-benefit standard used previously.211 It is unclear which standard currently applies and LIHTC partnerships have no guidance as to which one to follow. The result of the numerous tests and various factors relied upon by the IRS and courts in determining exempt status is a lack of comprehensive authority on protecting exempt status for nonprofit developers in LIHTC partnerships.

IV. THE IRS SHOULD ISSUE A REVENUE PROCEDURE TO GUIDE EXEMPT DEVELOPERS APPLYING THE PRIVATE-BENEFIT PRONG TO LIHTC PARTNERSHIPS

To protect the exempt status of developers in LIHTC partnerships, the IRS should issue a Revenue Procedure to provide binding authority for satisfying the private-benefit prong. Without such guidance, developers will be at risk of losing their exempt status, and there will be fewer exempt organizations willing to participate in LIHTC partnerships. First, the IRS should adopt the position that exempt organizations satisfy the requirement of furthering the charitable purpose through control of day-to-day management activities. Second, the charitable-assets test, which prohibits placing charitable assets at risk by providing guarantees to investors, should allow for an exception for guarantees that are standard in the development industry. Finally, the IRS should formally adopt the “overly benefit” rather than the “incidental benefit” measurement to

210. See id.
211. See id.
determine impermissible private benefits. Unless the IRS provides more guidance to protect the exempt status of housing developers, the success of the LIHTC program will be jeopardized.

A. The IRS Should Require Exempt Organizations To Maintain Control of Day-to-Day Management and Not Majority Voting Control on the LIHTC Partnership’s Board

Exempt organizations should be able to satisfy the control requirement of the private-benefit prong of the strict-scrutiny test if they retain control of the day-to-day management activities of the partnership rather than establish voting control of the board. The most recent IRS ruling on the exempt status of an organization in a partnership, Revenue Ruling 98-15, indicates that the IRS will not revoke the exempt status of an organization that maintains fifty-one percent voting control of the partnership board and is controlled by a provision in the partnership agreement obligating the partnership to further charitable purposes. On the other hand, the IRS finds that an organization without majority voting control and no statement of charitable purposes has jeopardized its exempt status. Revenue Ruling 98-15 should not apply to LIHTC partnerships because it confuses rather than clarifies the issue of control required by the private-benefit prong of the test. Securing voting control for certain major decisions, as required by the ruling, does not automatically increase an organization’s ability to further the exempt purpose through the partnership. Rather than applying the ruling to LIHTC partnerships, the IRS should provide binding authority granting exempt organizations the greatest opportunity for promoting charitable purposes without jeopardizing their exempt status.

The first reason the Ruling should not apply to LIHTC partnerships is that it does not consider the quality of the board membership when dictating a fifty-one percent control requirement. For example, if a board member appointed by the investor, but not by the developer, is an expert in housing for low-income persons and is not motivated by profit, the lack of voting control by the nonprofit should not be a critical issue. If

212. See supra note 111 and accompanying text.
214. See id.
215. See supra Part II.C.2.d.
the expert is actually voting to promote housing for low-income persons, the developer’s charitable purpose, then the exempt status of the developer should not be questioned. In General Counsel Memorandum 39,444, the IRS pointed to an independent committee as a favorable factor for exempt status because it would eliminate the degree of control of, and the potential for abuse by, the limited partners.217 Moreover, an internal IRS manual directs agents to look favorably upon the appointment of an independent board of directors, especially if the individuals are involved in low-income housing.218 Therefore, rather than focusing on the majority voting control requirement, the IRS should adopt the position that the appointment of an independent committee comprised of experts in low-income housing will be a favorable factor in protecting exempt status.

The second problem with Ruling 98-15 is that it does not specify the characteristics of major matters subject to the fifty-one percent control requirement. Although the ruling lists seven topics of major decisions, the ruling does not specify how to determine whether a decision is major and subject to the fifty-one percent requirement.219 A fifty-one percent control requirement is logical only if it applies either to all decisions or only to those decisions affecting the charitable purpose of providing housing to low-income persons. The requirement should not apply to resolutions, such as changing the name of the partnership, that would not have any effect on the provision of housing for low-income persons.

The third question arising from the application of Ruling 98-15 to LIHTC partnerships is whether less than fifty-one percent voting power would be acceptable to the IRS if the board members are bound by a statement that charitable purposes take precedent over profit motives.220 The fact pattern approved by the IRS contained such a statement and the scenario rejected by the IRS did not.221 However, the inclusion of a charitable-purpose statement should protect the exempt status of the developer whether or not it has voting control only if the investors prove that they are acting in accordance with the statement and that the partnership is furthering the charitable purpose of the exempt organization. The Tax Court has stated that it is the conduct of the

218. Kawecki & Friedlander, supra note 4, Part I(6).
220. See id.
221. Id.; supra notes 190 and 193 and accompanying text.
organization, not specific statements in the articles, that determine exempt status. IRS officials have also acknowledged that day-to-day management is more important than anything written in the governing documents. Therefore, although a statement of commitment to charitable purposes is favorable, a determination of the exempt status of an organization with less than majority voting control should be based primarily on whether the partnership is furthering the organization’s charitable purpose.

The final shortcoming of Revenue Ruling 98-15 is that if voting control is a safe harbor for exempt status, there is no guarantee that charitable purposes are being furthered. The operational test for evaluating whether an exempt organization’s status is at risk looks beyond the terms of agreements and examines whether the purpose is actually being furthered. If the fifty-one percent control standard is a safe harbor, LIHTC partnerships may maintain their exempt status through securing voting control without “materially participating” in the daily activities of the partnership as required by Housing Pioneers. The amount of control that satisfies the requirement of furthering an exempt purpose should be “material participation” through regular, continuous, and substantial activities as articulated in Housing Pioneers.

The focus of an IRS determination of an organization’s exempt status should be whether the organization “materially participates” in partnership activities and, more importantly, whether the partnership is furthering the charitable purposes of the exempt organization. The IRS should emphasize the importance of an independent committee and evidence that an organization’s activities are furthering an exempt purpose. Similarly, the IRS should not base its decision on whether or not the exempt organization maintains voting control over the decisions of the Board of Directors of the partnership.

222. See Am. Campaign Acad. v. Comm’r, 92 T.C. 1053, 1064 (1989); see also supra notes 140–141 and accompanying text.

223. Barbara Yuill, Tax Policy: Owens to Leave IRS for Private Sector; Updates Key Tax-Exempt and Health Care Issues, 209 D.T.R. G-8 (Oct. 29,1999) (“Day-to-day management . . . is perhaps the most important factor and possibly even more significant than control of the board or written agreements between the parties.”).

224. See 26 C.F.R. § 1.501(c)(3)–1(c) (2000); see also supra notes 140–141 and accompanying text.

225. Hous. Pioneers, Inc. v. Comm’r, 58 F.3d 401, 403 (9th Cir. 1995).

226. Id. at 403–04.
B. In Determining Exempt Status, the IRS Should Allow an Exception to the Charitable-Assets Test for Standard Guarantees by Exempt Organizations to Investors

When determining whether the partnership agreement obligates the nonprofit to place its charitable assets at risk, the IRS should exclude all guarantees between exempt organizations and for-profit investors that are standard within the housing industry. This component of the private-benefit prong has become increasingly important because guarantees and indemnifications are standard tools in providing security for investors in partnership agreements.\(^{227}\) However, the IRS prohibits guarantees that insulate the limited partner’s assets while increasing the risk to the charitable assets.\(^{228}\) To allow for more equal footing between for-profit and exempt organizations competing for the LIHTC, the test should be amended to allow an exception for standard guarantees such as loss reserves, environmental indemnifications, credit adjustments, and return of capital.

1. Loss Reserves Are a Necessary Component of LIHTC Partnerships and Should Not Jeopardize the Exempt Status of an Organization

Loss reserves should not affect the exempt status of organizations in LIHTC partnerships because they are a common method for preparing for the possibility of lower-than-estimated profits.\(^ {229}\) Loss reserves are necessary in LIHTC projects due to the collection of lower rents from low-income persons. In General Counsel Memorandum 39,862, the IRS found that loss reserves established by the exempt organization place charitable assets at risk.\(^ {230}\) The IRS should not view such agreements as placing charitable assets at risk if the exempt organization is entitled to a priority return of capital upon disposition of the property.\(^ {231}\) Under such an agreement, when the partnership sells the property, the assets contributed by the exempt organization would be returned to the organization first, then the remaining assets would be divided according to the partnership interests.

\(^{227}\) See supra notes 155–158 and accompanying text.


\(^{229}\) See supra note 163 and accompanying text.

\(^{230}\) See supra note 163 and accompanying text.

\(^{231}\) See SANDERS, supra note 7, § 17.6(c), at 567.
To accomplish the goals of loss reserves without jeopardizing exempt status, partnership agreements should establish reserves as specific capital contributions to the partnership rather than commitments to protect investors’ assets.232 If they are classified up-front as capital contributions to the partnership, the IRS will not view the reserves as set-asides from the exempt organization’s assets to protect the investor’s assets in the future. Alternatively, the source of the reserves could be the money earned by the exempt organization from developer’s fees rather than directly from the organization’s charitable assets.233

Whether the source of the reserve is developer’s fees or the exempt organization’s assets, the partnership agreement should provide that the exempt general partner is entitled to a priority return of capital upon disposition of the property.234 By prioritizing the return of the capital to the organization, the risk to the charitable assets is minimized. The assets of the exempt general partner that are placed in the loss reserve will be returned to the organization before the remaining assets are divided among the partners. Because the charitable assets are used as a particular contribution and not as a way to protect the assets of the for-profit partners, the IRS should allow LIHTC partnerships to establish loss-reserves agreements without jeopardizing their exempt status.

2. Environmental Indemnifications by an Exempt Organization for the Period of the Property’s Use by a Partnership Should Not Jeopardize Its Exempt Status

Environmental indemnifications for the period during which the property is used for low-income housing should not place the exempt status of an organization at risk because they are common tools in the development industry.235 The IRS should adopt the view that exempt developers who obtain environmental assessments to assure that property is free of contamination and who have control of the project to prevent future environmental damage will not be placing their assets at risk. The IRS’s disfavor of environmental indemnifications for “all losses [attributable] to the presence of hazardous materials at any time on or around the property, whether or not [the presence of the materials was] in

233. Sanders, supra note 171, at 101; see also SANDERS, supra note 7, § 17.6(c), at 567.
234. Sanders, supra note 171, at 101; see also SANDERS, supra note 7, § 17.6(c), at 567.
creates an unfair competitive advantage in favor of for-profit recipients of the LIHTC who can freely give environmental indemnifications.

The IRS has determined that environmental indemnifications for the current condition of property will not jeopardize exempt status, but that indemnifications for future environmental conditions do place exempt status at risk.\textsuperscript{237} The IRS has found that the warranty of the property by the exempt organization at the time of partnership formation is acceptable, especially after a Phase I environmental assessment is completed.\textsuperscript{238} On the other hand, the IRS has found that warranties of future environmental conditions are impermissible because they protect the assets of the investors while placing the exempt developer’s assets at risk.\textsuperscript{239} Because exempt organizations control the activities of partnerships and indemnifications by the general partner are common, the IRS should not consider environmental indemnifications for the period of use for low-income housing in determining whether an organization’s assets are at risk.

3. Guaranteed Returns Despite Credit Adjustments Should Not Affect an Organization’s Exempt Status Because Such Guarantees Are Common in LIHTC Partnerships

Minimum investment returns are standard in LIHTC partnerships to assure investors a reasonable return upon a credit adjustment\textsuperscript{240} and should not jeopardize an organization’s exempt status. As the IRS stated in Private Letter Ruling 97-31-038, minimum investment returns could lead to an organization failing the private-benefit prong and the loss of exempt status.\textsuperscript{241} Limited minimum investment returns should not jeopardize the exempt status of an organization because the LIHTC program, which encourages partnerships between nonprofit organizations and private investors, does not guarantee the amount of credits that a project will receive. If the exempt organization is prohibited from making such an agreement, investors will either turn to for-profit

\textsuperscript{236} Sanders, supra note 171, at 101 (internal citation omitted).
\textsuperscript{237} See Priv. Ltr. Rul. 97-31-038.
\textsuperscript{238} Id.; see also SANDERS, supra note 7, § 17.6(c), at 567.
\textsuperscript{239} Sanders, supra note 171, at 101.
\textsuperscript{240} Priv. Ltr. Rul 97-31-038.
\textsuperscript{241} See id.; see also supra notes 170–172 and accompanying text; Sanders, supra note 171, at 100.
recipients of the LIHTC who can guarantee a return or only purchase the credits at a lower price. Both situations will reduce the amount of capital available to the nonprofit developers.

To be consistent with the LIHTC’s preference for exempt developers, the IRS should promote partnerships between exempt and for-profit partners by allowing such guarantees if capped at the accumulated developer fees. If the guarantee is limited to the developer fees of the exempt organization, then the organization is only placing a limited amount of its charitable assets at risk. Furthermore, if the guarantee is treated as a capital contribution that will be returned to the charitable investor, the risk to the exempt organization’s assets is minimal. Thus, the IRS should not consider standard minimum-investment-return agreements when evaluating the private-benefit prong.


Return-of-capital provisions are a common method of obligating partners to meet partnership agreements and should not place the exempt status of an organization in a LIHTC partnership at risk. Examples of provisions that may result in return of capital are completion guarantees and guarantees of performance goals or activity levels. The IRS’s conclusion that these clauses impermissibly place charitable assets at risk if the provision obligates the developer to return capital out of its funds places an unfair burden on exempt organizations because for-profit developers are not subject to the same restrictions. Outright prohibitions on these provisions for exempt organizations would make partnerships with nonprofit developers less attractive to LIHTC investors and reduce the success of the LIHTC program. Standard provisions in partnership agreements such as loss reserves, environmental indemnifications, minimum investment returns, and

---

243. See SANDERS, supra note 7, § 17.6, at 565.
244. See id. at 568.
246. Sanders, supra note 171, at 100.
247. Sanders, supra note 171, at 100–01.
248. See Kawecki & Friedlander, supra note 4, Part II(2)(B).
return-of-capital agreements should not affect the IRS’s determination of exempt status regarding charitable assets being at risk.

C. The IRS Should Commit to the “Overly Benefit” Standard To Determine the Existence of Impermissible Private Benefit in LIHTC Partnerships

The IRS should formally adopt the “overly benefit” standard for LIHTC partnerships because it allows exempt developers and for-profit investors to join in developing much needed affordable housing for low-income households. In 1982, the Ninth Circuit upheld the Tax Court’s decision to abandon the per se prohibition against private inurement that prohibited joint ventures between exempt and for-profit investors for a new standard of incidental benefits. Although a 1997 Private Letter Ruling dealing with an LIHTC partnership introduced the relaxed standard of “overly benefit,” it is unclear whether the standard would apply to all LIHTC partnerships. Despite not using the “overly benefit” standard in later rulings, an IRS internal manual refers to this more relaxed standard.

The IRS should adopt the “overly benefit” standard prior to the end of the compliance periods of first tax credit properties in 2001. At that time, the partnerships will no longer need to maintain reduced rents, and the investors may wish to increase rents or sell the project. Although most LIHTC partnerships include provisions giving the right of first refusal to the “tax-exempt general partner or other tax-exempt entity,” the effect of sale proceeds on exempt developers is uncertain. As long as the exempt developer purchases the property or reinvests the proceeds in a charitable purpose, the sale will not automatically terminate exempt

249. See Plumstead Theatre, Inc. v. Comm’r, 675 F.2d 244 (9th Cir. 1982), aff’d 74 T.C. 1324 (1980).
250. See supra notes 88–93 and accompanying text.
251. See supra Part II.C.2.c.
253. In this Private Letter Ruling, the IRS focused on the organization’s ability to lessen the burdens on the government. SANDERS, supra note 7, § 4.2, at 143.
255. See Kawecki & Friedlander, supra note 4, Part II(2)(B).
257. SANDERS, supra note 7, § 12.2(l), at 438.
However, the proceeds from a sale realized by investors may be an impermissible private benefit. In order for the exempt organization to maintain its charitable purpose, the housing must be sold for a minimum price as detailed in the Code.\textsuperscript{259} Generally, the property is also subject to an additional fifteen-year compliance period that brings down the fair market value.\textsuperscript{260} If the fair market value is high enough that the investors recognize a considerable return, the exempt status of the developer may be jeopardized.\textsuperscript{261} The “overly benefit” standard would provide added security to developers seeking to protect their exempt status at the end of the compliance period.

This standard should apply to LIHTC partnerships because it will encourage partnerships between exempt developers and for-profit investors. The IRS should balance the public benefit from the LIHTC program and the societal benefits associated with an increased supply of affordable housing for low-income households against the private benefits of certain market-driven operation benefits.\textsuperscript{262} The “overly benefit” standard should be established through a binding Revenue Procedure to provide clear guidance to nonprofit developers and ensure that LIHTC partnerships will continue after the first fifteen-year compliance period.

IV. CONCLUSION

The Low-Income Housing Tax Credit is the primary source of funding for the development of housing for low-income persons. Despite the program’s role in funding such housing, exempt organizations awarded credits and entering into LIHTC partnerships do not have clear guidance for protecting their exempt status. Although the IRS has developed and applied a two-part test to determine whether an organization qualifies for exempt status, much of the authority is focused on the more general § 501(c)(3) requirements, non-binding, or not specific to housing organizations. Furthermore, the nature of the LIHTC program and partnership agreements in the development industry require board voting provisions and guarantees by the exempt organization that may put a

\begin{itemize}
  \item \textsuperscript{258} 26 C.F.R. § 1.501 (c)(3)-1(b)(4) (2000).
  \item \textsuperscript{259} 26 U.S.C. § 42(i)(7).
  \item \textsuperscript{260} SANDERS, supra note 7, § 12.2(1), at 438–39.
  \item \textsuperscript{261} See supra notes 143–144 and accompanying text.
  \item \textsuperscript{262} SANDERS, supra note 7, § 17.6(c), at 571.
\end{itemize}
nonprofit developer’s exempt status at risk. Losing this status would result in the partnership’s loss of the LIHTC and eventually a reduction in the amount of housing for low-income households.

To maintain the involvement of exempt organizations in LIHTC projects, the IRS should issue guidance through a Revenue Procedure. The Revenue Procedure would provide much needed binding authority to exempt organizations that will enable them to continue to partner with for-profit investors. The IRS’s position would not violate the commerciality doctrine because nonprofits would not directly compete with for-profits. In particular, the IRS should specify that exempt organizations in LIHTC partnerships need only to maintain control of daily management activities, may provide guarantees that are common within LIHTC partnerships and the development industry, and may allow for financial rewards that do not “overly benefit” the for-profit investors. If the IRS adopts such standards, the risk to the exempt status of developers in LIHTC partnerships will be minimized and the supply of housing available to low-income households will continue to increase.
THE EQUAL PAY ACT AS APPROPRIATE LEGISLATION UNDER SECTION 5 OF THE FOURTEENTH AMENDMENT: CAN STATE EMPLOYERS BE SUED?

Thane Somerville

Abstract: Congress may constitutionally abrogate state sovereign immunity only through legislation enacted pursuant to Section 5 of the Fourteenth Amendment to the U.S. Constitution. In *Kimel v. Florida Board of Regents*, the U.S. Supreme Court held the Age Discrimination in Employment Act to be inappropriate Section 5 legislation. *Kimel* was the first time the Court held an anti-discrimination statute enacted to protect civil rights inapplicable to the states. Based on the *Kimel* decision, other civil rights statutes, such as the Equal Pay Act (EPA), may face similar challenges. This Comment argues that the EPA is appropriate Section 5 legislation. Unlike recent statutes struck down as inappropriate Section 5 legislation, the EPA does not grant plaintiffs more substantive rights than the Constitution. The EPA is a narrowly tailored statute enacted to prevent gender-based wage discrimination that violates the Fourteenth Amendment. Congress reviewed substantial evidence of gender-based wage discrimination by state employers before it enacted the EPA. Based on this evidence, Congress enacted the EPA to provide a remedy for such prevalent discrimination. Courts should find that the EPA is appropriate legislation under Section 5 to enforce the substantive guarantees of the Fourteenth Amendment’s Equal Protection Clause.

In 1960, three years before Congress enacted the Equal Pay Act (EPA), women annually earned, on average, sixty-one percent as much as men. Factors that partially explained the wage differential included lack of employment opportunities for women and the historical tendency to push women into lower-paying administrative and secretarial positions. However, those factors do not explain why men and women employed in the same occupations were paid differently. After eighteen months of hearings, including testimony from working women and leaders of American industry, Congress enacted the EPA to remedy the serious problem of gender-based wage discrimination in the private sector.

3. *Id.* at 973–74.
Gender-based wage discrimination was not, however, limited to the private sector. In the early 1970s Congress heard extensive testimony that public employers engaged in rampant gender discrimination.\(^7\) Congress reacted to this information by enacting Title IX of the Education Amendments of 1972,\(^8\) which prohibited gender discrimination in all education programs receiving federal funds.\(^9\) Congress also extended the protection of Title VII of the Civil Rights Act of 1964\(^10\) to state employees.\(^11\) Finally, based on the substantial evidence of public-sector wage discrimination gathered by Congress over the preceding four years, Congress amended the EPA in 1974\(^12\) to apply to state employers.\(^13\)

Since the enactment of the EPA, the wage gap has closed, but at an extremely slow pace. In 1998, the median annual earnings for full time working women were seventy-three percent as much as men’s earnings.\(^14\) The availability of the EPA as a remedy for female employees who do not receive equal pay because of gender discrimination plays an important part in eliminating the wage gap. However, female plaintiffs employed in the public sector attempting to vindicate their right to equal pay have recently faced a new hurdle to overcome: the Eleventh Amendment defense of state sovereign immunity.\(^15\)

In *Seminole Tribe v. Florida*,\(^16\) the U.S. Supreme Court held that Congress could not abrogate sovereign immunity except through legislation enacted pursuant to Section 5 of the Fourteenth Amendment.

---

7. See infra note 262 and accompanying text.
13. 29 U.S.C. § 203(d), (x) (1994) (redefining employer to include “public agencies” which include “the government of a State or a political subdivision thereof”); see also id. § 203(c)(2)(c) (redefining employee as “any individual employed by a State, political subdivision of a State or an interstate governmental agency”).
to the U.S. Constitution. One year after *Seminole Tribe*, the Court, in *City of Boerne v. Flores*, limited the scope of appropriate Section 5 legislation to that designed to remedy or prevent unconstitutional conduct. Legislation is not appropriate under Section 5 if it grants plaintiffs more substantive rights than the Constitution because such legislation changes the substance of constitutional protection.

In *Kimel v. Florida Board of Regents*, the U.S. Supreme Court eliminated any assumption that legislation protecting civil rights would automatically be deemed appropriate legislation to enforce the Fourteenth Amendment. In *Kimel*, the U.S. Supreme Court held that Congress could not abrogate state sovereign immunity through the Age Discrimination in Employment Act of 1967 (ADEA). This was the first time that the Court held an anti-discrimination statute enacted to protect civil rights inapplicable to the states. Since *Kimel*, three courts of appeals have evaluated whether the EPA is appropriate Section 5 legislation. These courts concluded that states remain subject to suit under the EPA.

This Comment argues that lower courts have correctly upheld the EPA as appropriate Section 5 legislation. Part I outlines the substantive elements of the EPA. Part II introduces the constitutional doctrine of state sovereign immunity. Part III discusses the “congruence and proportionality” test used by the U.S. Supreme Court to determine what constitutes appropriate Section 5 legislation. Part IV examines to what extent gender discrimination violates the Fourteenth Amendment. Part V

17. See id. at 59.
20. See id.
22. Id. at 91–92.
25. See Varner v. Ill. State Univ., 226 F.3d 927 (7th Cir. 2000); Kovacevich v. Kent State Univ., 224 F.3d 806 (6th Cir. 2000); Hundertmark v. Fla. Dep’t of Transp., 205 F.3d 1272 (11th Cir. 2000).
argues that the EPA is appropriate Section 5 legislation and that EPA plaintiffs should remain able to sue state employers.

I. LEGISLATING AGAINST GENDER-BASED WAGE DISCRIMINATION: THE EQUAL PAY ACT

The EPA prohibits gender-based wage discrimination occurring when a man and a woman are found to do equal work for unequal pay.27 Prior to enacting the EPA, Congress heard substantial testimony regarding gender discrimination occurring in private-sector employment.28 Congress deemed legislation necessary to end gender-based wage discrimination, but industry leaders were concerned that Congress would disallow legitimate pay structures, such as those based on job value, seniority, or merit.29 Due to these concerns, Congress crafted a statute that would hold liable only those employers who engaged in gender-based wage discrimination.30

Congress drafted the EPA narrowly, making it applicable only to gender-based wage discrimination between men and women who do equal work. The EPA’s narrow scope becomes apparent after examining the prima facie requirements a plaintiff must satisfy. A plaintiff has the burden of proving that (1) two workers of the opposite sex, (2) in the same establishment, (3) are receiving unequal pay, (4) for equal work.31 The statute specifically states that equal work entails “equal skill, effort, and responsibility . . . performed under similar working conditions.”32

Addressing employers’ concerns, Congress gave employers four affirmative defenses that they could assert to show that any alleged wage discrimination was not based on gender.33 An employer can avoid liability under the EPA if it proves that unequal pay results from (1) a seniority system, (2) a merit system, (3) a system measuring earnings based on production, or (4) “any other factor other than sex.”34 The

30. See id. at 201.
31. 29 U.S.C. § 206(d)(1); see also Stanley v. Univ. of S. Cal., 178 F.3d 1069, 1074 (9th Cir. 1999) (describing plaintiff’s prima facie burden under EPA).
33. Id.; see also Brennan, 417 U.S. at 196–201.
34. 29 U.S.C. § 206(d)(1).

282
employer carries the burden of proving one of the affirmative defenses to avoid liability.\textsuperscript{35} Although a clear discriminatory intent may be revealed in many EPA cases due to the defendant’s inability to prove legitimate reasons for the unequal pay, the plaintiff is not required to prove discriminatory intent to prevail.\textsuperscript{36}

II. THE RESURGENCE OF STATE SOVEREIGN IMMUNITY

The Eleventh Amendment to the U.S. Constitution\textsuperscript{37} provides states with immunity from claims brought under federal law in both federal and state court.\textsuperscript{38} Congress has a limited power to abrogate this immunity through federal legislation.\textsuperscript{39} That power resides in Section 5 of the Fourteenth Amendment, which the U.S. Supreme Court has interpreted to be the only means available to Congress to abrogate sovereign immunity.\textsuperscript{40}

A. Judicial Construction of Sovereign Immunity

The literal text of the Eleventh Amendment provides states immunity only from suits brought under the citizen-state diversity jurisdiction of the federal courts.\textsuperscript{41} The U.S. Supreme Court, however, has not confined state sovereign immunity to the textual boundaries of the Amendment.\textsuperscript{42} Instead, the Court has held that established constitutional principles of sovereignty and federalism bar a citizen from suing the citizen’s own state in federal court even when the suit is based on federal question

\textsuperscript{35} Brennan, 417 U.S. at 196.

\textsuperscript{36} ZIMMER ET AL., supra note 2, at 1008.

\textsuperscript{37} U.S. CONST. amend. XI. (“The Judicial power of the United States shall not be construed to extend to any suit in law or equity, commenced or prosecuted against one of the United States by Citizens of another State, or by Citizens or Subjects of any Foreign State.”).


\textsuperscript{42} See Principality of Monaco v. Mississippi, 292 U.S. 313, 325–27 (1934).
jurisdiction. The Court has also extended the doctrine to bar suits brought under federal law against states in state courts.

B. Congressional Abrogation of Sovereign Immunity

Congress can only abrogate state sovereign immunity through legislation enacted pursuant to Section 5 of the Fourteenth Amendment. Section 5 authorizes Congress to “enforce, by appropriate legislation, the provisions of [the Fourteenth Amendment].” The Fourteenth Amendment provides in relevant part: “No State shall . . . deprive any person of life, liberty, or property, without due process of law; nor deny to any person . . . the equal protection of the laws.”

In Fitzpatrick v. Bitzer, the U.S. Supreme Court held for the first time that the Eleventh Amendment does not bar Congress from using Section 5 of the Fourteenth Amendment to authorize private lawsuits against states for money damages. Through ratification of the Fourteenth Amendment, the states empowered Congress to enforce that Amendment’s substantive guarantees. The Court saw no reason why that enforcement could not include providing for private suits against states, notwithstanding the constitutional doctrine of state sovereign immunity. The Court limited the Fitzpatrick holding to the validity of abrogating state sovereign immunity through the Section 5 power. The Court did not discuss whether congressional power to abrogate was confined exclusively to Section 5 legislation or whether other provisions of the Constitution granted Congress the power to subject states to suit. In Pennsylvania v. Union Gas Co., a plurality of the Court held that Congress could use its Commerce Clause power to subject states to suit.

43. See Hans v. Louisiana, 134 U.S. 1, 13–15 (1890).
46. Id. § 1.
48. Id. at 456.
49. Id. at 454–56.
50. Id. at 456.
51. See id.
53. U.S. CONST. art. I, § 8, cl. 3.
54. Union Gas, 491 U.S. at 23.
Seven years later, the U.S. Supreme Court expressly overruled Union Gas in Seminole Tribe v. Florida. The Court held that Union Gas was incorrectly decided and that no Article I power could be used to abrogate sovereign immunity. The effect of the Seminole Tribe holding is that only legislation enacted pursuant to the Fourteenth Amendment can be used to subject states to suit. The Fourteenth Amendment altered the pre-existing balance between state and federal power. States ratified the Fourteenth Amendment well after the Eleventh Amendment with the knowledge that they were ceding significant power to the federal government. The Court reaffirmed the validity of Fitzpatrick, finding that congressional power to abrogate state sovereign immunity is available, but limited to legislation enacted pursuant to the Section 5 power.

III. SECTION 5 ENFORCEMENT POWER: FROM FLORES TO KIMEL

The Seminole Tribe decision increased states’ ability to use sovereign immunity as a defense to suit. States cannot be sued under legislation that Congress has not enacted through the Section 5 power. Recent U.S. Supreme Court decisions have limited what can be considered appropriate Section 5 legislation, thereby further strengthening the sovereign immunity defense.

56. Id. at 72–73.
57. See id. at 59.
58. Id. at 65–66.
59. See id.
60. See id. at 59, 66, 72.
62. See id.
63. See id. at 504.
A. The Congruence and Proportionality Test: City of Boerne v. Flores

In *City of Boerne v. Flores*, the Court faced its first opportunity since the *Seminole Tribe* decision to discuss what constitutes valid Section 5 legislation. *Flores* involved a challenge to the Religious Freedom Restoration Act (RFRA), a statute that prohibited the government from burdening religious practices even if the burden resulted from neutral, generally applicable laws. The RFRA was enacted to supersede the holding in *Employment Division v. Smith*, where the Court held that religion-neutral laws that burden religious practices need not be supported by a compelling government interest. The purpose of the RFRA was restoration of the compelling interest test to all cases where the free exercise of religion is burdened.

The Court in *Flores* explained that Congress exceeded its Section 5 power when enacting the RFRA. Section 5 gives Congress the power to enact remedial legislation intended to remedy or prevent conduct that would independently violate the Fourteenth Amendment. Congress has wide latitude in determining what constitutes remedial legislation, but the U.S. Supreme Court will not uphold legislation that dilutes or expands the protection that the Constitution offers citizens. Because of the difficulty in determining where the line between appropriate remedial and inappropriate substantive Section 5 legislation lies, Congress may restrict some otherwise constitutional conduct if such restriction will help prevent the unconstitutional conduct primarily targeted by the legislation. If the legislation goes substantially beyond the substantive

64. 521 U.S. 507 (1997).
68. *Id.* at 885.
70. See *Flores*, 521 U.S. at 532–36.
71. See *id.* at 519.
72. *Id.* at 520, 536.
73. See *id.* at 519–20, 536.
74. *Id.* at 519.
75. *Id.* at 518.
Section 5 and the Equal Pay Act

protection of the Fourteenth Amendment, Congress must show, through the legislative record, that it was justified in its legislative response. 76

Rejecting the contention that the RFRA was appropriate remedial legislation, the Court formulated a test for lower courts to apply in determining whether legislation was appropriate under Section 5. 77 For legislation to be “appropriate” under Section 5, “[t]here must be a congruence and proportionality between the injury to be prevented or remedied and the means adopted to that end.” 78 In other words, the substantive rights Congress gives to litigants through Section 5 legislation must be substantially similar to the rights litigants would have if they brought suit under the Constitution. The means used by Congress to prevent the harm must be proportional to the harm Congress seeks to remedy. 79

Using the congruence and proportionality test, courts must examine whether Congress enacted the challenged legislation to remedy or prevent state action that would likely be unconstitutional. 80 The legislation must enforce actual substantive guarantees of the Fourteenth Amendment or be a proportional response to previously documented constitutional violations. 81 If the legislation is found to mirror the protection of the Constitution, the inquiry is at an end. However, if the Court finds that the legislation goes beyond the substantive protection of the Fourteenth Amendment, the Court will look into the legislative record for evidence suggesting that Congress had justification for enacting such broad legislation. 82

Applying the congruence and proportionality test to the RFRA’s compelling interest test, the Court found that the RFRA imposed the “most demanding test known to constitutional law” on states to justify facially neutral laws that burdened religious practices. 83 By raising the level of judicial scrutiny for neutral laws to the compelling interest standard, the RFRA dramatically increased the protection normally given to religious entities beyond what such entities would receive under the

76. See id. at 531–32.
77. See id. at 519–20.
78. Id. at 520.
80. Flores, 521 U.S. at 532.
81. Thro, supra note 61, at 504.
82. Kimel, 528 U.S. at 88.
83. Flores, 521 U.S. at 534.
Constitution. The RFRA made otherwise constitutional state legislation unlawful, and therefore the Act expanded the substance of the Constitution.\textsuperscript{84} In addition, the Court found no evidence in the legislative record that Congress had uncovered a widespread pattern of religious discrimination that would justify remedial legislation of any sort, let alone a broad statute like the RFRA.\textsuperscript{85}

The Court dismissed the argument that its previous decision in \textit{Katzenbach v. Morgan}\textsuperscript{86} approved of a “one way ratchet” theory of congressional authority.\textsuperscript{87} Under such a theory, Congress could permissibly expand but not dilute the substantive protection found in the Fourteenth Amendment.\textsuperscript{88} In \textit{Morgan}, the Court upheld a challenged provision of the Voting Rights Act (VRA)\textsuperscript{89} as appropriate Section 5 legislation.\textsuperscript{90} Interpreting \textit{Morgan}, the \textit{Flores} Court explained that the challenged VRA provision should be viewed either as a measure to remedy unconstitutional discrimination in governmental services or to remedy unconstitutional discrimination in establishing voter qualifications.\textsuperscript{91} Congress enacted the VRA to remedy state conduct that would be independently unconstitutional.\textsuperscript{92} Unlike the RFRA, the VRA did not give plaintiffs more substantive protection than found in the Constitution.\textsuperscript{93} The RFRA could not be considered appropriate Section 5 legislation because it was not targeted at remedying unconstitutional conduct; therefore, states were immune from suit under the RFRA.\textsuperscript{94}

\textsuperscript{84} See id. at 532–34.
\textsuperscript{85} Id. at 531.
\textsuperscript{86} 384 U.S. 641 (1966).
\textsuperscript{87} See Flores, 521 U.S. at 527–28.
\textsuperscript{88} GEOFFREY STONE ET AL., CONSTITUTIONAL LAW 261 (3d ed. 1996).
\textsuperscript{89} 42 U.S.C. § 1973b(e)(2) (1994) (providing that no person with sixth-grade education accredited by Puerto Rico could be prohibited from voting because of inability to read or write English).
\textsuperscript{90} Morgan, 384 U.S. at 646.
\textsuperscript{91} Flores, 521 U.S. at 528.
\textsuperscript{92} See id.
\textsuperscript{93} See id.
\textsuperscript{94} Id. at 532.
B. Flores and Seminole Tribe Together: Restricting the Ability To Sue States

The interplay between Flores and Seminole Tribe significantly reduces the ability of plaintiffs to sue states under a variety of federal statutes.95 If Congress did not enact a statute pursuant to Section 5 by specifically enforcing substantive constitutional guarantees, then the statute cannot abrogate sovereign immunity.96 Because the RFRA blatantly expanded the level of constitutional protection given to religious entities, it was unclear how the U.S. Supreme Court would apply the “congruence and proportionality” test when confronted with a closer case.

The Court’s opportunity to expand upon its explanation of the test came in 1999. In that year, the Court decided three cases regarding Section 5 legislation and sovereign immunity.97 In all three cases the Court held for the States without substantial discussion of the “congruence and proportionality” test.

The most extensive analysis of the congruence and proportionality test occurred in Florida Prepaid Postsecondary Education Expense Board v. College Savings Bank.98 In Florida Prepaid, a patentee brought an action against an agency of the State of Florida under provisions of the Patent and Plant Variety Protection Remedy Clarification Act (Patent Remedy Act),99 alleging patent infringement.100 The Court held that the Patent Remedy Act was not appropriate Section 5 legislation.101 Applying the Flores test, the Court determined that the Patent Remedy Act was not

96. Thro, supra note 61, at 504.
100. Florida Prepaid, 527 U.S. at 631.
101. Id. at 647.
designed to prevent or remedy unconstitutional state action.102 The Court rejected the contention that the act was designed to enforce the Due Process Clause of the Fourteenth Amendment.103 The Fourteenth Amendment offers no independent remedy to patent holders unless the infringement occurred without due process of law.104 Congress had “barely considered” the remedies available to patent holders in state courts, and the Court saw no evidence that Congress intended to prevent due process violations.105

Like the legislative findings in Flores, Congress had not identified a pattern of patent infringement by the states to justify federal intervention, nor was there evidence of any historical deprivation of constitutional rights in this area of the law.106 Due to the lack of evidence of constitutional violations by the states in the patent infringement area, the Patent Remedy Act could not be considered a proportional response designed to prevent unconstitutional state action.107

C. The Court Held that the ADEA Was Inapplicable to the States

In Kimel v. Florida Board of Regents,108 the U.S. Supreme Court applied the congruence and proportionality test to civil rights legislation for the first time, holding the Age Discrimination in Employment Act (ADEA)109 inapplicable to state employers.110 U.S. Supreme Court decisions prior to Kimel had held that states could constitutionally discriminate on the basis of age if the discrimination was rationally related to a state interest.111 The Court found that the ADEA impermissibly raised the level of judicial scrutiny applicable to age discrimination claims.112 Furthermore, the legislative record lacked

102. See id. at 639–41.
103. See id. at 641–43.
104. Id. at 643.
105. Id.
106. Id. at 645–46.
107. Id. at 646.
110. Kimel, 528 U.S. at 91.
evidence of state-enforced age discrimination, and therefore Congress was not justified in enacting such broad, substantive legislation.\textsuperscript{113}

The plaintiffs in \textit{Kimel}, employees of various state universities in Florida, brought suit alleging that the Florida Board of Regents had not fulfilled its promise to require universities to provide market adjustments to the salaries of eligible employees.\textsuperscript{114} The plaintiffs contended that this failure to allocate funds had a disparate impact on the base pay of older employees and therefore violated the ADEA.\textsuperscript{115} The ADEA makes it unlawful for employers, including state employers, “to fail or refuse to hire or discharge any individual or otherwise discriminate against any individual . . . because of such individual’s age.”\textsuperscript{116} The Eleventh Circuit held that the ADEA was not appropriate Section 5 legislation\textsuperscript{117} and the U.S. Supreme Court granted certiorari to resolve the split that had emerged in the circuit courts.\textsuperscript{118}

Applying the congruence and proportionality test, the Court examined whether classifications based on age generally violate the Equal Protection Clause.\textsuperscript{119} The Court had previously decided three cases regarding the constitutionality of age classifications.\textsuperscript{120} In each of those cases, the Court had held that the age classifications were constitutional.\textsuperscript{121} The legislatures had had a rational basis\textsuperscript{122} for each of the classifications at issue.\textsuperscript{123} In \textit{Kimel}, the Court specifically

\begin{footnotes}
\footnotetext{113. Id. at 89–91.}
\footnotetext{114. Id. at 69–70.}
\footnotetext{115. Id. at 70.}
\footnotetext{117. \textit{Kimel}, 528 U.S. at 71.}
\footnotetext{118. Id. The Eighth Circuit had previously held that the ADEA did not validly abrogate sovereign immunity. Humenansky v. Regents of Univ. of Minn., 152 F.3d 822, 827–28 (8th Cir. 1998). Six other circuit courts had found the ADEA to be appropriate Section 5 legislation. \textit{See Cooper v. N.Y. State Office of Mental Health}, 162 F.3d 770 (2d Cir. 1998); Migneault v. Peck, 158 F.3d 1131 (10th Cir. 1998); Coger v. Bd. of Regents, 154 F.3d 296 (6th Cir. 1998); Keeton v. Univ. of Nev. System, 150 F.3d 1055 (9th Cir. 1998); Scott v. Univ. of Miss., 148 F.3d 493 (5th Cir. 1998); Goshtasby v. Bd. of Trustees of the Univ. of Ill., 141 F.3d 761 (7th Cir. 1998).}
\footnotetext{119. \textit{See Kimel}, 528 U.S. at 82–84.}
\footnotetext{121. \textit{See supra} note 111 and accompanying text.}
\footnotetext{122. When a statute is evaluated under a “rational basis” standard, the court will accept almost any justification even if it was not the one that motivated the legislature when writing the statute. \textit{See McGowan v. Maryland}, 366 U.S. 420, 425–26 (1961).}
\footnotetext{123. \textit{See Kimel}, 528 U.S. at 84–86.}
\end{footnotes}
distinguished governmental age classifications from those based on race or gender under which the government must show a substantial or compelling interest to justify the classification.124 Unlike race or gender, states are allowed to discriminate on the basis of age if the classification is rationally related to a legitimate state interest.125

In *Kimel*, the Court found the ADEA to lack the necessary “congruence and proportionality” required for appropriate Section 5 legislation.126 The Court noted that a substantial number of state employment classifications that would not violate the Constitution would be unlawful under the ADEA.127 The Act also imposed a higher level of judicial scrutiny on age classifications than required by equal protection analysis.128 The Equal Protection Clause allows broad, general classifications based on age if a rational basis exists for the action.129 To the contrary, the ADEA required the employer to make individualized determinations regarding whether a specific older employee was suitable for employment.130 The ADEA therefore imposed a substantially heavier burden than rational basis review on state employers seeking to justify classifications based on age.131 The ADEA subjected age classifications to a higher level of judicial scrutiny than the Equal Protection Clause, and therefore it impermissibly expanded the substance of constitutional protection.

After finding that the ADEA expanded the substance of the Fourteenth Amendment, the Court looked for evidence in the legislative record that would justify the enactment of such broad legislation.132 The Court found that Congress had uncovered inconsequential evidence of age

---

124. *Id.* at 83–84.
125. See *id.* at 84.
126. *Id.* at 82–83.
127. *Id.* at 85–86.
128. See *id.* at 86–88.
129. See *id.* at 87–88.
130. *Id.* at 87.
131. *Id.* at 87–88.
132. See *id.* at 88–91.
Section 5 and the Equal Pay Act

discrimination by state employers.\textsuperscript{133} The limited findings made by Congress only related to discrimination that would not have amounted to equal protection violations, meaning that the age classifications were rationally related to legitimate state interests.\textsuperscript{134} Although Congress discovered age discrimination that was committed by private employers, this evidence was wholly irrelevant to whether the states had engaged in discrimination.\textsuperscript{135} Thus, the Court determined that the evidence before Congress did not justify the broad legislative response as applied to the states.\textsuperscript{136} The ADEA could not be justified as a proportional response to discrimination by the states against older Americans and was inappropriate Section 5 legislation.\textsuperscript{137}

D. Lower Court Challenges to the Equal Pay Act

The EPA has not been immune from challenges to its validity as Section 5 legislation. Prior to \textit{Kimel}, the only two circuit courts of appeal to consider whether the EPA was appropriate Section 5 legislation under the congruence and proportionality test upheld it as such.\textsuperscript{138} The U.S. Supreme Court vacated and remanded both decisions for reconsideration in light of \textit{Kimel}.\textsuperscript{139}

Pre-\textit{Kimel} challenges to the validity of the EPA as Section 5 legislation primarily focused on the fact that an EPA plaintiff need not prove discriminatory intent.\textsuperscript{140} For example, one district court held that

\begin{itemize}
  \item \textsuperscript{133} See \textit{id.} at 89–91.
  \item \textsuperscript{134} See \textit{id.} at 90–91.
  \item \textsuperscript{135} Id. at 91.
  \item \textsuperscript{136} See \textit{id.} at 91–92.
  \item \textsuperscript{137} See \textit{id.} at 91–92.
  \item \textsuperscript{138} See Anderson v. State Univ. of N.Y., 169 F.3d 117 (2d Cir. 1999), vacated and remanded by 528 U.S. 1111 (2000); Varner v. Ill. State Univ., 150 F.3d 706 (7th Cir. 1998), vacated and remanded by 528 U.S. 1110 (2000); see also Timmer v. Mich. Dep’t of Commerce, 104 F.3d 833, 842 (6th Cir. 1997) (upholding EPA as appropriate Section 5 legislation prior to \textit{Flores}).
  \item \textsuperscript{139} See State Univ. of N.Y. v. Anderson, 528 U.S. 1111 (2000); Ill. State Univ. v. Varner.
  \item \textsuperscript{140} See Varner, 150 F.3d at 714–15; see also Larry v. Bd. of Trustees, 975 F. Supp. 1447, 1449–50 (N.D. Ala. 1997) \textit{rev’d in part, vacated in part}, 211 F.3d 598 (11th Cir. 2000). \textit{Larry} was reversed and vacated by the Eleventh Circuit Court of Appeals two weeks after its decision in \textit{Hundertmark v. Florida Department of Transportation}, 205 F.3d 1272 (11th Cir. 2000). See \textit{Larry}, 211 F.3d 598.
\end{itemize}
the EPA was invalid Section 5 legislation on the sole basis that the EPA does not require the plaintiff to prove the discriminatory effects were the result of a state decision-maker’s discriminatory intent.\(^\text{141}\) That court found the EPA lacked the intent element required in equal protection claims, so it could not be viewed as a statute enforcing the Equal Protection Clause.\(^\text{142}\) Other decisions dismissed this line of attack stating that the EPA was enacted to prevent unconstitutional discrimination and that Congress can forbid conduct otherwise permissible to prevent or remedy unconstitutional conduct.\(^\text{143}\)

Three circuit courts of appeal have rendered decisions in suits challenging the EPA since the \textit{Kimel} decision.\(^\text{144}\) In each case, the court has held that the EPA is appropriate Section 5 legislation even in light of \textit{Kimel}, although each court’s reasoning and depth of analysis has varied. In \textit{Varner v. Illinois State University},\(^\text{145}\) the Seventh Circuit disagreed with the defendant’s contention that removing the burden of proving intent from an EPA plaintiff renders the EPA incongruent with the Equal Protection Clause.\(^\text{146}\) The court noted that compared to the ADEA, the EPA was narrow in scope,\(^\text{147}\) allowed employers a greater opportunity to escape liability,\(^\text{148}\) and targeted unconstitutional state action.\(^\text{149}\) The court found that in light of the prevalent problem of unconstitutional gender discrimination, the enacted remedial scheme was sufficiently congruent and proportional to the Fourteenth Amendment to be considered appropriate Section 5 legislation.\(^\text{150}\)

With only brief analysis, the Eleventh Circuit in \textit{Hundertmark v. Florida Department of Transportation}\(^\text{151}\) held that the EPA is congruent...
with the Equal Protection Clause solely because the aim of the EPA is to eradicate intentional gender discrimination. The court’s per curiam opinion was dedicated to describing how gender discrimination is a “problem of national import” and how, unlike age discrimination, gender discrimination is generally unconstitutional. The opinion did not discuss the different burdens of proof in EPA and Equal Protection Clause litigation, nor did it question whether Congress made sufficient findings of gender discrimination in the public sector before extending the EPA to the states.

The Sixth Circuit in Kovacevich v. Kent State University, also with limited discussion, upheld the EPA as appropriate Section 5 legislation. Relying heavily on dicta from Kimel, the court concluded that unlike the ADEA, the EPA does not prohibit substantially more state employment decisions than would likely be held unconstitutional. The court emphasized that in the Sixth Circuit EPA liability is equated with intentional gender discrimination and also noted that the affirmative defenses available to employers reduce the possibility that constitutional conduct is held unlawful under the EPA.

IV. GENDER DISCRIMINATION AND THE FOURTEENTH AMENDMENT

When analyzing whether the EPA is appropriate Section 5 legislation, a court must determine whether the EPA renders otherwise constitutional conduct unlawful or if it simply remedies unconstitutional state action. Therefore, courts must examine whether the conduct the EPA conceivably targets would violate the Fourteenth Amendment, specifically the Equal Protection Clause. Because the EPA is targeted at ending differential rates of pay based solely on an employee’s gender,
the appropriate initial inquiry is to what extent gender-based wage discrimination violates the Fourteenth Amendment.\textsuperscript{161}

The Fourteenth Amendment does not preclude state governments from classifying groups of people.\textsuperscript{162} When determining the constitutionality of a legislative classification a court will normally ask whether the classification is directed at the achievement of a legitimate governmental purpose and whether it rationally furthers that purpose.\textsuperscript{163} Courts examining the constitutionality of legislation under this "rational basis" standard are extremely deferential to the judgment of legislatures.\textsuperscript{164} If any conceivable set of facts could explain the basis for the legislation under the rational basis standard, a court will uphold the legislation as proper under the Fourteenth Amendment.\textsuperscript{165}

If legislatures classify individuals primarily on the basis of immutable characteristics such as race\textsuperscript{166} and gender,\textsuperscript{167} judicial scrutiny is heightened and the state bears a much heavier burden to justify the classification.\textsuperscript{168} If a state facially discriminates on the basis of gender, it has the burden of providing an "'exceedingly persuasive'" justification for the discriminatory classification.\textsuperscript{169} The state can meet this high standard only if the discriminatory classification "'serves 'important governmen-tal objectives and . . . [is] substantially related to the achievement of those objectives.'"\textsuperscript{170} Any classifications used to "create or perpetuate the legal, social, and economic inferiority of women" are invalid.\textsuperscript{171} If a

\begin{itemize}
  \item \textsuperscript{161} See Kimel v. Fla. Bd. of Regents, 528 U.S. 62, 83 (2000) (making initial inquiry into whether and under what circumstances age discrimination violates Fourteenth Amendment).
  \item \textsuperscript{162} See Personnel Adm’r v. Feeney, 442 U.S. 256, 271–72 (1979).
  \item \textsuperscript{165} See id.
  \item \textsuperscript{166} See Korematsu v. United States, 323 U.S. 214, 216 (1944). Racial classifications are subject to "strict scrutiny," meaning they can only be upheld if they are justified by a compelling state interest and are narrowly tailored to further that interest. See Bush v. Vera, 517 U.S. 952, 976 (1996).
  \item \textsuperscript{168} See Personnel Adm’r v. Feeney, 442 U.S. 256, 272–73 (1979).
  \item \textsuperscript{169} Virginia, 515 U.S. at 533 (quoting Miss. Univ. for Women v. Hogan, 458 U.S. 718, 724 (1982)).
  \item \textsuperscript{171} Id. at 534.
\end{itemize}
Section 5 and the Equal Pay Act

state uses gender to classify individuals for purposes of distributing benefits or burdens, such action will rarely be upheld.\textsuperscript{172}

Gender-neutral classifications that have discriminatory effects on women also violate the Equal Protection Clause, but only if enacted with a discriminatory purpose.\textsuperscript{173} Unlike facially gender-specific classifications, courts will not presume discriminatory intent behind such classifications, and the burden of proof remains on the plaintiff to prove the state action was enacted with discriminatory intent.\textsuperscript{174} Proof of the discriminatory effect alone will not be enough to make a prima facie showing of intent.\textsuperscript{175} If a plaintiff makes a prima facie showing that gender was a motivating factor in the decision-making process, the burden shifts to the defendant to establish that the same decision at issue would have been made even if gender had not been considered.\textsuperscript{176}

The “heightened scrutiny” test applied to gender-specific classifications also applies to neutral classifications where the plaintiff proves discriminatory intent.\textsuperscript{177} However, if gender discrimination is proven to be a motivating factor behind state action that has a discriminatory effect on women, the state cannot justify it under the heightened scrutiny test described above. For example, a facially neutral pay scale administered by state officials in a discriminatory fashion would violate the Equal Protection Clause if the plaintiff proved that gender and not a legitimate non-discriminatory factor was the basis behind the discriminatory treatment.\textsuperscript{178}

V. COURTS SHOULD UPHOLD THE EPA AS APPROPRIATE SECTION 5 LEGISLATION

The EPA is congruent and proportional with the Equal Protection Clause of the Fourteenth Amendment and therefore qualifies as appropriate Section 5 legislation. Unlike recent congressional acts struck down as inappropriate Section 5 legislation, the EPA does not grant

\textsuperscript{172} See id. at 533.
\textsuperscript{173} See Feeney, 442 U.S. at 273–74.
\textsuperscript{174} See id. at 276.
\textsuperscript{177} See Feeney, 442 U.S. at 272.
\textsuperscript{178} See id. at 276 (stating that dispositive question for determining whether Equal Protection Clause was violated was whether facially neutral statute was enacted with discriminatory purpose).
plaintiffs more substantive rights than the Constitution. The EPA also
does not raise the level of scrutiny given to gender-based classifications,
and the EPA is congruent with the Equal Protection Clause even though
a plaintiff need not prove discriminatory intent to prevail. Overall, due to
the narrow scope of the EPA, it is a proportional measure to combat
gender-based wage discrimination. In addition, Congress made the EPA
applicable to the states in response to substantial evidence that gender-
based wage discrimination was a serious problem in public
employment.\footnote{See infra Part V.B.} Thus, the EPA was enacted for a remedial purpose and is
valid Section 5 legislation.

A. The EPA Satisfies the Congruence and Proportionality Test

Under the congruence and proportionality test, a court must determine
whether the EPA targets the prevention of unconstitutional conduct or if it
impermissibly attempts to redefine the substance of constitutional
protection.\footnote{See Kimel v. Fla. Bd. of Regents, 528 U.S. 62, 82 (2000).} The EPA is congruent and proportional to the Equal
Protection Clause and is appropriate Section 5 legislation for three
reasons. First, the EPA does not subject gender-based classifications to a
higher level of judicial scrutiny than the Equal Protection Clause.
Second, although an EPA plaintiff need not prove discriminatory intent
to prevail, the affirmative defenses available under the EPA make it
unlikely in practice that employers who lack the discriminatory intent
necessary for an equal protection violation will be held liable under the
EPA. Third, the EPA is very narrow in scope and is limited to preventing
unequal compensation of men and women who do equal work.

1. The EPA Does Not Mandate Closer Scrutiny of Gender-Based
Classifications than the Equal Protection Clause

The EPA does not impose a higher level of judicial scrutiny on
gender-based classifications than would be given to such claims if
challenged under the Equal Protection Clause.\footnote{See Varner v. Ill. State Univ., 226 F.3d 927, 935 (7th Cir. 2000) (asserting that in some ways
Equal Protection Clause standard of liability is even more demanding than EPA); see also Kovacevich v. Kent State Univ., 224 F.3d 806, 820 (6th Cir. 2000) (“The standard for liability under the
EPA closely approximates the equal protection analysis for state-sponsored gender
discrimination.”). In Hundertmark v. Florida Department of Transportation, 205 F.3d 1272 (11th

\footnote{See infra Part V.B.}
Section 5 and the Equal Pay Act

Protection Clause, courts analyze gender classifications with a heightened level of scrutiny. \(^{182}\) Because of the heightened scrutiny given to gender-based classifications, it is unlikely that courts would find any gender-based wage structure unlawful under the EPA that would not also be independently unconstitutional.

A state wage structure expressly stating that females will be paid less than males for doing the same work would clearly violate the EPA. \(^{183}\) Proof of a facially discriminatory policy that required paying men and women unequally for equal work would easily satisfy the EPA’s prima facie requirements. \(^{184}\) The government employer would then have the burden of proving that the differential pay was based on a factor other than gender. \(^{185}\) The employer could not successfully invoke this defense because by definition a gender-specific classification is based on gender. \(^{186}\)

Assuming an analogous factual situation litigated under the Fourteenth Amendment, the only way the state would not violate the Equal Protection Clause is if it had an “exceedingly persuasive justification” for the policy. \(^{187}\) A state is unlikely to proffer any justification for a pay classification requiring unequal pay for men and women who did equal work persuasive enough for the U.S. Supreme Court to uphold the pay classification as constitutional. The Equal Protection Clause offers women extensive protection from gender-based classifications enacted by states and states carry a heavy burden to justify them.


\(^{183}\) See L.A. Dep’t of Water & Power v. Manhart, 435 U.S. 702, 704 (1978) (describing employment policy requiring women to contribute more to pension fund than men on sole basis that they are women); see also Equal Employment Opportunities Enforcement Act of 1971: Hearings Before the Subcomm. on Labor, 92d Cong. 433 (1971) (describing public school salary schedule in Salina, Kansas, that expressly paid male teachers more than female teachers).

\(^{184}\) See supra note 31 and accompanying text.


\(^{186}\) Judicial interpretations of the EPA suggest that gender “can provide no part of the basis for the wage differential.” ZIMMER ET AL., supra note 2, at 1007.

Alternatively, a state could argue that even though gender classifications are subject to heightened scrutiny, the EPA still holds employers to a higher standard than the Fourteenth Amendment. Under the EPA, wage structures that pay women unequally based on gender are illegal regardless of whether an important governmental purpose is served. Under the Equal Protection Clause, classifications based on gender are presumptively unconstitutional, but the presumption is rebuttable. For example, to justify the classification under the “exceedingly persuasive justification” test, the state has the opportunity to show that the classification serves “important governmental objectives” and is “substantially related to the achievement of those objectives.”

Such an argument underestimates the burden on an employer to justify a gender-based classification under the Equal Protection Clause. Courts reserve the most stringent level of judicial scrutiny for state classifications that discriminate on the basis of race. However, as a practical matter, the judicial scrutiny given to gender classifications is much more similar to the “strict scrutiny” given to racial classifications than to the extremely deferential “rational basis” standard used to evaluate all other legislative decisions. In fact, in his dissent in United States v. Virginia, Justice Scalia described the majority opinion as an

---

188. But see Varner v. Ill. State Univ., 226 F.3d 927, 935 (7th Cir. 2000) (“The Constitution demands an exceedingly persuasive justification for gender discrimination, while the [EPA] only requires an employer to offer some legitimate reason for a wage disparity other than sex.”).


190. See supra note 170 and accompanying text.


192. Under “strict scrutiny” review, the government must have a compelling state interest to justify the racial classification and there must be no less restrictive means to accomplish the objective. See Bush v. Vera, 517 U.S. 952, 976 (1996). Facially race-based classifications will almost always fail “strict scrutiny” and therefore be held unconstitutional. See STONE ET AL., supra note 88, at 601. But see Korematsu v. United States, 323 U.S. 214, 217–19 (1944) (holding national-security concerns enough to justify race-based classification).

193. See supra note 122.

194. Gender classifications are highly scrutinized because they are likely to reflect “archaic and overbroad” generalizations about women. Schlesinger v. Ballard, 419 U.S. 498, 507–08 (1975). However, gender classifications are not scrutinized as highly as race because, unlike with race, there are cases where a legitimate justification for the disparate treatment of men and women exists. See Michael M. v. Sonoma County Superior Court, 450 U.S. 464, 469 (1981).

application of “strict scrutiny” to gender classifications.\textsuperscript{196} A number of commentators have also noted that, in practice, there is little difference between the level of scrutiny given to gender and race classifications after Virginia.\textsuperscript{197} This similarity is significant because classifications based on race are considered almost impossible to justify.\textsuperscript{198}

States engaging in unequal gender-based pay are unlikely to have any explanation that would satisfy the exceedingly persuasive justification test. A state would have a difficult, if not impossible task explaining why underpaying a woman who did equal work as a man served an important governmental objective. A court would certainly classify any offered justification as rooted in the stereotypical and overbroad generalizations that are impermissible under the Equal Protection analysis of gender-based classifications.\textsuperscript{199}

Finally, the EPA does not suffer from the primary problem the Court found with the ADEA in Kimel v. Florida Board of Regents,\textsuperscript{200} which was that the ADEA effectively raised the level of judicial scrutiny given to age classifications.\textsuperscript{201} The ADEA imposed a level of scrutiny far more rigorous than the “rational basis” review normally given to presumptively constitutional age classifications.\textsuperscript{202} States may constitutionally restrict employment opportunity for individuals above a certain age so long as a rational basis exists for the action.\textsuperscript{203} The ADEA required courts to apply a higher level of judicial scrutiny than the Constitution would, making such otherwise constitutional age classifications presumptively illegal.\textsuperscript{204}

In summary, the EPA does not raise the level of constitutional scrutiny given to gender classifications as the ADEA did with age classifications. Unlike age classifications, the level of scrutiny given to gender

\textsuperscript{196} See Virginia, 518 U.S. at 574–75 (Scalia, J., dissenting).
\textsuperscript{198} No facially race-specific statute that disadvantaged a racial minority has been upheld since Korematsu v. United States, 323 U.S. 214 (1944). See STONE ET AL., supra note 88, at 601.
\textsuperscript{199} See Virginia, 518 U.S. at 533–34.
\textsuperscript{200} 528 U.S. 62 (2000).
\textsuperscript{201} See Varner v. Ill. State Univ., 226 F.3d 927, 934–35 (7th Cir. 2000).
\textsuperscript{202} See supra note 131 and accompanying text.
\textsuperscript{204} See Kimel, 528 U.S. at 86.
classifications under the Equal Protection Clause is so high that such classifications would rarely be constitutional. The EPA does not impose a higher level of judicial scrutiny on gender-based classifications than would be given to such claims if challenged under the Equal Protection Clause.

2. The EPA Is Congruent with the Equal Protection Clause Even Though an EPA Plaintiff Need Not Prove Discriminatory Intent To Prevail

A plaintiff does not have the burden of proving discriminatory intent to prevail under the EPA. That fact notwithstanding, the EPA is congruent with the Equal Protection Clause because the affirmative defenses available to employers make it unlikely that employers will be held liable under the EPA unless they had a discriminatory motive behind their pay structure. While the EPA requires no prima facie showing of intent, placing the burden of proof on the employer in EPA litigation does not make it inappropriate Section 5 legislation. The shifting of the burden of proof does not change the fact that employers who have legitimate non-discriminatory justifications for their wage structures will escape liability under the EPA.

When plaintiffs challenge a facially neutral pay structure that has the effect of discriminating against women under the Equal Protection Clause, the important issue is not the level of judicial scrutiny given to the pay structure, but whether the plaintiff can prove that it was enacted with discriminatory intent. In contrast, a plaintiff does not have to

205. See id

206. The burden-shifting issue was not addressed in Hundertmark v. Florida Department of Transportation, 205 F.3d 1272 (11th Cir. 2000). In Kovacevich v. Kent State University, 224 F.3d 806 (6th Cir. 2000), the court did not specifically discuss whether the burden-shifting effect of the EPA affects its congruence to the Equal Protection Clause. In Varner v. Illinois State University, 226 F.3d 927 (7th Cir. 2000), the state’s primary argument was based on the burden-shifting effect of the EPA. See id. at 934. The court agreed that the requirements of the EPA did not mirror the Equal Protection Clause and that some constitutional conduct was prohibited under the EPA. Id. The court noted, however, that the EPA’s affirmative defenses provide sufficient opportunity for an innocent employer to escape liability. Id.

207. See, e.g., EEOC v. Del. Dep’t of Health & Social Services, 865 F.2d 1408, 1414–16 (3d Cir. 1989) (showing that female nurses were paid less than male physician’s assistants even though their jobs entailed equal work satisfies prima facie case under EPA).

208. See supra note 173 and accompanying text.
Section 5 and the Equal Pay Act

prove discriminatory intent to prevail under the EPA. In an EPA claim, the defendant has the burden of persuading the trier of fact that the pay structure was based on legitimate, non-discriminatory factors.

Notwithstanding the differences in the plaintiff’s prima facie case and in the defendant’s burden of proof in EPA and equal protection claims, the EPA is appropriate Section 5 legislation. Although the plaintiff in an EPA claim does not have to prove discriminatory intent to prevail, the structure of EPA litigation ensures that employers will only be held liable when the court finds an impermissible motive of gender discrimination behind the pay structure. The EPA requires the employer to show that the wage differential was in fact the result of non-discriminatory intentions. By allowing the employer to escape liability with a showing that any “factor other than sex” motivated the wage structure, all possible motives for the wage differential other than gender will be eliminated and a court will be left with essentially a finding of gender-based discrimination. By definition, when a court finds an employer’s conduct to be based on gender and not motivated by non-discriminatory reasons, the court has found that the employer intended to discriminate.

The EPA is conducive to uncovering the same discriminatory intent necessary for Equal Protection violations. While the EPA does not expressly require evidence of intent, what separates many successful EPA defendants from unsuccessful ones is whether or not the court has reason to suspect intent to discriminate. Therefore, the lack of a prima facie intent requirement should not be enough to make the EPA inappropriate Section 5 legislation.

209. See supra note 36 and accompanying text.
210. See Varner, 226 F.3d at 934.
212. See Kouba v. Allstate Ins. Co., 691 F.2d 873, 876 (9th Cir. 1982); supra note 34 and accompanying text.
213. See Varner, 226 F.3d at 934.
214. In Korte v. Diemer, 909 F.2d 954 (6th Cir. 1990), the court held that a finding of liability under the EPA could be used in a later Title VII suit as evidence of intentional discrimination. Id. at 959. It noted that conduct found to violate the EPA could not be said in a later suit to lack intent to discriminate based on gender. Id.; see also Kovacevich v. Kent State Univ., 224 F.3d 806, 820 (6th Cir. 2000) (equating EPA liability with Title VII liability for intentional gender discrimination).
215. See Zimmer et al., supra note 2, at 1008; see also Martha Chamallas, Women and Part-Time Work: The Case for Pay Equity and Equal Access, 64 N.C. L. REV. 709, 747–49 (1986) (describing cases that suggest intent to discriminate is touchstone of liability under EPA).
The EPA’s structure prevents employers from being held liable for constitutionally permissible behavior despite the placement of the risk of non-persuasion on the employer. To ensure that the EPA would not interfere with legitimate pay structures, Congress gave employers four affirmative defenses to avoid liability if the wage discrimination was not based on gender. The fourth affirmative defense states that if the wage differential was based on “any other factor other than sex” the employer will escape liability.

While courts have not interpreted the EPA’s fourth affirmative defense to literally mean that any factor, no matter how obviously pretextual, will justify the wage differential, courts are deferential to an employer’s stated purpose. Courts “are not permitted to ‘substitute their judgment for the judgment of the employer . . . who [has] established and applied a bona fide job rating system, so long as it does not discriminate on the basis of sex.” The broad interpretation given to this affirmative defense makes it unlikely that a substantial number of employers would be held liable under the EPA for conduct that would not violate the Fourteenth Amendment.

The fact that the state bears the risk of non-persuasion in EPA litigation does not compel the result that the EPA is incongruent with the Fourteenth Amendment. The U.S. Supreme Court has granted Congress wide latitude in determining where the line between remedial and substantive legislation lies. If Congress enacts legislation to prevent or remedy unconstitutional conduct, it may simultaneously prohibit a “somewhat broader swath of conduct” including conduct not otherwise unconstitutional. Chief Justice Rehnquist, who has a very narrow view

---

216. See supra note 33 and accompanying text.
218. See Kouba v. Allstate Ins. Co., 691 F.2d 873, 876–77 (9th Cir. 1982).
221. The party who bears the burden of proof on an issue bears the risk that the trier of fact will not be persuaded. See 2 JOHN W. STRONG ET AL., MCCORMICK ON EVIDENCE 410 (5th ed. 1999).
of what constitutes appropriate Section 5 legislation,²²⁴ has suggested that placing the burden of proof on the government is permissible in the Section 5 context when legislation is otherwise focused on preventing unconstitutional conduct.²²⁵

The significance of placing the risk of non-persuasion on the state is limited to cases where the trier of fact remains uncertain who should prevail after all the evidence is presented.²²⁶ The law needs one party to bear the risk of non-persuasion only because a decision must be made; litigation cannot end in a tie.²²⁷ Who bears the risk is often a somewhat arbitrary judicial or legislative determination guided by principles of fairness and public policy.²²⁸ Congress presumably felt that it would be more fair for the employer shown to have a policy of paying men and women unequally for equal work to bear the risk of non-persuasion.

Considering that the purpose of the EPA is to eradicate gender-based wage discrimination,²²⁹ as well as the difficulties inherent in proving subjective discriminatory intent,²³⁰ and the broad defenses available to employers,²³¹ Congress justifiably placed the risk of non-persuasion on the employer instead of the employee.

Congress narrowly defined the substantive elements of the EPA so that only employers who pay discriminatory wages on the basis of gender are held liable.²³² Because the structure of EPA litigation often brings discriminatory motives to the forefront, the lack of a prima facie intent requirement should not be enough to make the EPA incongruent with the Fourteenth Amendment. Additionally, the congressional decision to shift the risk of non-persuasion to the employer is within the permissible bounds of the Section 5 power.


²²⁵. City of Rome v. United States, 446 U.S. 156, 214 (1980) (Rehnquist, J., dissenting) (“Congress could properly conclude that . . . it was necessary to place the burden of proving lack of discriminatory purpose on the [government].”).

²²⁶. See STRONG ET AL., supra note 221, at 410.

²²⁷. See id.


²³¹. See supra note 34 and accompanying text.

3. The EPA Is a Proportional Measure To Combat Gender-Based Wage Discrimination

Congress narrowly limited the scope of the EPA to prevent only unconstitutional gender-based wage discrimination. Legislation that is broadly over-inclusive is less likely to be a proportional response to the specific unconstitutional conduct at issue. The EPA is very narrow in scope and is a proportional measure to combat gender-based wage discrimination.

The EPA is focused on only one area of state responsibility: the relationship between state employer and employee. The EPA only regulates the compensation aspect of public employment. It does not deal with hiring, promotions, firing, or other wage discrimination issues relating to unequal or comparable work. It addresses only the rate of compensation between two employees who do equal work. This extremely narrow scope supports a finding that the EPA is a proportional response to gender-based wage discrimination.

Unlike the EPA, previous statutes challenged as inappropriate Section 5 legislation have broadly intruded into traditional areas of state responsibility. The Religious Freedom Restoration Act (RFRA) was a broad intrusion into the state legislative process. The RFRA affected potentially every piece of legislation that a state enacted. If state legislation would possibly burden a religious entity, the statute would have to satisfy the stringent “compelling interest” test required by the RFRA. Similarly, the ADEA intruded on all areas of state

233. Thro, supra note 61, at 505.
236. Zimmer et al., supra note 2, at 988.
237. Post-Kimel circuit court opinions have not emphasized the narrow scope of the EPA. It is clear, however, that the U.S. Supreme Court often focuses its attention on the scope of a challenged provision. See supra note 234 and accompanying text.
238. See, e.g., Kimel, 528 U.S. at 85–86; Flores, 521 U.S. at 532.
240. Flores, 521 U.S. at 534.
241. Id. at 533–34.
employment by imposing requirements on every aspect of the relationship between state employer and employee.

The EPA does not reach more broadly than is necessary to combat gender-based wage discrimination. The EPA only regulates the wages of individuals who do equal work but are paid unequally. It does not regulate other aspects of the employment relationship, or other state functions. Due to the EPA’s narrow reach and scope, courts should consider the statute a proportional measure to combat unconstitutional discrimination.

B. Congress was Justified in Applying the EPA to States Due to Substantial Evidence of Gender-Based Wage Discrimination by State Employers

Because it gathered substantial evidence that state employers engaged in gender-based wage discrimination, Congress was justified in applying the EPA to the states. Testimony heard throughout the early 1970s gave Congress substantial evidence that gender-based wage discrimination was a serious problem in public employment. This evidence strengthens the argument that Congress was justified in applying the protection of the EPA to state employees.

In its recent re-evaluation of sovereign immunity in Flores, Florida Prepaid, and Kimel, the Court has looked into the legislative record for evidence that the legislative response at issue was necessary to prevent constitutional violations by the states. “The appropriateness of remedial measures must be considered in light of the evil presented. Strong measures appropriate to address one harm may be an unwarranted

243. See Kimel, 528 U.S. at 85–87.
245. Of the post-Kimel opinions, only Varner v. Illinois State University, 226 F.3d 927 (7th Cir. 2000), adequately addressed the legislative findings issue. Id. at 935 (citing Congress’s clear understanding of state-enforced gender discrimination). In Hundertmark v. Florida Department of Transportation, 205 F.3d 1272 (11th Cir. 2000), the court stated that the lack of findings with respect to wage discrimination in the public sector was not dispositive because “gender discrimination is a problem of national import.” Id. at 1276. This directly contradicts language in Kimel stating that evidence of discrimination in the private sector is irrelevant when determining whether legislation is appropriate under Section 5. See Kimel, 528 U.S. at 90. The court in Kovacevich v. Kent State University, 224 F.3d 806 (6th Cir. 2000) did not address the legislative findings of Congress.
246. Infra note 262.
response to another, lesser one.”248 Courts must look to the legislative record for evidence showing whether Congress was justified in enacting the purportedly remedial statute.249

Prior to *Kimel*, lower courts upheld the EPA as appropriate Section 5 legislation based on the extensive evidence placed before Congress that *private* employers engaged in a long history of gender-based wage discrimination.250 This type of evidence, however, is irrelevant to whether Congress was justified in applying the EPA to the states.251 In *Kimel*, the U.S. Supreme Court expressly noted that a congressional finding of “substantial age discrimination in the private sector... is beside the point.”252 The inquiry into the legislative record should be limited only to evidence placed before Congress that the states themselves were engaged in unconstitutional discrimination.253

This inquiry need not be limited only to the hearings relating specifically to the EPA. When considering whether the EPA is appropriate Section 5 legislation, the Court may look to all the evidence placed before Congress to see if it could have rationally concluded that there was a problem.254 Courts cannot ignore the substantial evidence that Congress had gathered in hearings relating to the other legislation enacted in the early 1970s that dealt with gender-discrimination in public employment. Because members of Congress may use information learned from one set of hearings or debates as a basis for enacting other legislation,255 it is appropriate to consider the full range of evidence Congress heard regarding the extensive nature of sex discrimination by the states.

Before enacting the Fair Labor Standards Amendments, Congress heard extensive testimony regarding the need for the Fair Labor

248. *Flores*, 521 U.S. at 530 (citations omitted).
249. Id.
250. See, e.g., Varner v. Ill. State Univ., 150 F.3d 706, 716–17 (7th Cir. 1998).
251. *Kimel*, 528 U.S. at 90. Lack of legislative findings is not determinative of the Section 5 inquiry. Id. at 91.
252. Id. at 90.
253. See id.
255. Id.
Section 5 and the Equal Pay Act

Standards Act,256 of which the EPA is an amendment, to be applied to the states.257 The majority of the evidence Congress heard related to the need for extending the reach of the minimum-wage and child-labor provisions.258 However, Congress also called numerous witnesses to testify to the pervasive nature of gender-based wage discrimination in public schools and state universities, and therefore the need for making the EPA applicable to public employers.259 This evidence showed that, especially in the university and public school settings, women were paid less than men even though they performed equal work and that this wage disparity was attributable to intentional gender discrimination.260 The evidence showed that the problem was present in universities throughout the nation.261

In addition to the hearings relating specifically to the Fair Labor Standards Amendments, the early 1970s were a time when Congress heard much testimony regarding discrimination against women by the states.262 Before extending the EPA to the states, Congress enacted Title

257. Id.
259. 1971 FLSA, supra note 259, at 363 (testimony of Helen Bain) (testifying that female college professors “almost without exception, receive . . . substantially less for the same work than do their male counterparts” and that “[i]t is highly doubtful that these statistics are merely accidental”).
260. 1971 FLSA, supra note 259, at 322 (testimony of Dr. Bernice Sandler) (documenting studies showing unequal pay for equal work at University of Minnesota, University of Arizona, and Kansas State Teachers College); see also 1970 FLSA, supra note 259, at 547–50 (testimony of Dr. Bernice Sandler) (documenting complaints of unequal pay for equal work by female employees at University of Pittsburgh, University of California at Berkeley, and University of Michigan).
IX of the Education Amendments of 1972,\textsuperscript{263} and extended Title VII to state employers.\textsuperscript{264} Evidence gathered in the hearings relating to these pieces of legislation also revealed pervasive gender discrimination by public employers.\textsuperscript{265} The legislative hearings relating to the amendment of Title VII included extensive testimony regarding the pernicious nature of state-enforced employment discrimination.\textsuperscript{266} The evidence gathered at these hearings relating to anti-discrimination measures enacted just prior to the EPA indicates that Congress passed the EPA based on substantial evidence of gender-based wage discrimination in public employment.

As opposed to \textit{Flores}, where Congress had not documented modern instances of religious discrimination by states that would justify the RFRA,\textsuperscript{267} or in \textit{Kimel}, where Congress had no evidence the states engaged in unconstitutional age discrimination,\textsuperscript{268} Congress had evidence before it that the states did engage in gender-based wage discrimination.\textsuperscript{269} Not only did the hearings specifically relating to the Fair Labor Standards Amendments reveal this, but so did the extensive hearings relating to the other acts passed by Congress in the 1970s that were aimed at preventing women from being discriminated against by state employers.\textsuperscript{270} Considering all the evidence of state enforced gender-based wage discrimination placed before Congress in the early 1970s, Congress was justified in applying the EPA to the states.

VI. CONCLUSION

Unlike legislation recently held to be inappropriate Section 5 legislation, the EPA is congruent with the Fourteenth Amendment and is a proportional response to gender-based wage discrimination engaged in

\begin{footnotesize}
\begin{enumerate}
\item \textsuperscript{265} \textsuperscript{Supra} note 262.
\item \textsuperscript{267} City of Boerne v. Flores, 521 U.S. 507, 531 (1997).
\item \textsuperscript{268} \textit{Kimel v. Fla. Bd. of Regents}, 528 U.S. 62, 91 (2000).
\item \textsuperscript{269} \textsuperscript{Supra} note 262.
\item \textsuperscript{270} \textsuperscript{Supra} note 262.
\end{enumerate}
\end{footnotesize}
by the states. The Act does not raise the level of constitutional scrutiny given to gender-specific classifications, nor does it broadly interfere with state responsibility. The EPA prevents states from doing what they are prohibited from doing under the Fourteenth Amendment: paying a woman less than a man simply because of her gender. The EPA does not require a plaintiff to prove discriminatory intent to make out a prima facie case, but this does not automatically render the EPA inappropriate Section 5 legislation. The statute is narrowly designed to ensure that employers’ legitimate pay structures are upheld and employers who sexually discriminate are held liable. The EPA should be upheld as appropriate Section 5 legislation. State employees should remain able to vindicate their right to equal pay for equal work when state employers fail to meet the standards found not only in the Equal Pay Act, but also in the Fourteenth Amendment itself.
ATTORNEY-CLIENT CONFIDENTIALITY AND THE ASSESSMENT OF CLAIMANTS WHO ALLEGE POSTTRAUMATIC STRESS DISORDER

Robert H. Aronson,* Lonnie Rosenwald,** & Gerald M. Rosen***

Abstract: Posttraumatic Stress Disorder (PTSD) was first recognized by the American Psychiatric Association in 1980. A PTSD diagnosis requires an individual or individual’s loved ones to have experienced a traumatic event that was a threat to life or physical integrity and caused the individual to react to the incident with a specific number of avoidance, reexperiencing, and hyper-arousal symptoms. Obtaining a PTSD diagnosis can be of great value to a personal-injury plaintiff who claims damages due to a traumatic event. Further, if the traumatic event is unquestioned and the individual reports the classic symptoms, a PTSD diagnosis is relatively easy to apply and difficult to disprove. These plaintiffs will most often be examined and evaluated by mental-health professionals retained by the defendants. The question of whether the claimant was told or provided materials about common PTSD symptoms is crucial to the defense evaluator’s accurate PTSD assessment. One source of such information would be plaintiff’s counsel, but questions concerning information provided by counsel implicate the attorney-client privilege. This Article suggests that the policy bases underlying the attorney-client privilege and protecting a defendant’s right to test the validity of a plaintiff’s claims are best served by the creation of a narrowly drawn waiver or exception to the attorney-client privilege. Consistent with the patient-litigant exception to the physician-patient privilege, the proposed exception would be limited to those matters directly related to the nature, diagnosis, and symptoms of PTSD placed in issue by the plaintiff. The exception would also be limited to statements and materials about PTSD symptoms the attorney provided the client. This Article also notes the difficult ethical boundary between an attorney providing essential advice to a client about the nature of emotional and psychological damages versus improper coaching. The proposed exception would help discourage improper coaching and lead to the discovery of any improper coaching that had already occurred. Even where the information provided by the attorney was appropriate from an ethical standpoint, discovery of that information is essential to an accurate diagnosis and fairness to defendants.

Recent highly publicized trials and events have once again raised questions in the legal community and general public about attorney conduct and ethics. One area of concern is the extent to which attorneys have overstepped acceptable bounds by (over)zealously representing their clients. The legal system has seen abuses by prosecutors and

---

* Professor of Law, University of Washington; J.D., University of Pennsylvania, 1973; B.A., University of Virginia, 1969.
*** Clinical Professor, Department of Psychology, University of Washington; Ph.D., University of Wisconsin, 1972; B.A., University of Pittsburgh, 1967.
lawyers for both plaintiffs and defendants. Because unethical conduct on the part of attorneys most often occurs at the instance or with the consent of the client, it follows that unethical discussions between attorney and client do occur but are at least ostensibly protected by privilege. In such instances, should the opposing party be able to discover, or even introduce at trial, information provided by the attorney to the client? Commentators have suggested a number of situations in which they believe that an exception to attorney-client confidentiality should exist.¹

This Article addresses a relatively recent phenomenon: plaintiffs claiming damages based on allegations of Posttraumatic Stress Disorder (PTSD). Part I indicates the difficulty of accurately diagnosing PTSD, the potential for abuse by claimants in forensic settings, and the psychiatric evaluator’s need to know whether the claimant’s attorney provided information about PTSD symptoms prior to the evaluation. Part II reviews the historical rationales for the attorney-client privilege, along with the established exceptions and waivers, and concludes that despite the unquestioned value of attorney-client confidentiality, either existing exceptions or a new exception based on overriding policy concerns should be employed to permit the discovery, and possibly the admission, of certain information. Part III establishes the bases and parameters for a narrowly drawn exception for information provided to a PTSD claimant by the claimant’s attorney, when such information is necessary for an accurate psychiatric evaluation. Finally, this Article expresses concern that some lawyers may be playing an unethical role by coaching their clients regarding how to present PTSD symptoms to the forensic evaluator.

POSTTRAUMATIC STRESS DISORDER: THE DIFFICULTY OF ACCURATE DIAGNOSIS AND THE NEED TO CONSIDER INFORMATION PROVIDED TO CLAIMANTS

Posttraumatic Stress Disorder (PTSD) was recognized in 1980 by the American Psychiatric Association in the Third Edition of its Diagnostic and Statistical Manual. The diagnosis has undergone revision since that time but continues to maintain a basic assumption regarding causation whereby a defined class of traumatic events is linked to a defined class of symptoms. Thus, to obtain a PTSD diagnosis, an individual or individual’s loved ones must experience a traumatic event that was a threat to life or physical integrity (the Stressor Criterion), and the individual must react to the incident with a specific number of avoidance, reexperiencing, and hyper-arousal symptoms (the Symptom Criteria). This causal link between the traumatic event and a set of symptoms makes the a PTSD diagnosis particularly attractive in the furtherance of a personal-injury claim. Unlike other psychiatric diagnoses such as depression, where causes may be multiple or even biologically determined with no external precipitant, the cause of PTSD is clarified by the very act of making the diagnosis.

Soon after the establishment of the PTSD diagnosis, commentators expressed concerns about the potential for abuse in forensic and

---


4. The DSM-IV-TR provides that “[t]he essential feature of Posttraumatic Stress Disorder is the development of characteristic symptoms following exposure to an extreme traumatic stressor.” DSM-IV-TR, supra note 3, at 463.

5. Id. at 463–68.

6. Id. Ralph Slovenko has observed that:

In tort litigation, PTSD is a favored diagnosis in cases of emotional distress because it is incident specific. It tends to rule out other factors important to the determination of causation. Thus plaintiffs can argue that all of their psychological problems issue from the alleged traumatic event and not from myriad other sources encountered in life. A diagnosis of depression, in contrast, opens the issue of causation to many factors other than the stated cause of action.”

Ralph Slovenko, Legal Aspects of Post-Traumatic Stress Disorder, 17 PSYCHIATRIC CLINICS N. AM. 439, 441 (David A. Tomb ed., 1994).
disability settings. These concerns can be raised at several levels. First, there is the issue of whether a traumatic event actually occurred and if the claimant was indeed present. In support of these concerns the professional literature provides case reports on war-related PTSD claims by individuals who never saw combat, and sting operations have videotaped citizens scrambling onto a bus after a staged accident so they could present themselves as injured passengers.

Even if an event has occurred and the individual was present, there is the issue of how that individual was affected. This is a difficult issue to determine because the symptom criteria for diagnosing PTSD are generally subjective, easily coached, and easily simulated. Thus, an individual who has experienced a traumatic event needs only to report problems such as bad dreams, trouble sleeping, efforts to avoid thinking about the event, loss of interest in hobbies, and perhaps two or three other subjective complaints to fulfill criteria for the diagnosis.
The mental-health professional who serves as forensic expert faces a difficult task when assessing the validity of alleged problems resulting from trauma. The expert cannot assume that an event has occurred as the claimant describes or that the claimant even participated in the event as alleged. The expert also cannot assume a claimant actually has a posttraumatic stress disorder simply because the defining symptoms are reported. These problems are recognized in the American Psychiatric Association’s Diagnostic and Statistical Manual, which specifically cautions professionals to rule out malingering when PTSD presents in forensic and disability settings.\footnote{12}

Michael Trimble reflected on the attraction of a PTSD diagnosis in forensic settings and the problem of malingering when he noted that attorneys might coach clients in the furtherance of a claim: “[PTSD,] sanctioned so neatly by the DSM-III, clearly has both conceptual and medico-legal implications. . . . [It] will give a great deal of leverage to those seeking compensation and the counting off of symptoms in checklist fashion will become routine practice in many a lawyer’s office.”\footnote{13}

Recently, Trimble’s concerns received support when twenty survivors of a major marine disaster filed personal-injury claims for emotional damages, resulting in a high incidence rate of diagnosed PTSD.\footnote{14} As it

\footnote{12. The text of the DSM-IV states: “Malingering should be ruled out in those situations in which financial remuneration, benefit eligibility, and forensic determinations play a role.” DSM-IV, supra note 3, at 467.}


\footnote{14. Gerald M. Rosen, The Aleutian Enterprise Sinking and Posttraumatic Stress Disorder: Misdiagnosis in Clinical and Forensic Settings, 26 PROF. PSYCHOL.: RESEARCH & PRACTICE 82, 82–87 (1995). Records, reports, and depositions of treating doctors showed that nineteen of twenty litigating survivors had been seen by one or more mental-health professionals and all had presented with the hallmark features of PTSD. All mental-health professionals provided a diagnosis of PTSD, and these diagnoses were maintained for six months or longer. The survivors’ presentation of symptoms yielded a total incidence rate for diagnosed PTSD of eighty-six percent if it was conservatively assumed that non-assessed survivors did not have problems. Rosen pointed out that nowhere in the scientific literature has such a high incidence rate been reported for survivors of accidental trauma. Instead, epidemiological surveys find that fewer than twenty-five percent of accident survivors report reactions of sufficient severity to warrant a PTSD diagnosis. See, e.g., Naomi Breslau et al., Traumatic Events and Posttraumatic Stress Disorder in an Urban Population of Young Adults, 48 ARCHIVES GEN. PSYCHIATRY 216, 216–22 (1991); Fran H. Norris, Epidemiology of Trauma: Frequency and Impact of Different Potentially Traumatic Events on Different Demographic Groups, 60 J. CONSULTING & CLINICAL PSYCHOL. 409, 409–18 (1992). Recent methodological refinements in surveying techniques suggest rates of PTSD after accidents that are even lower. See, e.g., Naomi Breslau et al., Trauma and Posttraumatic Stress Disorder in the Community, 55 ARCHIVES GEN. PSYCHIATRY 620, 626–32 (1998).}
turned out, the high incidence of diagnosed PTSD among litigating survivors was explained, at least in part, by reports of attorney coaching and symptom sharing. Thus, several survivors discussed in follow-up interviews how they had received instructions from their counsel.\textsuperscript{15} Three reported being told they did not need to work.\textsuperscript{16} Another explained that he was told it might be worth his while to see a doctor every week.\textsuperscript{17} Another survivor reported that attorneys had explained to several crew members how people with PTSD had sleep problems, nightmares, and fears.\textsuperscript{18} Rosen noted in his report on the \textit{Aleutian Enterprise} sinking: “All told, 6 of 20 survivors provided unambiguous reports that some form of attorney advice had occurred with regard to symptoms of PTSD, not going back to work, or the merits of seeing a doctor.”\textsuperscript{19} The survivors reported that crew members shared this information with others in the group.\textsuperscript{20}

Findings from this maritime case do not represent an isolated instance. J. R. Youngjohn has documented a case of attorney coaching prior to the neuropsychological evaluation of a mild head-injury client.\textsuperscript{21} Further, in one survey of attorneys, sixty-three percent of those questioned felt it was appropriate to provide clients with information about psychological-test validity measures, when separate studies find that such coaching can reduce the performance of validity checks.\textsuperscript{22} Paul Lees-Haley has reviewed these findings and observed that “insofar as examinees’ responses to tests and interviews reflect preparation by an attorney for litigation, instead of the psychological status of the

---

\textsuperscript{15} Rosen, \textit{supra} note 14, at 82–87.
\textsuperscript{16} Id. at 84.
\textsuperscript{17} Id.
\textsuperscript{18} Id.
\textsuperscript{19} Id.
\textsuperscript{20} Id.
\textsuperscript{22} See Martha W. Wetter & Susan K. Corrigan, \textit{Providing Information to Clients About Psychological Tests: A Survey of Attorneys’ and Law Students’ Attitudes}, 26 PROF. PSYCHOL.: RES. & PRACTICE 474, 474–77 (1995). The finding that forty-seven to forty-eight percent of those surveyed “always or usually” provided clients with information about psychological-test validity measures, and sixteen percent “sometimes” provided such information, raises concerns for psychologists who also have found that coaching about validity scales can reduce the effectiveness of these scales. See Ruth A. Baer et al., \textit{Effects of Information About Validity Scales on Underreporting of Symptoms on the MMPI-2: An Analogue Investigation}, 2 PSYCHOL. ASSESSMENT 189, 189–200 (1995); see also Joanne Storm & John R. Graham, \textit{Detection of Coached General Malingering on the MMPI-2}, 12 PSYCHOL. ASSESSMENT 158, 158–65 (2000).
Confidentiality and Posttraumatic Stress Disorder

Examinee, the validity of psychologists’ expert opinions may be compromised. 23

In addition to professional peer-review articles that speak to the issue of attorney coaching, insurance “sting operations” have demonstrated that attorneys can play an active role in furthering the presentation of false claims. 24 In light of such findings, a responsible forensic evaluator would want to know whether a claimant’s attorney has provided information about psychiatric symptoms that might pertain to the claimant’s clinical presentation. Unfortunately, efforts to obtain such information inevitably run up against claims of attorney-client confidentiality.


24. See, e.g., Kerr, supra note 9, at A1. New Jersey transit companies reported that when buses had collisions in urban areas,

they would often be surrounded by ‘runners’ for doctors and lawyers who would get on the bus, hand out leaflets with phone numbers, and encourage passengers to say they suffered from back or neck injuries that are hard to disprove. The most sophisticated runners spent their days scanning police radio frequencies for word of traffic accidents. . . . One tape made inside a bus crashed in September 1992, shows a man jumping onto the bus three minutes after the accident and declaring to the passengers on board: “All you people who want to get paid you stay right there, stay down. Wait for the ambulance to come. Your neck hurts, your legs hurt. All of that. You’ll get some money. Stay there. They pay.”

Id. at D2.

Sting operations have contributed to public awareness that coaching occurs, to the extent that the phenomenon has now appeared in popular books of fiction, much to the detriment of attorneys' professional image. See G. M. Rosen, Posttraumatic Stress Disorder, Pulp Fiction, and the Press, 24 BULL. AM. ACAD. PSYCHIATRY & LAW, 267, 267–69 (1996), who cites Carl Hiassen’s Strip Tease (1993). In that pulp fiction book one of the leading characters places a cockroach in a container of yogurt and resesals the package, preparing to make a claim for PTSD. The character meets with his attorney and states:

“Tell me about your ace shrink,” to which counsel responds, “A good man. I’ve used him on other cases. You should start seeing him as soon as possible, and as often as possible. . . . It’s important to document your pain and suffering. It will help determine the final damages. . . . You might even consider quitting your job. . . . Lost income would greatly enhance a jury award.

Hiassen, supra, at 52.
II. ATTORNEY-CLIENT CONFIDENTIALITY: APPLICABILITY TO A PTSD CLAIM

A. Historical Perspective: Confidentiality To Promote Free Flow of Communication

Professor Wigmore stated the classic conception of the attorney-client privilege:

(1) Where legal advice of any kind is sought (2) from a professional legal advisor in his capacity as such, (3) the communications relating to that purpose, (4) made in confidence (5) by the client, (6) are at his instance permanently protected (7) from disclosure by himself or by the legal advisor, (8) except the protection be waived.25

The privilege is rooted in sixteenth-century common law.26 At that time, although parties were not allowed to testify at trial, their attorneys could be called to testify.27 Steeped in notions of the attorney as a gentleman, whose “first duty . . . is to keep the secrets of his client,”28 the privilege originally protected the attorney from the unwonted compromise of his “honor” that compulsory testimony regarding client confidences would entail.29 At its inception, the privilege belonged to the attorney, who could waive it.30

By the eighteenth century, ownership of the privilege shifted to the client, as did the ability to waive it.31 The privilege came to be seen as a

---


26. See 9 W.S. HOLDSWORTH, A HISTORY OF ENGLISH LAW 198–203 (1926); JOHN H. WIGMORE, EVIDENCE § 2291 (3d ed. 1940). It has been suggested that the privilege reflects “ideas at least as old as Roman Law.” CHARLES W. WOLFRAM, MODERN LEGAL ETHICS § 6.1.2, at 242 (1986) (citing Radin, The Privilege of Confidential Communication Between Lawyer and Client, 16 CAL. L. REV. 487 (1928)).

27. WOLFRAM, supra note 26, § 6.1.2, at 243.


29. Id. at 530–31; WOLFRAM, supra note 26, § 6.1.2, at 243.

30. WIGMORE, supra note 25, § 2290, at 544–45; see also ROBERT H. ARONSON & DONALD T. WECKSTEIN, PROFESSIONAL RESPONSIBILITY IN A NUTSHELL 196 (2d ed. 1991); WOLFRAM, supra note 26, § 6.1.2, at 243.

31. See In re Colton, 201 F. Supp 13, 15 (S.D.N.Y. 1961), aff’d, 306 F.2d 633 (2d Cir. 1962); see also Wolfram, supra note 26, § 6.1.2, at 243 (citing Annesley v. Earl of Anglesea, 17 How. St. Tr. 1139, 1225 (Ex. 1743) (opinion of Mountenay, B.)).
means of safeguarding the client’s freedom in consulting and benefiting from the assistance of counsel. The existence of the privilege was thought to encourage greater candor on the part of the client and to enhance the truth-seeking functions of the adversary system. The privilege protected both the client’s statements to the attorney and the attorney’s statements to the client. This was deemed by some to be essential because any statements by the attorney could reveal the nature of the client’s disclosures and could also be interpreted as admissions by the client.

**B. Definition of the Privilege and Rationale for Its Continued Use**

1. **Privilege: The Evidentiary Obligation To Protect Client Communications from Disclosure**

   Today, the common law privilege has been statutorily codified in most states and a number of U.S. Supreme Court decisions have interpreted and upheld the privilege, finding that its purpose is to encourage full and frank communication between attorneys and their clients and thereby promote broader public interests in the observance of law and administration of justice. The privilege recognizes that sound legal advice or advocacy

---


33. *Upjohn*, 449 U.S. at 389; *Fisher*, 425 U.S. at 403; *see* Trammel v. United States, 445 U.S. 40, 51 (1980) (noting that privilege assures that “advocate and counselor [can] know all that relates to the client’s reasons for seeking representation if the professional mission is to be carried out”).

34. *See*, e.g., *In re LTV Sec. Litig.*, 89 F.R.D. 595, 602-03 (N.D. Tex. 1981). *But see* Wells v. Rushing, 755 F.2d 376, 379 n.2 (5th Cir. 1985) (noting that content of communication may not be privileged to the extent that it contained communications from attorney to client). Some courts have noted that advice given by a lawyer to his client is privileged only if the advice is based on, or would reveal, confidential information furnished by the client. Mead Data Central, Inc. v. U.S. Dept. of Air Force, 566 F.2d 242, 254 (D.C. Cir. 1977); SMC Corp. v. Xerox Corp., 70 F.R.D. 508 (D. Conn. 1976); J.P. Foley & Co., Inc. v. Vanderbilt, 65 F.R.D. 523, 526 (S.D.N.Y. 1974).

35. Zacharias, *supra* note 1, at 358; *see also* ARONSON & WECKSTEIN, *supra* note 30, at 196. *But see*, e.g., *SCM Corp.*, 70 F.R.D. at 522.

36. For example, in Washington, the privilege is codified at REVISED CODE OF WASHINGTON 5.60.060(2) (2000): “An attorney or counselor shall not, without the consent of his or her client, be examined as to any communication made by the client to him or her, or his or her advice given thereon in the course of professional employment.” *See also*, e.g., CAL. EVID. CODE § 954 (2000); N.Y. C.P.L.R. 4503 (McKinney 2000).
serves public ends and that such advice or advocacy depends upon the lawyer’s being fully informed by the client.37

In spite of the privilege’s general acceptance, it has not been without its critics. Early commentators noted that if lawyers warned clients from the start not to discuss criminal or other improper conduct, the end result would be positive—that is, the truly guilty would receive less legal assistance.38 Others believed the argument that clients would not confide in lawyers absent a broad privilege was not compelling because surveys have shown much of the public does not believe an absolute privilege exists.39 Lawyers themselves are not completely comfortable with a rule that arguably has evolved as a tool to protect information unnecessarily when disclosure would be desirable.40 Professor Wigmore cautioned against an overly broad application of the privilege and urged that it be employed as “an exception to the general duty to disclose” that should be “strictly confined within the narrowest possible limits.”41

2. Privilege: The Ethical Obligation To Protect Client Communications from Disclosure

In addition to the evidentiary attorney-client privilege, there is also a general ethical duty of attorney-client confidentiality springing from the law of agency.42 While there is considerable overlap between the

37. Trammel, 445 U.S. at 51; see also Upjohn, 449 U.S. at 392; Fisher, 425 U.S. at 403–04. Some authorities have suggested that the privilege helps to prevent crime and other misconduct by encouraging clients to disclose contemplated wrongdoing, giving attorneys a chance to discourage such acts. See, e.g., Upjohn, 449 U.S. at 392. But cf. Zacharias, supra note 1, at 369–70 (indicating that there is no empirical evidence to support this argument).

38. Wigmore, supra note 25, at 549 (quoting Jeremy Bentham, 5 Rationale of Judicial Evidence 302–04 (1827)).

39. For example, a 1962 Yale Law School study found that 40 of 108 laypeople questioned believed that attorneys could be compelled to disclose all confidences in court. Zacharias, supra note 1, at 394.

40. In a survey of lawyers in upstate New York, more than one-third said they had asserted the privilege when they believed they should have been required to disclose information. Zacharias, supra note 1, at 382.

41. Wigmore, supra note 25, § 2292, at 554.

42. The ethical rule is codified in the ABA’s Model Rules of Professional Conduct (MRPC), Rule 1.6 (2000), which provides in part: “(a) A lawyer shall not reveal information relating to representation of a client unless the client consents after consultation, except for disclosures that are impliedly authorized in order to carry out the representation, [with limited exceptions].”
Confidentiality and Posttraumatic Stress Disorder

privilege and the ethical duty, there are notable differences between the two. As an evidentiary rule, the attorney-client privilege applies in judicial and other proceedings in which a lawyer may be called as a witness or otherwise compelled to produce evidence about a client. It also is limited to those matters communicated in confidence to an attorney by his or her client or by the attorney or client to an intermediary necessary for effective communication between the attorney and client.\footnote{See, e.g., United States v. Kovel, 296 F.2d 918, 920–21 (2d Cir. 1961); In re Consol. Litig. Concerning Int’l Harvester’s Disposition of Wis. Steel, 666 F. Supp. 1148, 1157 (N.D. Ill. 1987) (collecting cases).}

The lawyer’s ethical duty to maintain client confidentiality is much broader, encompassing the full range of obligations owed by an agent to a principal. Not limited to situations in which evidence is sought from the lawyer under the compulsion of law, the ethics rule forbids unauthorized disclosure of client confidences in any forum and regardless of the source of the information.\footnote{See, e.g., United States v. Kovel, 296 F.2d 918, 920–21 (2d Cir. 1961); In re Consol. Litig. Concerning Int’l Harvester’s Disposition of Wis. Steel, 666 F. Supp. 1148, 1157 (N.D. Ill. 1987) (collecting cases).} Within its scope, therefore, is all “information relating to the representation,” not just confidences shared by the client.\footnote{MODEL RULES OF PROF’L CONDUCT R. 1.6.} Under the more capacious ethical obligation, the lawyer as a trusted “agent” truly is the keeper of the client’s secrets.

C. Exceptions to the Privilege: Crime or Fraud

Although quite expansive, attorney-client confidentiality as embodied in both the evidentiary privilege and the ethical rules of professional conduct is subject to certain limited exceptions. Most notably, the attorney-client privilege does not attach to advice sought by the client in furtherance of a crime or fraud.\footnote{United States v. Zolin, 491 U.S. 554, 563 (1989); Clark v. United States, 289 U.S. 1, 15 (1933) (stating that attorney and client communications are no longer privileged where client seeks to commit fraud); see also Burton v. R.J. Reynolds Tobacco Co., 167 F.R.D. 134, 142–43 (D. Kan. 1996) (noting that evidence sufficient to establish prima facie showing that tobacco company attorneys were used to facilitate perpetration of continuing fraud to deceive American public about hazards of smoking was sufficient for crime or fraud exception to attorney-client privilege); Haines v. Liggett Group, Inc., 140 F.R.D. 681, 697 (D.N.J. 1992), order vacated on other grounds, 975} The rationale for this exception is that

A majority of jurisdictions has adopted this rule, with some modifications. ABA/BNA LAWYER’S MANUAL ON PROF’L CONDUCT § 55:104–107 (1993); see also MODEL CODE OF PROF’L RESPONSIBILITY DR 4-101; ARONSON & WECKSTEIN, supra note 30, at 197; Zacharias, supra note 1, at 361–62.


44. MODEL RULES OF PROF’L CONDUCT R. 1.6 cmt. 5 (2000); see also ARONSON & WECKSTEIN, supra note 30, at 198–200; ABA/BNA LAWYERS’ MANUAL ON PROFESSIONAL CONDUCT § 55:902.

45. MODEL RULES OF PROF’L CONDUCT R. 1.6.

46. United States v. Zolin, 491 U.S. 554, 563 (1989); Clark v. United States, 289 U.S. 1, 15 (1933) (stating that attorney and client communications are no longer privileged where client seeks to commit fraud); see also Burton v. R.J. Reynolds Tobacco Co., 167 F.R.D. 134, 142–43 (D. Kan. 1996) (noting that evidence sufficient to establish prima facie showing that tobacco company attorneys were used to facilitate perpetration of continuing fraud to deceive American public about hazards of smoking was sufficient for crime or fraud exception to attorney-client privilege); Haines v. Liggett Group, Inc., 140 F.R.D. 681, 697 (D.N.J. 1992), order vacated on other grounds, 975
a lawyer advising a client in the commission of a crime or fraud is not acting within his or her rightful capacity as a legal advisor, and the client is not entitled to invoke the privilege in such an instance. Similarly, if the communications regarding the commission of a crime or fraud are initiated by the lawyer but the client acts in accordance with the fraudulent advice, the privilege would be vitiated because the privileged relationship would not exist. The proponent of the crime or fraud exception must make a foundational showing of crime or fraud to obtain the otherwise privileged communications. The U.S. Supreme Court has held that before a trial court can even engage in an *in camera* review of the privileged materials, the proponent of the exception must make a showing adequate to support “a good faith belief by a reasonable person” that review of the materials will produce evidence sufficient to warrant application of the exception.

Under the American Bar Association (ABA) Model Rules, a lawyer may not knowingly counsel or assist a client in conduct that is criminal or fraudulent, nor may the lawyer use false evidence or commit a fraud upon the court. Also, a lawyer who learns a client intends to commit a criminal act that is likely to result in “imminent death” or “substantial bodily harm” has the discretion to reveal the information in order to

---


48. See Zolin, 491 U.S. at 572.

49. Id. at 571.

50. See id. at 15.

51. MODEL RULES OF PROF’L CONDUCT R. 1.2(d).

52. MODEL RULES OF PROF’L CONDUCT R. 3.3(a)(4). But see id. at R. 1.2(d): “[A] lawyer may discuss the legal consequences of any proposed course of conduct with a client and may counsel or assist a client to make a good faith effort to determine the validity, scope, meaning or application of the law.”
Confidentiality and Posttraumatic Stress Disorder

prevent commission of the crime.53 A lawyer may also reveal client confidences to the extent necessary to protect the lawyer from any liability that may emerge from the representation of the client.54 Some states have added exceptions for fraud,55 information involving possible bodily harm,56 and past client improprieties involving the lawyer’s services.57

D. Waiver of the Privilege

The attorney-client privilege may also be waived by the client. This waiver may occur either expressly or implicitly. An express waiver may be effected either in discovery or at trial when a client clearly and voluntarily agrees to waive the privilege.58 Also, in some instances, an “implied” waiver of the privilege is found to occur.59 When, for

53. Id. at R. 1.6(b)(1). In some states, this portion of the rule has been modified and expanded to allow a lawyer (in his or her discretion) to reveal client confidences to prevent the client from committing any crime. There is no requirement that the crime be one that is likely to result in “imminent death” or “bodily harm.” ABA/BNA LAWYER’S MANUAL ON PROFESSIONAL CONDUCT § 55:104–107 (1993). Arizona, Connecticut, Florida, Illinois, Nevada, New Jersey, North Dakota, Texas, and Wisconsin all have versions of RPC 1.6 that make mandatory disclosure of information reasonably necessary to prevent a client from committing a crime likely to result in death or substantial bodily harm. Id.; cf. RESTATEMENT (THIRD) OF THE LAW GOVERNING LAWYERS, § 117A (Proposed Final Draft No. 2 1998); ABA Ethics 2000 Commission’s Proposed Revision of RPC 1.6(b)(1).

54. MODEL RULES OF PROF’L CONDUCT R. 1.6(b)(2).


56. See, e.g., RULES REGULATING THE FLA. BAR R. 4-1.6(b)(2) (2000).

57. See, e.g., CONN. RULES OF PROF’L CONDUCT R. 1.6(c)(2) (2000); MD. RULES OF PROF’L CONDUCT R. 1.6(b)(2) (2000); NEV. SUPREME COURT RULES R. 156(3)(a) (2000).

58. See, e.g., G.S. Enters., Inc. v. Falmouth Marine, Inc., 571 N.E. 2d 1363, 1368–69 (Mass. 1991) (finding defendant’s express waiver of attorney-client privilege after formal discovery period had elapsed was not too late to preclude reliance on defense).

59. See, e.g., Chi. Title Ins. Co. v. Superior Court, 220 Cal. Rptr. 507, 512 (Cal. Ct. App. 1985) (finding that implied waiver where communication “goes to the heart of the claim”); Mountain States Tel. & Tel. v. DiFede, 780 P.2d 533, 543 (Colo. 1989) (finding that implied waiver results when party places in issue confidential communication going to heart of claim); League v. Vanice, 374 N.W.2d 849, 856 (Neb. 1985) (stating party impliedly waives privilege by raising issue central to his ability to maintain action); Jakobleff v. Cerrato, Sweeney & Cohn, 97 A.D.2d 834, 835 (N.Y. App. Div. 1983) (noting that privilege is impliedly waived when necessary to determine validity of claim and when application would deprive adversary of vital information); Kammerer v. W. Gear Corp., 96 Wash. 2d 416, 420, 635 P.2d 708, 711 (1981) (noting that when client intends to waive privilege, waiver cannot be delayed until trial because it would unduly limit discovery); see also United States v. Sanders, 979 F.2d 87, 92 (7th Cir. 1992) (noting that attorney-client-privilege objection is waived where defendant fails to make timely and specific objection to admission of privileged statement at trial); Hollins v. Powell, 773 F.2d 191, 196–97 (8th Cir. 1985) (noting that
example, the client discloses the communications to a third person,\textsuperscript{60} chooses to testify regarding privileged communications,\textsuperscript{61} or fails to object when others reveal the privileged communications.\textsuperscript{62}

The privilege is also waived when, as part of a claim or defense, the client places the subject matter of the confidence in issue.\textsuperscript{63} Reasoning that privileged materials ought not to be used as both a shield and a sword, courts have employed a “fairness doctrine”\textsuperscript{64} in finding a waiver of the privilege for matters in issue.\textsuperscript{65}

This waiver of the attorney-client privilege has been found where a client brings a claim of legal malpractice\textsuperscript{66} or asserts reliance on the attorney-client privilege waived for failure to object to questioning that led client to testify concerning portions of attorney-client communication); Love v. United States, 386 F.2d 260, 265 (8th Cir. 1967) (noting that failure of defendant to object to testimony of attorney who represented him in another case waived assertion of attorney-client privilege on appeal).

60. See, e.g., In re von Bulow, 828 F.2d 94, 101 (2d Cir. 1987) (holding that client’s acquiescence in attorney’s publication of widely read book containing privileged materials from client’s criminal case was sufficient to waive privilege for those matters in subsequent civil case); Denver Post Corp. v. Univ. of Colo., 739 P.2d 874, 881 (Colo. Ct. App. 1987) (holding that attorney-client privilege was waived where university disclosed to state auditor information gathered from internal investigation regarding contracts with Saudi Arabia); Dutton v. State, 452 A.2d 127, 145 (Del. 1982) (holding that revelation to close friend waived privilege); State v. Shire, 850 S.W.2d 923, 931 (Mo. Ct. App. 1993) (finding attorney-client privilege waived where defendant’s daughter was present during conversations between attorney and client and daughter’s presence was not necessary or essential to communication).

61. See State v. Ingels, 4 Wash. 2d 676, 104 P.2d 944 (1940); State v. Vandenberg, 19 Wash. App. 182, 187, 575 P.2d 254, 257 (1978) (holding client’s testimony as to fact of communication to attorney waives privilege for whole of that communication and all other communications to attorney on same subject matter).

62. E.g., Love, 386 F.2d at 265.


64. In re von Bulow, 828 F.2d at 101.

65. See Chevron Corp. v. Pennzoil Co., 974 F.2d 1156, 1162–63 (9th Cir. 1992) (holding that privilege was waived when client asserted affirmative defense of reliance on advice of counsel); Indep. Prods. Corp. v. Loew’s, Inc., 22 F.R.D. 266, 277 (S.D.N.Y. 1958); Wade’s Canadian Inn, 638 N.Y.S.2d at 801; Republic Ins. Co. v. Davis, 856 S.W.2d 158, 161–63 (Tex. 1993) (regarding “offensive use” exception).

66. E.g., Pappas v. Holloway, 114 Wash. 2d 198, 208, 787 P.2d 30, 36 (1990); WIGMORE, supra note 25, § 2327, at 638; see also Elia v. Pifer, 977 P.2d 796, 804 (Ariz. 1999) (holding that by filing legal malpractice action against domestic-relations attorney, client impliedly waived
advice of counsel as an affirmative defense. 67 In Hearn v. Rhay, 68 for example, the court required disclosure of legal advice given to defendants who asserted the defense of qualified immunity from a claim for damages under the Civil Rights Act based on their status as “public officials.” 69 Because the defendants contended that they acted in good faith and on the advice of legal counsel, the court found the content of their communications with counsel was “inextricably merged” with the affirmative defense and “inhere[d] in the controversy itself.” 70 The court stated the benefit to be gained from disclosure outweighed any potential damage to the attorney-client relationship because access to the materials was required for a “fair and just determination” of the issues. 71

The Hearn court applied a three-part test to determine whether the privilege had been implicitly waived:

1. assertion of the privilege was a result of some affirmative act, such as filing suit, by the asserting party;
2. through this affirmative act, the asserting party put the protected information at issue by making it relevant to the case; and
3. application of the privilege would have denied the opposing party access to information vital to his defense. 72

A number of courts have followed Hearn. 73 Other courts, however, have criticized Hearn for creating an overly broad “placing-at-issue” exception and thereby undermining the legislative intent of the privilege by requiring a balancing of the interests

67. Pennzoil, 974 F.2d at 1161.
68. 68 F.R.D. 574 (E.D. Wash. 1975).
69. See id. at 578, 581.
70. Id. at 582.
71. Id.
72. Id. at 581.
of the privilege holder against those of the opponent.\textsuperscript{74} For example, the Third Circuit criticized \textit{Hearn} in \textit{Rhone-Poulenc Rorer, Inc. v. Home Indemnity Co.}\textsuperscript{75} The court noted it is essential that clients know from the start what confidences will never be revealed.\textsuperscript{76} By making later disclosure dependent on the particular facts of future litigation, the Third Circuit asserted that the \textit{Hearn} court had destroyed that predictability.\textsuperscript{77} The court instead adopted a narrower waiver rule: “The advice of counsel is placed in issue where the client asserts a claim or defense, and attempts to prove that claim or defense by disclosing or describing an attorney-client communication.”\textsuperscript{78}

However, \textit{Hearn} is not, or need not be, as open-ended as the Third Circuit implied. First, the decision places a heavy burden on the party seeking disclosure by requiring it to demonstrate the necessity of disclosure.\textsuperscript{79} Courts can add further safeguards by requiring that a party also show that it has exhausted every reasonable alternative source of the information sought and that it describe the information with reasonable specificity.\textsuperscript{80}

The “at issue” exception to attorney-client privilege corresponds to the exceptions to the physician-patient and psychotherapist-client privileges that apply in personal-injury and wrongful-death actions. Several states have enacted exceptions to those privileges in situations where the patient puts his or her physical, mental, or emotional state at issue in litigation. For example, under Rule 510 of the Texas Rules of Evidence, a plaintiff who seeks damages for alleged injuries, such as mental anguish, may not suppress communication or records relevant to his or her mental and emotional condition.\textsuperscript{81}

\begin{footnotesize}
\begin{enumerate}
\item \textsuperscript{74} \textit{E.g.}, Succession of Smith v. Kavanaugh, Pierson & Talley, 513 So. 2d 1138, 1145–46 (La. 1987).
\item \textsuperscript{75} 32 F.3d 851 (3d Cir. 1994).
\item \textsuperscript{76} \textit{See id.} at 863–84.
\item \textsuperscript{77} \textit{Id.} at 864.
\item \textsuperscript{78} \textit{Id.} at 863.
\item \textsuperscript{79} \textit{Hearn v. Rhay}, 68 F.R.D. 574, 581–82 (E.D. Wash. 1975).
\item \textsuperscript{80} \textit{See, e.g.}, Greater Newburyport Clamshell Alliance v. Pub. Serv. Co. of N.H., 838 F.2d 13, 18 (1st Cir. 1988) (holding, in case involving prosecution for criminal trespass, that only privileged communications that were made while undercover informant was present were subject to discovery).
\item \textsuperscript{81} \textit{Texas Rules Ann. R. 510(d)(5)} (Vernon 2001); \textit{see also Haw. R. Evid. 504.1(d)(3)} (1995) (establishing psychologist-client privilege but stating: “There is no privilege under this rule as to a communication relevant to the physical, mental, or emotional condition of the client in any proceeding in which the client relies upon the condition as an element of the client’s claim or
Confidentiality and Posttraumatic Stress Disorder

states, there is an automatic waiver of the litigant’s physician-patient privilege after the commencement of the action. Likewise, the privilege is waived in the criminal context when a psychiatrist, formerly a member of the defense team, is called as a witness by the State, after the defendant has raised the insanity defense. By placing his or her mental state in issue, a defendant waives any privilege with respect to that issue.

Even when the attorney-client privilege is applicable, the privilege may be waived. Such a waiver may be express or implied. A widely recognized implied waiver of the physician-patient privilege occurs when the patient places his or her physical or mental state in issue. A similar waiver of the attorney-client privilege would be appropriate in civil cases.

E. Distinction Between Civil and Criminal Cases

While the fundamentals of attorney-client confidentiality are the same in criminal and civil suits, it is important to distinguish the underlying rationales for confidentiality in these two very different realms. In the criminal context, the defendant may bolster the underlying rationales of the attorney-client privilege with constitutional protections against self-

defense.”); LA. REV. STAT. ANN. § 37.2363 (West 2000) (establishing psychologist-client privilege except “[w]here the patient or client puts his mental state in issue by alleging mental or emotional damages or condition in any judicial or administrative proceedings”); MD. CODE ANN. § 9-109(d)(t) (2001) (establishing psychiatrist-client and psychologist-client privilege except when “the patient introduces his mental condition as an element of his claim or defense”); NEB. REV. STAT. ANN. § 71-1, 206.29 (Michie Supp. 2000) (noting that psychologist-client privilege exists except “[w]hen the client or patient, by alleging mental or emotional damages in litigation, puts his or her mental state in issue”); TENN. CODE ANN. § 24-1-207 (Supp. 2000) (stating that communication between psychiatrist and patient is privileged except in proceedings “in which the patient raises the issue of the patient’s mental or emotional condition;” if patient raises issue, communication shall be disclosed only after finding by court that “the evidence cannot be obtained from another source” or that “the evidence sought is necessary to avoid substantial injustice to the party seeking it, and either that the disclosure will result in no significant harm to the patient or that it would be substantially unfair as between the requesting party and the patient not to require the disclosure”).

82. See, e.g., WASH. REV. CODE § 5.60.060(4)(b) (2000) (stating that privilege is waived ninety days after commencement of action); see also CAL. EVID. CODE ANN. § 996 (West 2001) (codifying exception to physician-patient privilege for personal-injury actions); MICH. COMP. LAWS ANN. § 600.2157 (West 2000) (stating that patient waives physician-patient privilege by bringing personal-injury action or malpractice suit), § 600.2912f (stating that person who has given notice or commenced action alleging medical malpractice waives physician-patient privilege).

the attorney-client privilege with constitutional protections against self-incrimination under the Fifth Amendment, 84 the right to the assistance of counsel under the Sixth Amendment, 85 and the presumption of innocence. Further, zealous advocacy by defense counsel is necessary to offset the resources the police and prosecutors can marshal to prove the defendant’s guilt. Therefore, even when a criminal defendant admits guilt to his or her attorney, the adversary system assumes that greater justice for more people will be achieved if the attorney is not permitted to reveal the information. 86 This rationale supports the justice system’s fundamental assumption that it is preferable to let ten guilty people go free than to have one innocent person convicted. 87

In the civil realm, by contrast, these constitutional considerations do not come into play. There is, for example, no presumption of innocence and no privilege against self-incrimination. Consequently, the civil litigant may be required to produce evidence about the facts underlying an attorney-client communication, even if incriminating, where the criminal defendant would be protected from doing so under the Fifth Amendment. Likewise, different considerations apply when the privilege is used to shield from scrutiny a plaintiff’s claim for damages due to emotional stress, particularly when the accurate PTSD diagnosis is so dependent on whether the symptoms occurred after “prompting” by one or more third persons.

In spite of these inherent distinctions, many of the Rules of Professional Conduct, including the strict rule of confidentiality, are premised on a criminal-law vision of lawyers’ roles and obligations. Noting that the Rules of Professional Conduct state rules applicable to lawyers in all fields of practice, resulting in rules that in some cases are irrelevant and in others are contrary to certain highly specialized practice areas, some legal ethicists have proposed the adoption of more

85. United States v. Henry, 447 U.S. 264, 295 (1980) (Rehnquist, J., dissenting) (“[T]he Sixth Amendment, of course, protects the confidentiality of communications between the accused and his attorney.”).
86. See People v. Belge, 372 N.Y.S.2d 798, 802 (1975); MODEL CODE OF EVID. R. 210 cmt. a (1942).
87. Furman v. Georgia, 408 U.S. 238, 368 n.158 (1972) (Marshall, J., concurring) (quoting Justice William O. Douglas in Foreword, J. FRANK & B. FRANK, NOT GUILTY, 11–12 (1957)). But see 7 THE WORKS OF JEREMY BENTHAM 474–75 (J. Bowring ed. 1843) (suggesting that lawyer whose representation enabled client whom he knew to be guilty to escape punishment was virtually accessory after fact to crime).
flexible rules of attorney-client confidentiality to account for these differences.88

In harmony with this approach, the American Academy of Matrimonial Lawyers’ (AAML) Bounds of Advocacy elaborates on, and occasionally modifies, the Rules of Professional Conduct in the family-law area.89 Its standards specifically provide for an exception to the rule of confidentiality, and hence, permit disclosure that an attorney reasonably believes necessary to prevent substantial physical or sexual abuse of a child.90 Similarly, many states have carved out exceptions to various evidentiary privileges and duties of confidentiality in their child abuse reporting laws. Some states have compelled attorney reporting,91 while others have left it in the attorney’s discretion.92 Most states’ statutes expressly state that reporting of child abuse will not constitute a violation of the confidential communications privileges of most professions but make no mention of the attorney-client privilege.93 As the AAML’s Bounds of Advocacy suggests, however, allowing a family

88. See, e.g., William H. Simon, Ethical Discretion in Lawyering, 101 HARV. L. REV. 1083, 1090 (1988); Stanley Sporkin, The Need for Separate Codes of Professional Conduct for the Various Specialties, 7 GEO. J. LEGAL ETHICS 149 (1993); David B. Wilkins, Legal Realism for Lawyers, 104 HARV. L. REV. 468 (1990); Zacharias, supra note 1, at 400; see also Aronson, supra note 1, at 150–51.


90. AAML BOUNDS OF ADVOCACY (2000) Goal 6.5 (“An attorney should disclose information relating to a client or former client to the extent the lawyer reasonably believes necessary to prevent substantial physical or sexual abuse of a child.”); AAML BOUNDS OF ADVOCACY (1991) Standard 2.26 (“An attorney should disclose evidence of a substantial risk of physical or sexual abuse of a child by the attorney’s client.”). A footnote recognizes, however, that such disclosure would be prohibited in some jurisdictions as “conduct adverse to the client and based on confidential information.” BOUNDS OF ADVOCACY (2000) n.79; BOUNDS OF ADVOCACY (1991) n.42.


93. See, e.g., WASH. REV. CODE § 26.44.060(3) (2000) (“[C]onduct conforming with the reporting requirements of this chapter shall not be deemed a violation of the confidential communication privilege of RCW 5.60.060 (3) [priest/clergyman] and (4) [physician], 18.53.200 [optometrist] and 18.83.100 [psychologist]”); see also Mary Harter Mitchell, Must Clergy Tell? Child Abuse Reporting Requirements Versus the Clergy Privilege and Free Exercise of Religion, 71 MINN. L. REV. 723, 734 n.53 (1987) (“[T]he express retention of the attorney-client privilege [in most statutes] is noteworthy in light of the frequently conceded weakness of its rationale.”).
lawyer the discretion to reveal confidences in order to protect the best
interests of a child would truly serve the interests both of the family law
system and the public in keeping children from harm. 94 Even if it were
necessary to safeguard the abuser’s criminal system protections by
prohibiting use of such confidential information in any criminal
proceeding, at least the child would be better protected and any custody
determination would be made on the basis of all relevant facts.

Constitutional support for the attorney-client privilege of criminal
defendants is not applicable to civil penalties. In the criminal context, a
broad privilege with narrow exception preserves the defendant’s
presumption of innocence and Fifth and Sixth Amendment rights. The
principles do not support preventing discovery of information about
PTSD provided to a civil litigant claiming damages based on PTSD.

F. Bases for Finding PTSD Claims

A waiver or exception to the attorney-client privilege for information
about PTSD symptomology provided to a plaintiff claiming damages
due to PTSD could be justified on a number of bases. First, since the
attorney-client privilege shields relevant information from the trier of
fact, it should be narrowly construed consistent with its purpose:
encouraging clients to confide fully in their attorneys. As such, it would
not include materials and information about PTSD diagnosis and

94. AAML BOUNDS OF ADVOCACY, Goal 6.5 cmt. As stated in the Comment to the ABA Ethics
2000 Commission’s Proposed Revision of RPC 1.6(b)(1):

Although the public interest is usually best served by a strict rule requiring lawyers to preserve
the confidentiality of information relating to the representation of their clients, the
confidentiality rule is subject to limited exceptions. In becoming privy to information about a
client, a lawyer may foresee that the client intends serious harm to another person. However, to
the extent a lawyer is required or permitted to disclose a client’s purposes, the client will be
inhibited from revealing facts which would enable the lawyer to counsel against a wrongful
course of action. The public is better protected if full and open communication by the client is
couraged than if it is inhibited. Paragraph (b)(1) recognizes the overriding value of life and
physical integrity and permits disclosure reasonably necessary to prevent reasonably certain
death or substantial bodily harm. Substantial bodily harm includes life-threatening or
debilitating injuries and illnesses and the consequences of child sexual abuse. Such harm is
reasonably certain to occur if it will be suffered imminently or if there is a present and
substantial threat that a person will suffer such harm at a later date if the lawyer fails to take
action necessary to eliminate the threat.

See also RESTATEMENT (THIRD) OF THE LAW GOVERNING LAWYERS § 117A (Proposed Final Draft
No. 2 1998).
I see no reason to broaden the privilege beyond the narrow standard as set forth in United Shoe. The purpose of the privilege is to ensure that the client may confide in his attorney to obtain legal advice. Unless the legal advice reveals what the client has said, no legitimate interest of the client is impaired by disclosing the advice. Wigmore’s discussion of the broader standard concedes that it is not based on “any design of securing the attorney’s freedom of expression.” 8 Wigmore, supra, § 2320, at 629. His rationale is twofold: to prevent use of the attorney’s statements as admissions of the client or as revelations of the client’s communication. The latter reason, of course, is safeguarded by the explicit requirement of the narrow standard. That same requirement seems adequate to protect against admissions entitled to protection. If the attorney’s statement is based upon a fact communicated by the client, its use as an admission is barred by the availability of the privilege, even under the narrow standard. If the attorney makes an admission as agent for his client, without revealing a fact confidentially communicated from the client, no reason appears why such admission should be protected. Thus, Wigmore’s reasons for adopting the broad standard are adequately satisfied by the narrow standard. 96

Second, where a prima facie case of fraudulent conduct can be established, the information could be obtained under the crime or fraud exception to the attorney-client privilege. Third, the claim for damages based on PTSD could be deemed a waiver of the attorney-client privilege, as is the case with medical records. By placing the plaintiff’s symptoms of PTSD in issue, fairness to the defendant would require access to information the client received about PTSD prior to reporting those symptoms.97 Such information is necessary to the defendant’s mental-health expert in evaluating the plaintiff’s condition and as a basis for evaluating the plaintiff’s credibility at trial.

95. 70 F.R.D. 508 (D. Conn. 1976).
96. Id. at 522.
Further, to the extent that plaintiffs will testify about their symptoms, they arguably waive the confidentiality of information provided to them prior to trial. Under Federal Rule of Evidence 612, the opposing party may examine a document used “before testifying, if the court in its discretion determines it is necessary in the interests of justice.” The authorities are split as to whether review by the witness of attorney-prepared documents waives the attorney-client privilege and work-product protection.98

In jurisdictions that grant the trial judge discretion to authorize discovery of otherwise privileged materials provided to a witness, materials concerning PTSD symptoms would present the strongest case for permitting discovery. As noted above, it would be virtually impossible for the defendant’s mental-health evaluator to make an accurate PTSD determination without access to materials provided to the claimant.99 Similarly, the claimant’s testimony at trial concerning the symptoms would be inextricably intertwined with materials read concerning PTSD symptoms prior to claiming to suffer such symptoms.100 In jurisdictions where the above bases for finding a waiver of or exception to the attorney-client privilege would not be applicable, however, a new exception for information about PTSD diagnosis and symptoms should be established.

III. A NEW EXCEPTION TO ATTORNEY-CLIENT PRIVILEGE

A. The Need for an Exception when Assessing PTSD Claims

When psychiatrists and psychologists evaluate a patient in a clinical setting for PTSD symptoms, they generally rely on the patient’s statement of symptoms and their causes as the primary, if not sole, source of data.101 This is understandable and meets the standard of care because it is assumed that the patient is motivated to report accurately in

99. See supra notes 7–24 and accompanying text.
100. See infra notes 104–05 and accompanying text.
order to obtain an appropriate diagnosis and treatment.\textsuperscript{102} In forensic and disability settings, additional motivational factors may influence what an individual reports, and reliance on an individual’s self-reporting of symptoms is ill advised.\textsuperscript{103}

It is in the context of these additional motivational factors that concerns about a malingered PTSD diagnosis apply. Individuals can malinger PTSD symptoms on their own, with the assistance of relevant reading materials, or with the benefit of coaching by relatives, friends, or counsel. In every case, it is necessary to evaluate these sources of influence. After all, the goal of an independent exam is for the expert to assess for the court the plaintiff’s emotional status. To accomplish this goal, the examiner must assess all factors that have contributed to the client’s presentation of symptoms, including the possibility of extraneous factors that serve to educate the individual on symptoms not truly present. In the forensic setting, all of these sources of influence can be assessed by the expert except information provided by the attorney. Thus, for example, in \textit{Nelsen v. Research Corp. of the University of Hawaii},\textsuperscript{104} a treating doctor provided information on psychiatric symptoms of PTSD to the plaintiff Nelsen. The court stated: “I am not persuaded that Nelsen in fact suffered the full range of PTSD symptoms, based on the evidence indicating a lack of candor on the part of Nelsen, and in light of the testimony at trial that Dr. Poletti provided Nelsen with the DSM-III-R list of PTSD symptoms.”\textsuperscript{105} In this context, an exception

\textsuperscript{102} See Greenberg & Shuman, \textit{supra} note 101, at 53; Strasburger et al., \textit{supra} note 101, at 450–51.

\textsuperscript{103} See Greenburg & Shuman, \textit{supra} note 101; Strasburger et al., \textit{supra} note 101. The role of a therapist is considered incompatible with the role of a forensic examiner because of the differing approach to data demanded by each function. A similar concern underlies common law restrictions on the statements-of-physical-condition exception to the hearsay rule. See, e.g., Christopher B. Mueller & Laird C. Kirkpatrick, \textit{Evidence} § 8.42, at 938–39 (2d ed. 1999).

The main reason to admit statements made for purposes of getting treatment is that they are trustworthy. Usually they are made by the patient to his physician, and they describe past and present physical sensations relating to his condition, so risks of misperception and faulty memory are minimal. He knows his description helps determine treatment, so he has reason to speak candidly and carefully, and risks of insincerity and ambiguity are minimal.

Statements for purposes of diagnosis offer no similar assurance of candor. The speaker is likely to be trying either to prepare a doctor to testify or aid in litigation, or to get some other personal benefit (job or insurance coverage). He is tempted to maximize or minimize, overstate or understate injuries, pains, symptoms, and disabilities.

\textit{Id.}; see also 2 John W. Strong et al., \textit{McCormick on Evidence} § 277, at 233 (5th ed. 1999).


\textsuperscript{105} Id. at 844–45.
to the attorney-client privilege is needed when mental-health professionals are asked to determine the validity of PTSD claims.

An unusual turn of events in one personal-injury case illustrates the problem faced by experts attempting to assess claims of PTSD. In Mallory v. Supreme Alaska Seafoods, Inc., the plaintiff was asked at deposition if he had read anything about PTSD, a condition that had been diagnosed after plaintiff’s attorney had referred his client to a psychologist. Plaintiff’s counsel objected to the question on the basis of attorney-client privilege. The issue was taken before the judge, who ruled that the nature of materials provided were privileged but the materials were not. The plaintiff then disclosed that his attorney had provided deposition transcripts of other clients who, on referral by this same attorney to the same psychologist, had all presented the classic PTSD symptoms.

B. Policy Bases and Parameters for a New Exception

1. The Exception Should Be Narrowly Drawn To Avoid Detrimental Effect on Attorney-Client Relationship

As discussed above, one of the primary rationales underlying attorney-client confidentiality is the facilitation of frank and open client-attorney communications. The attorney-client privilege’s aim is both to facilitate the full development of facts necessary for effective representation and to help a lawyer advise clients in the proper exercise of their legal rights and the avoidance of any violation of the law. In the long run, lawyers will be better able to assist clients to act properly if they have all of the relevant information.

None of the policy rationales in support of the privilege would be frustrated by the adoption of a limited waiver of attorney-client confidentiality in the situation of a plaintiff claiming PTSD as a source of damages. The scope of the waiver would be extremely limited,

107. See id. at 2.
108. Id. Because written and oral information provided by any party other than counsel would not be privileged, only coaching provided by attorneys is subject to privilege.
109. See supra notes 32–33 and accompanying text.
Confidentiality and Posttraumatic Stress Disorder

including only those matters directly related to the nature, diagnosis, and symptoms of PTSD. Because the waiver would take effect only when a litigant placed those matters in issue, it would be consistent with the patient-litigant exception to the physician-patient privilege. In the context of a claim for damages resulting from PTSD, it makes little sense to require disclosure of the plaintiff’s communications with psychological professionals while shielding communications from the plaintiff’s attorney that preceded the consultations with the mental-health professionals and very likely influenced them.

In addition, the waiver could be limited to those statements and materials about PTSD symptoms provided by the attorney to the client. Some jurisdictions already limit the privilege to information communicated by the client to the attorney, rather than the reverse.\footnote{111 See, e.g., SCM Corp. v. Xerox Corp., 70 F.R.D. 508, 520–23 (D. Conn. 1976).} There is no evidence that such a limitation has negatively affected the attorney-client relationship.

2. Information Is Essential to Accurate Diagnosis and Assessment of Credibility

Once a plaintiff claims damages based on PTSD, the defendant has the right to have the plaintiff evaluated by a mental-health professional. Finding out whether a plaintiff’s symptoms appeared spontaneously or after he or she was provided with information about PTSD symptoms is essential to an accurate evaluation. The importance of evaluating the effect of information about PTSD symptoms is not diminished if the information was provided by the attorney.

Consider how a mental-health professional would go about assessing the possibility of attorney influence on a plaintiff’s presentation of symptoms. The issue could be approached by first asking a general question such as: “Have there been any sources of information that have influenced your knowledge of post-trauma reactions or your presentation of symptoms today?” This type of question covers the possibility that the plaintiff has acquired information from news articles, books, friends, or other sources, including counsel. Only an affirmative response to the general question would be followed by more focused ones such as, “Have any of your friends who are war veterans spoken to you about their post-trauma experiences?” In a similar vein the evaluator might ask, “With specific reference to your attorney, has he or she advised...
how you might present your problems today?,” or “Has counsel ever advised you or provided information concerning how people react to traumatic events?” Both the general question and more specific questions open the door for a plaintiff to discuss what may be considered privileged communications, though only in a very limited area of concern—and only if the attorney has in fact provided information about PTSD symptoms.

The arguments noted above for applying the “at issue” exception to attorney-client confidentiality would also justify a new exception.112 Because the proposed waiver or exception would only take effect when a litigant places those matters in issue, it would be in line with the matter-in-issue exception. By asserting PTSD as a source of damages, the plaintiff would be deemed to have placed any oral or written statements by the attorney regarding the nature, diagnosis, and symptoms of PTSD in issue and thus should be considered to have waived the privilege regarding those particular communications.

Just as a personal-injury plaintiff who waives his or her physician-plaintiff privilege on bringing the action is not thereby precluded from seeking medical assistance, so too the proposed waiver should not prevent a client from seeking a lawyer’s advice. Lawyers may ethically provide information concerning the symptoms of PTSD to clients prior to their psychological evaluation.113 But if they do provide such information, the evaluator must have access to the information to make an accurate evaluation. At the very least, the evaluator should be permitted to ask about prompting or bases for the symptoms other than spontaneous occurrence, and then be permitted to indicate, as part of the basis for the evaluation, if the plaintiff had refused to answer. This background information would also be vital for the jury to make an accurate assessment of the credibility of the plaintiff’s condition and the expert’s evaluation thereof.

3. A Limited Exception Is Necessary in Fairness to Defendants

While attorney-client confidentiality serves many laudable goals, there are strong countervailing policies favoring disclosure when litigants assert PTSD as a source of damages and there is a possibility of attorney coaching concerning the symptoms. For one, assertion of the

112. See supra notes 81–83 and accompanying text.
113. See infra Part III.B.4.
Confidentiality and Posttraumatic Stress Disorder

privilege in this instance blocks the defendant’s access to materials that are crucial to the claim asserted against the defendant. Courts in similar situations have weighed the equities and come out on the side of a waiver.114 Starting its analysis with a “presumption in favor of preserving the privilege,” one court found that

In a civil damages action, however, fairness requires that the privilege holder surrender the privilege to the extent that it will weaken, in a meaningful way, the defendant’s ability to defend. That is, the privilege ends at the point where the defendant can show that the plaintiff’s civil claim, and the probable defenses thereto, are enmeshed in important evidence that will be unavailable to the defendant if the privilege prevails. The burden on the defendant is proportional to the importance of the privilege. The court should develop the parameters of its discovery order by carefully weighing the interests involved, balancing the importance of the privilege asserted against the defending party’s need for the information to construct its most effective defense.115

Following this reasoning, a PTSD plaintiff’s claim may be “enmeshed” in the content of the attorney-client communications. Accurate diagnosis of PTSD is dependent on the client’s truthful reporting of certain key symptoms and the psychotherapist’s knowledge of whether the client was provided prior information about those symptoms. Therefore, it is impossible for defendants to construct a defense against coached manifestations of PTSD if they are prevented from gaining access to the very communications that contain the relevant evidence. Where the evidence is crucial to the defendant’s ability to defend against a claim and the disclosure is narrowly limited to information about PTSD symptoms provided to a plaintiff asserting PTSD as a basis for a damages award, the balance of equities in such a personal-injury action should tip toward disclosure.

114. See, e.g., Greater Newburyport Clamshell Alliance v. Pub. Serv. Co. of N.H., 838 F.2d 13, 20 (1st Cir. 1988) (holding, in case involving prosecution for criminal trespass, that only privileged communications which were made while undercover informant was present were subject to discovery); State v. Pawlyk, 115 Wash. 2d 457, 463–65, 800 P.2d 338, 341–42 (1990).

115. Newburyport, 838 F.2d at 20.
4. An Exception Will Assist in Discouragement and Discovery of Fraud

In addition to fairness to defendants, an exception to the privilege would serve to discourage fraudulent claims of PTSD. If plaintiffs (and their lawyers) are aware that placing their condition in issue by asserting the claim will make information they have seen or heard regarding the symptoms of PTSD nonprivileged, they may be less likely to fabricate the elements of the claim and attorneys may be less likely to prompt clients with information they know can be revealed to the evaluator.

As the exceptions suggest, the privilege does not have a firm constitutional basis as it is applied in civil litigation. The Sixth Amendment provides some protection from disclosure in criminal cases. But in civil cases, the policy favoring liberal discovery—a relatively recent development compared to the attorney-client privilege—weighs heavily in favor of limiting the privilege.

In an adversary system, the attorney is often placed in an extremely difficult position. On the one hand, the attorney has an obligation to give clients the benefit of the attorney’s knowledge and expertise. In the context of a client involved in a traumatic event, the client is likely to inquire about the likelihood and nature of potential damages. The attorney would be remiss in failing to advise that, in addition to damages for physical injury, lost wages, and harm to personal property damages may be awarded for emotional and psychological trauma.


The concept that a spirit of cooperation and forthrightness during the discovery process is necessary for the proper functioning of modern trials is reflected in decisions of our Court of Appeals. In Gammon v. Clark Equip. Co., 38 Wash. App. 274, 686 P.2d 1102 (1984), aff’d, 104 Wash. 2d 613, 707 P.2d 685 (1985), the Court of Appeals held that a new trial should have been ordered because of discovery abuse by the defendant. Then Court of Appeals Judge Barbara Durham wrote for the court:

The Supreme Court has noted that the aim of the liberal federal discovery rules is to “make a trial less a game of blindman’s bluff and more a fair contest with the basic issues and facts disclosed to the fullest practicable extent.” The availability of liberal discovery means that civil trials no longer need be carried on in the dark. The way is now clear . . . for the parties to obtain the fullest possible knowledge of the issues and facts before trial.

Id.
Suppose, for example, that the client was injured in a plane crash. In the ensuing lawsuit against the airline, the client would seek damages, including damages for psychological and emotional trauma. Lawyers may properly provide information about the diagnosis and symptoms of PTSD so that clients might determine whether they have experienced such symptoms. A plane-crash victim who has had nightmares or is afraid to take airplane trips may not realize such effects can form the basis for a PTSD diagnosis or resulting damages. Discussion with the client’s attorney about potential symptoms may be necessary to protect the client’s rights. For example, the attorney might ask specific questions about unusual fears, dreams, relationship problems, inability to engage in certain activities, and other difficulties that can occur post-trauma.

While such questions may be permissible, an attorney actually telling a client to study a list of symptoms, a book on PTSD, or sample depositions of clients who successfully obtained a PTSD diagnosis would appear to constitute impermissible coaching. There is a fine line between necessary and desirable preparation and unethical coaching; between helping the client understand and express existing trauma reactions and assisting the client to create symptoms that did not previously exist. To the extent that lawyers assist in the creation of symptoms that did not otherwise exist, they violate the Rules of Professional Conduct.

In addition to providing the forensic evaluator information essential to an accurate diagnosis, disclosure of information provided by the attorney would serve as a disincentive to “create” symptoms that did not otherwise exist. At the least, a client who described symptoms in the precise language of a list, book, or deposition transcript provided by the

---

118. As Professor Wolfram stated:

The line between legitimate and disreputable preparation is largely one of the lawyer’s intentions. A lawyer may legitimately attempt to refresh a witness’ memory, to assist the witness to testify in a straightforward and effective way, and to help the witness be prepared to meet improper or suggestive lines of hostile examination. On the other hand, a lawyer may not assist or school a witness in twisting or distorting the witness’ subjective memory and, thus, the truth as far as the witness knows it.

WOLFRAM, supra note 26, § 12.43, at 648 (footnotes omitted) (commenting that witness must also “testify in his or her own words” and not “simply to parrot words and phrases supplied by a lawyer”).

119. See, e.g., MODEL RULES OF PROF’L CONDUCT R. 1.2(d) (assisting client in conduct that is fraudulent); MODEL RULES OF PROF’L CONDUCT R. 3.3(a)(4) (fraud upon court); MODEL RULES OF PROF’L CONDUCT R. 3.4 (fairness to opposing party and counsel).
attorney would lose credibility and the likelihood of a diagnosis of PTSD. And, in extreme cases, fraudulent conduct would be exposed.\footnote{An example of how the attorney-client privilege in the civil context obscured the truth for too long can be found in the tobacco litigation. For nearly forty years, the tobacco companies succeeded in hiding a massive fraud on the American public by invoking the attorney-client privilege. When the veil of deception was finally pierced, it was revealed that counsel for the tobacco companies participated in the creation of an elaborate public-relations scheme to encourage the public to continue smoking cigarettes by pretending to investigate the health hazards of smoking. In reality, the tobacco companies were already in possession of information that smoking was hazardous to the health, which they deviously suppressed by even going so far as to prohibit scientists on their payroll from publishing study results. See, e.g., Burton v. R.J. Reynolds Tobacco Co., 167 F.R.D. 134, 142–43 (D. Kan. 1996) (finding evidence sufficient to establish prima facie showing that tobacco-company attorneys were used to facilitate perpetration of continuing fraud to deceive American public about hazards of smoking); Haines v. Liggett Group, Inc., 140 F.R.D. 681, 697 (D.N.J.), order vacated on other grounds, 975 F.2d 81 (3d Cir. 1992) (finding that intentional shielding of evidence that scientific research might tend to prove smoking caused disease was viable theory of fraud entitling plaintiffs to discovery of withheld documents under crime or fraud exception to attorney-client privilege).}

Permitting a limited exception for information about PTSD symptoms provided to the client is one way of discouraging unethical coaching and determining those instances in which the lawyer has gone beyond proper counseling to assist a client in committing fraud. Allowing the fact-finder access to the communications should make clear the nature of the attorney’s conduct. That is, it should be apparent from the content of the communications whether the claim of PTSD was manufactured or was in fact mainly elicited from the client who was already experiencing the symptoms without knowing what they were.

Even if evidence of the information is not permitted at trial, it is still desirable to require disclosure during discovery. Under Federal Rule of Evidence 703, it is well established that an expert (here, a diagnosing psychiatrist or psychologist) can base his or her testimony on evidence that would not be admissible at trial. In the court’s discretion, such bases of expert testimony are discoverable.\footnote{Fed. R. Civ. P. 26(b)(1) provides:}

(b) Discovery Scope and Limits. Unless otherwise limited by order of the court in accordance with these rules, the scope of discovery is as follows:

(1) In General. Parties may obtain discovery regarding any matter, not privileged, that is relevant to the claim or defense of any party, including the existence, description, nature, custody, condition, and location of any books, documents, or other tangible things and the identity and location of persons having knowledge of any discoverable matter. For good cause, the court may order discovery of any matter relevant to the subject matter involved in the action. Relevant information need not be admissible at the trial if the discovery appears reasonably calculated to lead to the discovery of admissible evidence.

Id.
“develop the parameters of its discovery order by carefully weighing the interests involved, balancing the importance of the privilege asserted against the defending party’s need for the information to construct its most effective defense.”122

As noted above, it is essential in some cases for the defendant’s forensic mental-health expert to know whether the claimant received information about PTSD symptoms in order to make an accurate diagnosis. Because the exception would be limited to information provided by the attorney—and not communications the client made in confidence—there would be a very limited impact on the attorney-client relationship.

C. An Exception May Be Found or Established by the Courts or Enacted by the Legislature: Washington, a Case Study

Evidentiary privileges in Washington are established by the legislature and the courts do not have discretion to establish new privileges.123 Exceptions to existing privileges, on the other hand, may be created either by the legislature124 or by the courts.125 In fact, the

122. Greater Newburyport Clamshell Alliance v. Pub. Serv. Co. of N.H., 838 F.2d 13, 20 (1st Cir. 1988). See also United Jersey Bank v. Wolosoff, 483 A.2d 821 (N.J. 1984), where the court held that within the framework of the pretrial discovery process, the attorney-client shield may be pierced when confidential communications are made a material issue by virtue of the allegations in the pleadings and where such information cannot be secured from any less intrusive source. The court stated:

Assuming that all or some of the documents fall within the purview of the attorney-client privilege, it must then be determined whether overriding public policy concerns nevertheless compel disclosure. The “necessary foundations to the valid piercing of [the] privilege” were described by our Supreme Court in In re Kozlov, 79 N.J. 232, 398 A.2d 882 (1979). There, the Court adopted a tripartite test in determining whether the privilege must yield to other important societal concerns. First, “[t]here must be a legitimate need . . . to reach the evidence sought to be shielded.” Id. at 243, 398 A.2d 882. Second, “[t]here must be a showing of relevance and materiality of that evidence to the issue before the court.” Id. at 243–244, 398 A.2d 882. Lastly, the party seeking to bar assertion of the privilege must show “by a fair preponderance of the evidence including all reasonable inferences that the information [cannot] be secured from any less intrusive source.” Id. at 244, 398 A.2d 882.

Wolosoff, 483 A.2d at 826 (alterations in original).


124. See, e.g., WASH. REV. CODE § 5.60.060(4) (2000) (establishing physician-patient privilege, “except as follows: (a) In any judicial proceedings regarding a child’s injury, neglect, or sexual abuse or the cause thereof; and (b) Ninety days after filing an action for personal injuries or wrongful death, the claimant shall be deemed to waive the physician-patient privilege”)

Washington Supreme Court recently stated that “[l]egislative grants of testimonial privilege conflict with the inherent power of the courts to compel the production of relevant evidence and are, therefore, strictly construed.” The Washington appellate courts have established a number of exceptions to the attorney-client privilege when they found the public interest outweighed the benefits the privilege was intended to promote.

The Washington Supreme Court, in determining whether to create an exception to the attorney-client privilege, stated:

The court is faced with the task of balancing society’s interest in the free and open flow of communication between attorney and client, which the privilege promotes, against society’s interest in the administration of justice by our courts on the basis of a full disclosure of the facts and with the affirmative assistance of attorneys, which the privilege discourages.

Similarly, in Pappas v. Holloway, the court held that the attorney-client privilege does not prevent disclosure of confidential communications between the client and the third-party-defendant attorney in a malpractice action where the information sought was relevant to the malpractice issue raised by the client and where the application of the

---

126. C.J.C. v. Corp. of the Catholic Bishop of Yakima, 138 Wash. 2d 699, 717, 985 P.2d 262, 271 (1999) (refusing, with respect to reports from priest’s mental-health counselors sent to Diocese, to “fashion a rule similar to the ‘joint defense’ or ‘common interest’ rule under which communications exchanged between multiple parties engaged in a common defense remain privileged under the attorney-client privilege”).

127. See, e.g., Dike v. Dike, 75 Wash. 2d 1, 11, 448 P.2d 490, 496 (1968). Dike established the exception that an attorney may be compelled, under certain circumstances, to disclose the whereabouts of his client: “The exceptions which have arisen are the result of a balancing process in which the courts have had to weigh the benefits of the privilege against the public interest in the criminal investigation process . . . .” Id. (quoting Formal Opinion 91 (1933), ABA, Opinions on Prof. Ethics 339 (1967)); see also State v. Pawlyk, 115 Wash. 2d 457, 800 P.2d 338 (1990) (finding that attorney-client privilege does not prevent disclosure of confidential communications between client and psychiatrist retained by defense counsel to evaluate defendant for purposes of insanity defense, even if defendant never calls psychiatrist as witness at trial); State v. Chervenell, 99 Wash. 2d 309, 316, 662 P.2d 836, 840 (1983) (finding that attorney-client privilege does not prevent criminal defendant’s former attorney from testifying to advice given regarding defendant’s waiver of his constitutional rights); State v. Sheppard, 52 Wash. App. 707, 763 P.2d 1232 (1998) (noting that fee arrangements between attorney and client not protected from disclosure by attorney-client privilege, unless disclosure would convey substance of confidential attorney-client communication); Whetstone v. Olson, 46 Wash. App. 308, 732 P.2d 159 (1986) (furthest crime or fraud); State v. Metcalf, 14 Wash. App. 232, 540 P.2d 459 (1975) (same).

128. See Dike, 75 Wash. 2d at 14, 448 P.2d at 498.

Confidentiality and Posttraumatic Stress Disorder

privilege would deny the attorney an adequate defense.130 The court applied the *Hearn v. Rhay*131 three-part test to determine whether an applied waiver of the attorney-client privilege should be found.132 The court noted a Louisiana court’s criticism of the *Hearn* test on the basis that it “improperly undermines the legislatively established attorney-client privilege by causing courts to reassess the privilege by weighing the individual privilege-holder’s interests against his opponent’s need for evidence whenever the privilege is attacked.”133 The Washington Supreme Court responded to the criticism of *Hearn* by noting that:

> While it is true that the attorney-client privilege is statutory in nature, it is also true that this court has held that the privilege itself should be strictly limited for the purpose for which it exists. [citing *Dike*]. In this instance, the Holloways cannot counterclaim against Pappas for malpractice and at the same time conceal from him communications which have a direct bearing on this issue simply because the attorney-client privilege protects them. To do so would in effect enable them to use as a sword the protection which the Legislature awarded them as a shield.134

In *State v. Bonds*,135 the court appointed a psychiatrist to assist defense counsel and to evaluate the defendant for purposes of an insanity plea.136 When the defense decided not to call the psychiatrist as a witness at the trial, the prosecution sought the psychiatrist’s testimony.137 The defendant claimed that the psychiatrist’s testimony was protected under the attorney-client privilege because the doctor had been part of the defense team at the juvenile proceedings.138 The court rejected this contention on the basis that the public interest in full disclosure outweighed the privilege. As noted in *State v. Pawlyk*,139 the *Bonds* court

130. *See id.* at 203–04, 787 P.2d at 34.
133. *Id.* at 208, 787 P.2d at 36 (quoting *Succession of Smith v. Kavanaugh*, Pierson & Talley, 513 So. 2d 1138, 1145 (La. 1987)).
134. *Id.*
135. 98 Wash. 2d 1, 653 P.2d 1024 (1982).
136. *Id.* at 18, 653 P.2d at 1034.
137. *Id.*
138. *Id.*
found persuasive the reasoning in Saltzburg, *Privileges and Professionals: Lawyers and Psychiatrists*, 66 Va. L. Rev. 597, 635-42 (1980), where the author “argues that the defense psychiatrist’s examination of the defendant is likely to be more accurate on the issue of insanity than that of the prosecution’s. . . . Moreover, a defendant might benefit by undergoing several psychiatric examinations, examining reports of psychiatrists unfavorable to his insanity defense, and tailoring his responses in subsequent examinations more favorably to his defense. . . .” Saltzburg argues, and we agree, that for these reasons all available evidence of defendant’s mental condition should be put before the jury. 140

These bases are also applicable to materials about PTSD symptoms provided to a claimant prior to evaluation by a defense mental-health professional. That evaluation is likely to be more accurate if the evaluator has access to potential bases for the claimant’s responses, and the claimant might well use the information about symptoms to tailor his responses to the evaluator’s questions more favorably to support the PTSD claim.

IV. CONCLUSION

Increasingly, plaintiffs have claimed emotional and psychological damages on the basis of their having developed PTSD. Essential to an accurate assessment of PTSD claims is the evaluator’s ability to determine all sources of information about related symptoms that may have been available to the plaintiff and that may impact that individual’s clinical presentation. At present, however, the plaintiff’s assertion of attorney-client privilege and work-product protection is likely to prevent discovery of information provided by the plaintiff’s attorney.

Lawyers and courts determine the “rules of the game,” and they can modify them when supported by convincing policy rationales. Thus, for example, the civil discovery rules have been amended on a number of occasions in an attempt to establish the proper balance between competing considerations, including attorney work product and client privilege, encouraging the parties to zealously pursue all possible factual and legal bases for their positions, avoiding fraudulent claims while encouraging efforts to seek changes in the law, and, most important, deciding cases accurately based on all relevant facts.

140. Id. at 462–63, 800 P.2d at 341 (quoting Bonds, 98 Wash. 2d at 21, 653 P.2d at 1035).
Confidentiality and Posttraumatic Stress Disorder

Consideration of similar policy rationales warrants discovery and admission of information about PTSD symptoms provided by attorneys whose clients claim damages on the basis of PTSD. Existing authority, on the basis of waiver or an exception to the privilege, supports allowing access to such information. However, to the extent that existing authority within a particular jurisdiction is deemed not to support such a waiver or exception, a new exception is justified: the very minor (if any) impact on the attorney-client relationship of a limited exception for materials provided by the attorney regarding PTSD diagnosis and symptoms is substantially outweighed by the need of the mental-health evaluator to have access to that information. As the Washington Supreme Court stated with respect to a criminal defendant’s asserted right to prevent access to communications with a psychiatrist retained by defense counsel to evaluate the defendant for purposes of an insanity defense:

   defendant’s asserted right . . . reflects the “bygone philosophy that for an attorney’s investigations to be effective they must be shrouded in secrecy.” If defendant asserts an insanity defense, evidence pertaining to that defense must be available to both sides at trial. There is thus no need for the confidentiality defendant maintains is required.141

Limited access to information provided to a PTSD claimant by his or her attorney is essential to a fair and accurate mental-health evaluation. In addition, knowledge that such information will be discoverable will discourage improper coaching by overzealous counsel.

141. State v. Pawlyk, 115 Wash. 2d at 470, 800 P.2d at 345 (quoting State v. Craney, 347 N.W.2d 668, 677 (Iowa 1984)).
CONVERSION AND MERGER OF DISPARATE BUSINESS ENTITIES

Robert C. Art*

Abstract: Legislation permitting a business organized in one form, such as a corporation, to merge with a business of a different form, such as a limited liability company, is relatively recent, but reasonable and beneficial. A logical extension of this legislation is to permit a single business entity to convert its organizational form without involving a second entity. Recognition of these cross-entity transactions flows naturally from the expansion of organizational options in recent years, particularly the introduction of limited liability companies and limited liability partnerships. Conversion and merger of disparate entities are already available in a few states, with varying degrees of liberality, and are likely to become increasingly common. Both the Revised Uniform Partnership Act and the Model Business Corporation Act support the key principles behind conversion and merger. This Article examines the policies, principles, and drafting issues of cross-entity conversion and merger legislation. It focuses on Oregon’s comprehensive statutory scheme and urges other states to emulate that approach.

I. INTRODUCTION ..........................................................................350

II. THE EXPANSION OF BUSINESS ORGANIZATION OPTIONS .................................................................353

A. The Traditional Entities: Corporation, Cooperative, Partnership, and Limited Partnership .........................................................355

B. The Traditional Dilemma: Limited Liability or Pass-Through Tax Treatment .................................................................359

C. New Options: LLC, LLP, and LLLP ..........................................................360

D. Current Choices and Prospects for the Future .....................................365

III. CONVERSION AND MERGER UNDER TRADITIONAL LAW .................................................................368

A. Transactions and Problems in the Absence of Statutes ..........................................................368

B. Benefits of Conversion and Merger Statutes ..........................................................372

*Professor of Law, Willamette University; J.D., DePaul, 1978; LL.M., Columbia, 1981. The author was a member of the Task Force on Multi-Entity Conversion and Merger of the Business Law Section, Oregon State Bar, and served as its principal drafter. Professor Art has also served on the Business Law Section’s task forces and committees that drafted Oregon’s corporation act in 1989, its subsequent amendments, and its partnership act of 1997. He has also served on the Section’s legislative committee.
C. Limited Availability of Conversion and Merger:  
   Oregon as an Example ............................................................373
D. The Issue of Personal Liability Entities Changing to  
   Limited Liability Entities .........................................................377
E. A Situation Ripe for Reform .....................................................378

IV. THE TREND TOWARD CONVERSION AND  
MERGER OF DISPARATE ENTITIES ..................................379
   A. Use of Foreign Law To Accomplish What Domestic  
   Law Fails To Provide ..............................................................380
   B. The “Junction Box” Model ...............................................381
   C. The Self-Contained Model ......................................................382

V. OREGON’S BUSINESS ENTITY CONVERSION AND  
MERGER LEGISLATION ...........................................................383
   A. Legislative History ..........................................................383
   B. The Central Premise: Continuity of Entity Existence ..........385
   C. Principles and Goals ...............................................................386
   D. Types of Business Entities Included .....................................388
   E. Procedures for Approving the Transaction .........................391
   F. Creditors’ Rights and Owners’ Liabilities .........................393
   G. Owners’ Rights ........................................................................399
   H. Property ...................................................................................402
   I. Documentation and Public Filings...........................................405
   J. Income Tax Consequences .....................................................407
   K. Early Experience with Conversion  
      and Merger of Disparate Business Entities .........................408

VI. CONCLUSION ..............................................................................409
APPENDIX ...............................................................................................410

I. INTRODUCTION

Why should merger of two corporations be authorized and facilitated by statute, while merger of a corporation with a limited liability company is cumbersome, complex, and expensive? Why should a general partnership be allowed to readily convert to a limited liability partnership, but not to a limited liability company? States increasingly are recognizing that only historical happenstance, not reasoned policy, accounts for such anomalies, and are updating their statutes to enhance
the opportunities for business entities and their counsel to alter their organizational form.\(^1\)

This salutary development follows logically from the recent introduction of new forms of business organization, especially the limited liability company (LLC) and limited liability partnership (LLP), and from concepts incorporated in the Revised Uniform Partnership Act (RUPA).\(^2\) Modern statutes both extend the range of permissible mergers to disparate entities (such as a corporation merging with a partnership), and authorize conversions from one form to another (a transformation not involving a second business).\(^3\) These changes are variously known as cross-entity,\(^4\) cross-species,\(^5\) multi-entity,\(^6\) or inter-entity\(^7\) conversions and mergers. Oregon legislation effective in 2000 provides a particularly good model for other states to follow.\(^8\)

Combining two business entities of unlike form has always been possible, but has been cumbersome, complex, and expensive, without policy justification. A corporation seeking to combine with an existing limited liability company, for example, could not merge because no

---

1. See, e.g., Robert R. Keatinge, Inter-Entity Mergers and Conversions and Other Changes Under SB 97-233, COLO. LAW., Jan. 1998, at 43. Colorado, Delaware, Nevada, and Texas are among the states at the forefront of the trend toward authorization of mergers of disparate business entities and conversion of one form to another. Citations and further detail are provided infra notes 177–84.

2. REV. UNIF. PARTNERSHIP ACT (1994).

3. See, e.g., infra notes 177, 181, 195, 199–200 (citing statutes from Colorado, Delaware, Nevada, and Texas). Also, the Model Business Corporation Act was amended in 1999 to authorize mergers of corporations with “other entities” but the reform did not extend to conversions. MODEL BUS. CORP. ACT ANN. §§ 11.01–.02 (Supp. 1999); The Committee on Corporate Laws, Changes in the Model Business Corporation Act—Fundamental Changes, 54 BUS. LAW. 685, 695–97 (1999) (indicating and explaining new cross-entity merger provisions, but making no provision for cross-entity conversion).

4. An article on Colorado’s “inter-entity” mergers, for example, also refers to them as “cross-entity” mergers. Keatinge, supra note 1, at 44.


6. For example, the Oregon State Bar committee that drafted Oregon’s current legislation was known as the Task Force on Multi-Entity Conversion and Merger. The “multi” reference, however, emphasizes the number of entities (such as three corporations) that can merge, rather than the type (such as a corporation and a limited liability company (LLC)), which is the more significant innovation of the legislation.

7. See Keatinge, supra note 1 (as reflected in the article’s title).

8. 1999 Or. Laws 362. The principal provisions of the legislation are codified in thirty-eight sections of OR. REV. STAT. chapters 60 (corporation), 62 (cooperative), 63 (LLC), 67 (partnership), and 70 (limited partnership) (1999). The Appendix to this Article provides a chart of citations by subject.
statute authorized it, but could accomplish the same pragmatic results. As one possibility, the corporation could sell its assets to the LLC in exchange for LLC membership interests, then dissolve and issue the membership interests to its shareholders as a liquidating distribution. As a result, the two businesses become one, the corporation disappears, and the former shareholders become members of the surviving LLC—all of the characteristics of a cross-entity merger. 9 No apparent policy supported the law’s failure to provide a more efficient, direct means.

Once cross-entity merger is accepted, authorization of cross-entity conversion is the next logical step. Otherwise, existing businesses are deprived of the flexibility and opportunities available to newly organized enterprises, and relegated to the traditional, burdensome means of altering their organizational form. A partnership or limited partnership wishing to become an LLC, for example, could dissolve and distribute its net assets (in cash or in kind) to partners, who could then organize a new LLC and contribute to it the former partnership assets in exchange for LLC membership interests. 10 Alternately, in a jurisdiction allowing cross-entity merger but not cross-entity conversion, the goal can be achieved in two steps. First, the partnership would organize an LLC, briefly holding all of the membership interests. Then the partnership would merge into the LLC, disappearing but leaving its former partners with membership interests in the surviving LLC. 11 No reason appears for the law not to authorize the direct, single-step option of conversion, involving only one entity and one document, to accomplish the same result.

Authorizing conversion and merger of disparate entities raises numerous issues, involving each of the constituencies within or affected by a business. Most importantly, rights of creditors and litigants must be protected, so that conversion and merger cannot be used to evade existing obligations. Transfer of business property must be perfectly clear. The rights of owners must be preserved in the procedures to

9. This approach and others are described in Mark D. Lubin, Selected Aspects of the California Limited Liability Company Act, in THE BEST ENTITY FOR DOING THE DEAL, 937 PLI/Corp. 343, 361–64 (1996); Lion & Chacon, supra note 5, at 216–22.

10. David C. Culpepper, Limited Liability Companies, in ADVISING OREGON BUSINESSES ch. 39A (1989 & Supp. 1997); Lion & Chacon, supra note 5, at 169, 218. A second method is for the partners to contribute their partnership interests to the LLC in exchange for LLC membership interests. The partnership then liquidates, distributing its assets to the LLC. A third alternative is for the partnership to exchange its assets for LLC membership interests and then liquidate, distributing the membership interests to the former partners. Id. at 216–21.

11. The proposed transaction is a variation of one suggestion by Lion & Chacon, supra note 5, at 213 (“Method 3”), 218 (chart IV), but substituting conversion for organizing a new entity.
authorize the transactions and in the consequences of the transaction. Administrative procedures must allow the state to record and report on the current status of a business, without undue expense or complexity. Businesses that are organized in other states entail further issues of coordination, reciprocity, and choice of laws.

Several states have wrestled with these issues, producing alternate means of resolution. Oregon has a particularly direct, comprehensive, and flexible set of statutes, effective in 2000, which other states should emulate. The statutes authorize conversion or merger among any of the six business entity forms: corporation, professional corporation, cooperative, partnership (including general partnership and limited liability partnership), limited partnership, and limited liability company. Notably, the legislation allows conversion or merger of forms providing limited liability to owners (such as the corporation and LLC) into forms imposing personal liability on owners (such as the general partnership)—a policy and drafting decision that required particularly careful consideration.

This Article will outline the plethora of business organization alternatives presently available and the means by which business entities have traditionally been allowed to combine or to change organizational structure. It will describe the progress and alternative approaches used by states that have been leaders in enacting statutory provisions authorizing and facilitating mergers of disparate business entities and conversion of a business into a different organizational form. Finally, this Article will focus on Oregon’s contribution to this field of law, a model that other states should follow.

II. THE EXPANSION OF BUSINESS ORGANIZATION OPTIONS

The choices of business organization form for businesses has expanded over the past several years. Traditionally, Oregon, like other states, offered the options (in addition to sole proprietorship) of

12. For example, Colorado, Delaware, Nevada, and Texas allow cross-entity conversion and merger. See infra notes 177–82, 199–200, and accompanying text.
13. OR. REV. STAT. §§ 60.470–.497 (corporation), 60.605–.623 (cooperative), 63.467–.497 (LLC), 67.340–.365 (partnership), 70.500–.540 (limited partnership) (1999).
corporation\textsuperscript{15} (and the variations of professional corporation\textsuperscript{16} and cooperative),\textsuperscript{17} general partnership,\textsuperscript{18} and limited partnership.\textsuperscript{19} In general, the forms providing a shield from personal liability for business debts also entailed taxation at both the company and the owner levels.\textsuperscript{20} More recent additions are the limited liability company\textsuperscript{21} and limited liability partnership,\textsuperscript{22} providing benefits both in terms of protection from liability and avoidance of company-level taxation.

A thorough presentation of the characteristics of each organizational form is complex, including myriad differences in individual states,\textsuperscript{23} and is left to sources other than this Article.\textsuperscript{24} However, mention of a few of

\begin{footnotesize}
\begin{enumerate}
\item[15.] Oregon Business Corporation Act, 1987 Or. Laws 52 (codified at OR. REV. STAT. ch. 60 (1999)).
\item[16.] Oregon Professional Corporation Act, 1969 Or. Laws 592 (codified at OR. REV. STAT. ch. 58 (1999)).
\item[17.] The Oregon Cooperative Corporation Act, 1957 Or. Laws 716, is codified at OR. REV. STAT. ch. 62 (1999).
\item[19.] Oregon’s UNIFORM LIMITED PARTNERSHIP ACT, 1985 Or. Laws 677, codified at OR. REV. STAT. ch. 70 (1999), is based on the UNIFORM LIMITED PARTNERSHIP ACT. See generally James M. Kennedy & Turid L. Owren, Limited Partnerships, in ADVISING OREGON BUSINESSES ch. 6 (1989).
\item[20.] See infra notes 54–60 and accompanying text.
\item[21.] The Oregon Limited Liability Company Act, 1993 Or. Laws 173, is codified at OR. REV. STAT. ch. 63. See generally Culpepper, supra note 10.
\item[22.] Provisions governing limited liability partnerships (LLPs) are not in a separate chapter, as is each of the other forms of business entity. Instead, in 1997, the legislature appended a number of provisions applicable to LLPs, OR. REV. STAT. §§ 67.500–770, to the general partnership act, OR. REV. STAT. ch. 67. 1997 Or. Laws 775.
\item[23.] A few examples will suggest the complexity of the comparison. An LLP has some tax advantages over an LLC in terms of wealth-transfer tax valuation and self-employment tax. In Florida, the corporate income tax applies to an LLC but not a limited partnership. Oesterle & Gazur, supra note 14, at 119 n.101. In Texas, the corporate franchise tax applies to an LLC but not a limited partnership. Id.
\end{enumerate}
\end{footnotesize}
the most salient features will indicate some of the major reasons that businesses organized in one form may wish to convert or merge into another form.

A. The Traditional Entities: Corporation, Cooperative, Partnership, and Limited Partnership

Prior to a key change in federal income tax regulations in 1988, entrepreneurs organizing a business more complex than a sole proprietorship generally chose from a short list of well-established forms. To obtain protection from personal liability, they could organize a corporation, the professional corporation variant if they provided professional services, or a cooperative. Partnership, which exposes owners to full individual liability for business debts, was the default choice. Limited partnership, protecting some but not all owners, completed the traditional list of organizational options.

The corporate form was originally designed primarily for large enterprises, separating ownership from control, but was extended to small, closely held businesses and professionals. Corporations have centralized management, with a statutorily prescribed structure of directors elected by shareholders and officers appointed by directors. Shareholders are deprived of agency power to bind the corporation, but are also protected from liability to creditors except in the rare cases of “piercing the corporate veil.” Directors are subjected to the fiduciary duties of care and loyalty to shareholders, which cannot be waived, although most states permit exculpation from monetary liability for

---


25. In 1988, the Internal Revenue Service released the “check-the-box” regulations allowing non-corporate entities, such as limited liability companies, to select pass-through tax treatment. Rev. Rul. 88-76, 1988-2 C.B. 360. Until this key decision, the LLC had been an isolated form in Wyoming; after 1988 it rapidly spread throughout the country.


27. OR. REV. STAT. § 60.151; REV. MODEL BUS. CORP. ACT § 6.22.

breach of the duty of care. 29 Those factors, plus the familiarity of the corporate form to investors, make corporations the typical business entity form for going public.

The professional corporation (PC), is simply a corporation organized for the purpose of practicing law, medicine, accounting, or a number of other professional services specified in the governing statute. 30 A few specialized provisions cover such matters as liabilities of employees, shareholders, and the corporation, and the relationship of the PC to professional licensing agencies. 31 However, in all other matters, including conversion and merger, PCs are governed by the Business Corporation Act. 32

Cooperative corporations (usually called cooperatives) are a distinctly different form, governed by a separate statute. 34 Cooperatives have “members” and may or may not have shares and shareholders. 35 Members commonly purchase supplies or market products through the cooperative, and receive net proceeds or other benefits in proportion to the volume of their purchases or marketing, rather than in proportion to their capital contributions. 36 Governance is typically based on the principle of one vote per member, rather than one vote per share. 37


32. Id. §§ 58.325–389.

33. Id. § 58.045.


35. Or. Rev. Stat. § 62.513(c). Members own shares, or pay membership fees, or meet other qualifications as set forth in the articles and bylaws. Id. § 62.175(2). Both members and shareholders are shielded from liability. Id. § 62.215.

36. Id. § 62.415. For example, a farmer who contributed 17% of a cooperative’s capital, but 5% of the grain stored in the cooperative’s silo, will receive 5% of the proceeds of the silo operation.

37. Id. § 62.265(1).

38. Id. § 62.265(1) (“Shares of stock as such shall not be given voting power except in the specific instances authorized by this chapter.”).
Cross-Entity Conversion and Merger

cooperaive form can be used for nearly any purpose but seems most common in agriculture, consumer organizations (such as buying clubs), and multi-family housing.

Partnership (sometimes called general partnership to distinguish it from limited partnership and limited liability partnership) stands in sharp contrast to the corporate form. Unless the partnership agreement provides otherwise, every owner participates equally in management. Moreover, every partner has apparent agency authority to bind the partnership, even if the partnership agreement denies actual agency authority—an aspect that makes the form unsuitable when centralized management is desired. Partners are exposed to joint and several liability to creditors, though creditors typically must first exhaust their remedies against the business.

Partners are subject to fiduciary duties that cannot be waived but can be defined in the partnership agreement. General partnership is the only

39. Cooperatives can engage in any lawful business except banking or insurance. Id. § 62.115. For a discussion of common uses, see Simon, supra note 34, at 4.

40. For example, a group of farmers might operate a grain storage and loading facility, or a commodity distribution and marketing system, in cooperative form. As one indication of the agricultural connection, the Oregon Cooperative Corporation Act states a public policy in favor of "the efficient production and distribution of agricultural and other products derived from natural resources or labor resources," and provides protection from antitrust liability. OR. REV. STAT. §§ 62.005, .845. For discussion of development of the cooperative form in Pennsylvania, see William H. Clark, Jr., What the Business World is Looking for in an Organization Form: The Pennsylvania Experience, 32 WAKE FOREST L. REV. 149, 158–61 (1997).

41. OR. REV. STAT. §§ 67.140(7), 68.310(5) (1999); UNIF. PARTNERSHIP ACT § 18(e) (1914); REV. UNIF. PARTNERSHIP ACT § 401(f) (1994).

42. OR. REV. STAT. §§ 67.090(1), 68.210(1); UNIF. PARTNERSHIP ACT 9(1); REV. UNIF. PARTNERSHIP ACT § 301(1) (noting that every partner can bind partnership, unless partner in fact has no authority and that fact is known to outside party). See generally ALAN R. BROMBERG & LARRY E. RIBSTEIN, BROMBERG AND RIBSTEIN ON PARTNERSHIP § 4.01(b) (2000 & Supp. 2001) [hereinafter BROMBERG PARTNERSHIP].

43. OR. REV. STAT. §§ 67.105, 68.270; REV. UNIF. PARTNERSHIP ACT § 306; UNIF. PARTNERSHIP ACT § 15.

44. The Uniform Partnership Act labeled partners “fiduciaries” of co-partners but did not define the fiduciary duties, except as to certain narrow issues, such as the duty to render information, OR. REV. STAT. § 68.340; UNIF. PARTNERSHIP ACT § 21, leaving the development of fiduciary duties to courts under the common law. The revised act sought to define and limit the duties. RUPA section 404 provides that “the only fiduciary duties are the duties of loyalty and care.” REV. UNIF. PARTNERSHIP ACT § 404; see also OR. REV. STAT. § 67.155(1). For discussion of fiduciary duties in partnerships and citations to the very extensive debate between proponents of strict fiduciary duties and proponents of freedom of contract, see J. Dennis Hynes, Freedom of Contract, Fiduciary Duties, and Partnerships: The Bargain Principle and the Law of Agency, 54 WASH. & LEE L. REV. 439 (1997). States adopting RUPA have modified the fiduciary duty provision with multiple variations, seriously detracting from uniformity in this field of law. Allan W. Vestal, "Assume a Rather Large
form of business entity (other than sole proprietorship) that can be formed without any explicit agreement, filing with the state, or other formality. These factors contribute to its popularity, especially among smaller and less sophisticated businesses. Partnership interests can be assigned but the transferee does not become a partner unless the others agree. Death or withdrawal of any partner can dissolve the partnership, though the remaining partners can continue the business if they compensate the dissociated partner or that person’s estate.

Limited partnership, typically governed by a separate statute, divides partners into two distinct classes—general and limited partners—producing a structure combining elements of the other traditionally recognized business forms. General partners have exclusive authority to manage and owe fiduciary duties somewhat like corporate directors. However, general partners are exposed to joint and several liability to creditors, as in a partnership. The limited partners select the general

45. UNIF. PARTNERSHIP ACT section 6, REV. UNIF. PARTNERSHIP ACT section 202, and OREGON REVISED STATUTES section 67.055 provide that a partnership is created by an association of two or more persons to carry on as co-owners a business for profit, imposing no requirement for a filing or even a writing.

46. BROMBERG PARTNERSHIP, supra note 42, § 1.01(b).

47. OR. REV. STAT. §§ 67.200, 68.440 (stating that transferee receives only profits to which partner would otherwise be entitled); REV. UNIF. PARTNERSHIP ACT § 503; UNIF. PARTNERSHIP ACT § 27.

48. OR. REV. STAT. § 68.530; UNIF. PARTNERSHIP ACT § 31. In the RUPA, death or withdrawal is deemed a dissociation that leads to a buyout of the partner’s interest rather than a windup. OR. REV. STAT. § 67.220; REV. UNIF. PARTNERSHIP ACT § 601.

49. OR. REV. STAT. §§ 67.250, 68.600; REV. UNIF. PARTNERSHIP ACT § 701; UNIF. PARTNERSHIP ACT § 38.

50. The REVISED UNIFORM LIMITED PARTNERSHIP ACT (RULPA), was approved by the National Conference of Commissioners on Uniform State Laws in 1976, and substantially amended in 1985.

51. REV. UNIF. LTD. PARTNERSHIP ACT § 403.

52. Id. § 9(1), 6A U.L.A. 346 (1995) (“A general partner shall have all the rights and powers and be subject to all the restrictions and liabilities of a partner in a partnership without limited partners . . . .”). In some states, however, legislation protects the general partner in a limited partnership from liability, Oesterle & Gazur, supra note 14, at 105 (citing COLO. REV. STAT ANN. §§ 7-62-101(12), -403(2) (West 1999); DEL. CODE ANN. tit. 6, § 17-214 (Supp. 2000); MICH. STAT. ANN. §§ 20.46, 10.1403(b) (Michie 1960 & Supp. 1996); TEX. REV. CIV. STAT. ANN. art. 6132a-1, § 2.14(a) (Vernon Supp. 1995)).
partners but otherwise have little or no agency power or managerial authority, except as specifically granted in the partnership agreement, and are not subject to personal liability. In those respects, limited partners resemble corporate shareholders.

B. The Traditional Dilemma: Limited Liability or Pass-Through Tax Treatment

In selecting among the business organizational forms that were available prior to 1988, entrepreneurs and their counsel generally were required to balance the importance of limited liability with the value of pass-through tax treatment, selecting one and forfeiting the other. The Internal Revenue Code classifies businesses either as “corporations” or “partnerships.” The term “corporation” includes “association,” covering all businesses that resemble corporations, even if not formally incorporated under state law. The term “partnership,” for tax purposes, includes all multi-owner businesses that are not corporations, trusts, or estates.

Corporations, professional corporations, and cooperatives provided limited liability for all equity owners, but were subject to “double taxation”; business income was taxed at the entity level, and all profits distributed to owners were taxed at the individual level. Alternatively,

One means of avoiding personal liability to any individual, while still providing pass-through tax treatment for the limited partners, is to designate a corporation as a general partner. BROMBERG PARTNERSHIP, supra note 42, § 12.11(b).

53. Although limited partners generally are not at risk of personal liability, they can be exposed if they participate in control of the enterprise. Amendments to the RULPA in 1985 provided safe harbors against liability for participation, including one for “matters related to the business . . . which the partnership agreement states in writing may be subject to the approval or disapproval of limited partners.” REV. UNIF. LTD. PARTNERSHIP ACT § 303(b)(6)(ix), 6A U.L.A. 145 (1995). Liability extends only to persons who “reasonably [believe], based upon the limited partner’s conduct, that the limited partner is a general partner.” Id. § 303(a), 6A U.L.A. 144. At least one state has eliminated the potential for liability of limited partners for participating in control of the business, and others may follow that lead. GA. CODE ANN. § 14-9-303 (1994), cited in Clark, supra note 40, at 176 (stating that similar legislation has been proposed in Pennsylvania).


55. Id. § 7701(a)(2).


57. I.R.C. § 7701(a)(2).

58. Id. §§ 11, 301, 302; BITTKER & EUSTICE, supra note 56, ¶ 1.03.
a corporation could elect Subchapter S treatment\textsuperscript{59} to avoid the entity-level tax, but only subject to restrictions and limitations.\textsuperscript{60}

Partnerships and limited partnerships received pass-through tax treatment, avoiding taxation at the entity level, but at the price of exposure of at least some owners to personal liability for business debts. All of the partners in a general partnership and the general partners in a limited partnership are personally liable.\textsuperscript{61}

Thus, none of the traditional forms offered the ideal combination of limited liability and pass-through tax treatment. Businesses sought new options and state legislatures responded.

C. New Options: LLC, LLP, and LLLP

During the late 1980s and the 1990s, new business-organization alternatives became available: the limited liability company, limited liability partnership, and a variation, the limited liability limited partnership. State legislatures intended to provide businesses as many corporate characteristics as possible (especially limited liability for investors) coupled with the pass-through treatment under federal tax regulations available to partnerships,\textsuperscript{62} enabling businesses to avoid the entity-level tax applicable to corporations.\textsuperscript{63} The new forms also offer

\textsuperscript{59} I.R.C. §§ 1361–1379.

\textsuperscript{60} S corporations, for example, cannot have more than seventy-five shareholders, more than one class of stock, or nonresident alien shareholders. 26 U.S.C. § 1361(b) (1994). For discussions of other factors, see Wayne M. Gazur & Neil M. Goff, \textit{Assessing the Limited Liability Company}, 41 CASE W. RES. L. REV. 387, 459 (1991); John W. Nickel et al., \textit{Subchapter S Taxation, in ADVISING OREGON BUSINESSES} ch. 11 (1989 & Supp. 1997); and Oesterle & Gazur, \textit{supra} note 14, at 121 n.109.

\textsuperscript{61} See \textit{supra} notes 43, 50, and accompanying text.

\textsuperscript{62} Internal Revenue Service regulations, known as the \textit{Kintner} regulations, determined whether a business entity was taxable as a corporation (or “association”) or instead as a partnership. Four characteristics were relevant: centralized management, free transferability of interests, continuity of existence, and limited liability. An entity with three or more of those characteristics was taxed as a corporation, while one with two or fewer was a partnership. 26 C.F.R. 301.7701-1, -2(a) to -2(e) (repealed in 1996). The regulations were promulgated in response to \textit{United States v. Kintner}, 216 F.2d 418 (9th Cir. 1954).

\textsuperscript{63} Many commentators confirm that statutes creating the LLC and LLP were specifically intended to provide businesses the federal income tax advantages of partnerships coupled with the advantages, especially limited liability, of the corporate form. \textit{See}, e.g., Susan Saab Fortney, \textit{Seeking Shelter in the Minefield of Unintended Consequences—The Traps of Limited Liability Law Firms}, 54 WASH. & LEE L. REV. 717, 722–24 (1997); Carol R. Goforth, \textit{The Rise of the Limited Liability Company: Evidence of a Race Between the States, But Heading Where?}, 45 SYRACUSE L.
more flexibility in selecting the form of internal management.\textsuperscript{64} Once the new forms became available in some states, the others followed quickly, none wanting to be left behind.\textsuperscript{65}

The limited liability company originated in Wyoming in 1977,\textsuperscript{66} and remained an isolated anomaly for several years.\textsuperscript{67} The pivotal event providing the impetus for extensive state legislation was the Treasury Department’s revenue ruling in 1988 allowing an LLC pass-through tax treatment.\textsuperscript{68} The company’s income was allocated directly to its owners, labeled “members,” avoiding the double taxation applicable to most corporations. After the revenue ruling the LLC spread swiftly throughout the nation and by 1996 had become available in all states.\textsuperscript{69}

In addition to the favorable tax treatment, the LLC form shields all of its members and managers from liability for obligations of the company\textsuperscript{70} and provides a choice of management structure. The default provision is member management, whereby all members participate in management and have agency power to bind the entity,\textsuperscript{71} much like

\begin{thebibliography}{99}
\item REV. 1193, 1272 (1995). Pennsylvania was an especially clear example, stating that motivation explicitly in the statute. Clark, \textit{supra} note 40, at 162.
\item 64. The corporate system delegates centralized control to the board of directors, with investors deprived of participation or agency power except for power to elect directors. That arrangement is better suited to publicly traded corporations than to closely held corporations. See \textit{generally} DOUGLAS A. BRANSON, CORPORATE GOVERNANCE § 4.12 (1993) (noting that owners of smaller corporations serving as directors find traditional forms of governance inappropriate). Of course, shareholder agreements provide a means of modifying corporate governance norms to meet the preferences within a particular corporation, but entail certain costs for negotiation and preparation.
\item 65. Fortney, \textit{supra} note 63, at 722; Goforth, \textit{supra} note 63, at 1272.
\item 66. WYO. STAT. ANN. §§ 17-15-102 to -144 (Michie 1999) (effective June 30, 1977); Gazur & Goff, \textit{supra} note 60, at 389. The statute was enacted as special legislation to assist a mineral company, providing limited liability to all owners and pass-through tax treatment. \textit{Id.} The IRS issued a private letter ruling classifying the company as a partnership for tax purposes. Priv. Ltr. Rul. 81-06-082 (Nov. 18, 1980); Fortney, \textit{supra} note 63, at 722 (citing Goforth, \textit{supra} note 63, at 1199–200).
\item 67. For eleven years, only Florida had followed Wyoming’s lead. Fortney, \textit{supra} note 63, at 722 n.21; Oesterle & Gazur, \textit{supra} note 14, at 105 n.23.
\item 70. OR. REV. STAT. § 63.165 (1999).
\item 71. \textit{Id.} § 63.130.
\end{thebibliography}
partners in a general partnership. Alternatively, an LLC may provide in its articles of organization for management by or under the authority of one or more managers who need not be members, thereby creating a structure comparable to a corporation or a limited partnership. Fiduciary duties likely apply, but the operating agreement may eliminate or limit the liability of a member or manager to the company or its members. Members have the same protection from liability to outsiders as shareholders in corporations. The combination of attributes has made the LLC remarkably popular.

The most recent major addition to the list of organizational options is the limited liability partnership, or LLP (sometimes labeled a registered limited liability partnership). The form originated in Texas in 1991 and was designed primarily to protect attorneys and other professionals from vicarious liability for the negligence of their partners in the wake of the

---

72. A member-managed LLC, like a general partnership, may designate one or more persons to be responsible for routine management, but this designation without more will not deprive the other members of apparent agency authority to bind the company to outsiders. To do that requires a provision in the articles of organization specifying management by managers. OR. REV. STAT. § 63.140(2); Culpepper, supra note 10, § 39A.19.

73. OR. REV. STAT. § 63.135.

74. The existence and extent of fiduciary duties within an LLC varies from state to state and is not yet fully defined. In Delaware, for example, the statute does not explicitly create fiduciary duties for LLC members or managers, but it is likely that they exist (by analogy to limited partnerships and corporations) when one person controls the entity’s property for the benefit of other owners. Leyden, supra note 69, at 57. In Oregon, a statutory provision imposes duties of good faith and care on managers but is silent on duties of non-manager members. OR. REV. STAT. § 63.155.

75. OR. REV. STAT. § 63.160 (excluding breaches of manager’s duty of loyalty, intentional misconduct, knowing violation of law, unlawful distribution, or improper personal benefit).

76. The extent of liability of LLC members to outsiders has yet to be fully litigated, but one expert’s evaluation is that liability protection will be fairly uniform among the states despite differences in statutory language, and will be similar to shareholders in corporations. See Thompson, supra note 69, at 7. A couple of similarities are that, even under an LLC act that carefully protects members from nearly all liabilities, members (like shareholders) can be compelled to fulfill their undertakings to make contributions to the entity, and may be required to return improper distributions received from the entity. Leyden, supra note 69, at 57 (citing Delaware’s LLC act, DEL. CODE ANN. tit. 6, § 18-101 (effective 1992)); see also CALLISON & SULLIVAN, supra note 24, § 8.7.

77. As one indication, in Delaware, more than 100,000 LLCs were formed between 1992 and 2000, compared to approximately 290,000 corporations organized over a dramatically longer period. Leyden, supra note 69, at 51.

collapse of many Texas financial institutions. The form spread rapidly and was adopted by 1999 in all fifty states and the District of Columbia.

LLPs are generally governed not by a separate statute (as is the case of each of the other types of business entity) but by the general partnership act, with a few specialized provisions regarding public filings and owners’ liability for business obligations. For that reason, and because the LLP is within the general partnership act’s definition of partnership, it actually is not a separate form of entity.

Part of the appeal of the LLP was that the partnership form was well-established while the LLC innovation was not well known, and the “partner” title was far more appealing to some than the “member” designation. Another advantage was the ease of conversion, in that existing partnerships could gain the limited liability shield without having to change their form to an LLC. Further, the partnership form...

79. The initial bill, which covered only professionals, appeared to be a “help-a-lawyer-bill,” and was heavily criticized by partnership expert Professor Alan Bromberg. See Steven A. Waters & Matthew D. Goetz, Annual Survey of Texas Law—Partnerships, 45 SW. L.J. 2011, 2022 (1992), cited in Fortney, supra note 63, at 725 n.35. Revisions removed the limitation to professionals, denied protection to partners for misconduct of those under their supervision, required disclosure of LLP status in the firm name, and required liability insurance. Id. For descriptions of the origins of the LLP, see Alan R. Bromberg & Larry E. Ribstein, Bromberg and Ribstein on Limited Liability Partnerships and the Revised Uniform Partnership Act § 1.01(a), at 3 (2000 ed.) [hereinafter BROMBERG LLP] and Robert W. Hamilton, Registered Limited Liability Partnerships: Present at the Birth (Nearly), 66 U. COLO. L. REV. 1065, 1065 (1995).

80. Fallany O. Stover & Susan Pace Hamill, The LLC versus LLP Conundrum: Advice for Businesses Contemplating the Choice, 50 ALA. L. REV. 813, 816 (1999) (providing complete list of citations to LLC statutes); see also BROMBERG LLP, supra note 79, § 1.01(e).

81. See, e.g., OR. REV. STAT. §§ 67.500–.770 (1999). Most states follow the same pattern of including LLP within the general partnership act. Exceptions exist, such as Pennsylvania, but even there LLPs are governed generally by the general partnership law. Clark, supra note 40, at 162 n.60.

82. Fortney, supra note 63, at 726 (noting that LLP owners can continue tradition of holding themselves out as partners); Keatinge, supra note 24, at 80 (recognizing that some entities select LLP or even convert LLC to LLP in order to label owners as “partners”); Oesterle & Gazur, supra note 14, at 105 (noting that, among other things, forms for partnership agreements were more available and trusted than forms for LLC operating agreements); id. at 113 (finding that lawyers prefer to be “partners” rather than “members” of law firms).

83. A general partnership wishing to become an LLP need only file its application for registration with the Secretary of State and amend its partnership agreement to add LLP to the name. See, e.g., OR. REV. STAT. §§ 67.590, .625. A partnership that registers as an LLP continues to be the same entity. OR. REV. STAT. § 67.050(2). There is no need to transfer assets and liabilities to a new entity as might be necessary if the entity changed to an LLC. Clark, supra note 40, at 164; Fortney, supra note 63, at 725–26.
was recognized in all states, which for a period was not true for LLCs—a matter of particular importance to nationwide accounting firms.84

When a partnership registers as an LLP, its partners gain protection from liability for obligations of the partnership, although they always remain liable for their own negligence and for the negligence of those they supervise.85 Many states, including Oregon, protect against both contract and tort liabilities, while some others restrict the liability shield to contract debts.86 Some statutes condition limited liability on maintenance of a specified level of insurance as a means of providing some protections for creditors.87 Oregon protects LLP partners who are professionals against liabilities only above a specified dollar amount.88 This limited liability has some negative consequences not always recognized by the participants but is nevertheless highly attractive.89

The characteristics of an LLP are highly similar, though not identical, to those of an LLC,90 leading some to suggest that the LLC would not

86. OR. REV. STAT. § 67.105(3)(a) (stating that obligation of LLP, “whether arising in contract, tort or otherwise, is solely the obligation of the partnership” rather than its partners). Some LLP statutes eliminate vicarious liability for malpractice by co-partners in professional service firms, such as accounting and law firms, while others provide a shield for both contract and tort liabilities. The trend appears to be to protect LLP partners for all liabilities. Larry E. Ribstein, Possible Futures for Unincorporated Firms, 64 U. CIN. L. REV. 319, 321–22 (1996). For an extensive discussion, see Thompson, supra note 69, at 22–24; Hamilton, supra note 79, at 1066–68, 1087–90 (comparing Texas’s original LLP statute with narrow protection for owners against business liabilities with Minnesota and New York statutes offering more liberal protection).
88. The liability provision in the LLP act, OR. REV. STAT. § 67.105(4), imposes liability on professionals to the same extent provided in the professional corporation act for shareholders who are professionals, Id. §§ 58.185, .187. The PC act limits joint and several liability to $300,000 for one shareholder and $2 million for two or more shareholders on a single claim of negligent or wrongful acts or omissions, per calendar year. Id. § 58.185(5). The amounts are indexed to inflation using the Consumer Price Index. Id. § 58.187.
89. Because partners in an LLP are liable for their own negligence and that of anyone they supervise, but not for negligence of co-partners, the incentive is for partners not to consult or oversee the work of others within the firm. Fortney, supra note 63, at 730–37; Vestal, supra note 44, at 516.
90. For a detailed comparison, see Stover & Hamill, supra note 80, at 821–38. Among the significant differences are that partners (including those in LLPs) have apparent authority to bind the partnership to third parties, while members in manager-managed LLCs do not, partners can dissociate at any time while members cannot, voting rights differ, and fiduciary duties differ, at
have developed had the LLP come first. The form is especially attractive to lawyers, accountants, and other professionals. In Oregon and some other states, the LLP form is available only to professionals.

A variation of the LLP is the limited liability limited partnership (LLLP). An LLLP is a limited partnership (organized under the limited partnership statute) that has registered as a limited liability partnership (under provisions of the general partnership statute). Oregon does not authorize LLLPs, but some states specifically provide for them and others implicitly allow them by including limited partnership within the definition of “partnerships” entitled to register as LLPs.

In an LLLP, management is centralized in the general partner as in a limited partnership. The limited partners are better shielded from liability than in a standard limited partnership in which participation in control can cause loss of limited liability. In many ways, an LLLP is comparable to a manager-managed LLC.

D. Current Choices and Prospects for the Future

The new forms of business organization—LLC, LLP, and LLLP—are in a sense artifacts of a federal income tax policy that has since been

least for non-managing participants. Ribstein, supra note 86, at 319–40; Stover & Hamill, supra note 80, at 845–46. LLCs shield owners from all types of business debts, while LLP statutes sometimes protect only from contract liabilities. Ribstein, supra note 86, at 321–27. Oregon’s LLC act allows perpetual existence while the LLP provisions do not. Culpepper, supra note 10, § 30A.11.

91. Oesterle & Gazur, supra note 14, at 105–06 (suggesting that “the LLC is a historical accident”); Thompson, supra note 69, at 5.

92. OR. REV. STAT. § 67.500(1) (stating that LLP must either render professional service or be affiliated with LLP that renders professional service); see also N.Y. PARTNERSHIP LAW § 121-1500(a) (McKinney Supp. 1997) (stating that LLP is available only for individuals and services providing professional services).

93. BROMBERG LLP, supra note 79, § 5.04, at 191. For a full discussion of the LLLP, see id., ch. 5.

94. Under Oregon’s Revised Partnership Act, “a partnership, not including a limited liability partnership, may register as a limited liability partnership.” OR. REV. STAT. § 67.500; see also BROMBERG LLP, supra note 79, at 185 n.1 (citing to uncodified Oregon law).

95. E.g., PA. CONS. STAT. ANN. § 820(a) (West 1995), cited in BROMBERG LLP, supra note 79, at 185 n.1.

96. BROMBERG LLP, supra note 79, § 5.04, at 191.

97. Id. at 191.

98. Id. §5.01, at 186. The two forms are not identical, however. For example, the shield from liability for a limited partner in a limited partnership may be different from the shield for a general partner in an LLLP. Id.
abandoned. Legislatures authorized the new forms primarily to allow closely held businesses limited liability together with pass-through tax treatment (and without the restrictions imposed on S corporations). Subsequent changes to the federal tax regulations rendered that original tax motivation moot.

“Check-the-box” regulations effective in 1997 abandoned the traditional pattern of categorizing non-corporate business entities as either “associations” (like corporations) or partnerships, based on the specified characteristics. Instead, an unincorporated entity can simply elect the tax treatment it prefers. Some differences in tax treatment continue to apply, however, and can be significant in certain circumstances.

Nevertheless, the LLC, LLP, and LLLP forms not only continue to exist but grow in popularity. The combination of characteristics that the newer entities offer, especially limited liability and flexibility in tax treatment and management structure, have great appeal, notwithstanding the change in tax regulations. Indeed, some wonder why any well-advised, closely held business would select any of the older forms.


100. The definition of corporation for purposes of the check-the-box regulations includes any entity organized under a state law describing it as incorporated, a corporation, body corporate, body politic, joint-stock company, or joint-stock association. Id. § 301.7701-2(b)(1), (3).

101. See supra note 62 and accompanying text.

102. An “eligible entity” (meaning one that is not a trust or a corporation) with two or more members can elect classification as either a partnership or an association taxed as a corporation. An eligible entity with only one owner can elect classification as either a corporation or a “disregarded entity.” Treas. Reg. § 301.7701-3(a), -3(b)(1).


104. Ribstein, supra note 24, at 331 (“In the wake of the development of the LLC and the LLP and the ‘check the box’ rule there is a new rule of thumb for entrepreneurs who are considering whether to incorporate: Don’t.”); Vestal, supra note 44, at 515 (“[I]t is difficult to conceive of a reason why knowledgeable and well-represented participants would organize a firm as a general partnership or limited partnership.”).
Still more business entity forms are available in some states or have been proposed. Some commentators have suggested that the entity legislation be reorganized, placing elements common to all forms of business organization into a single statute, defining alternatives more clearly, and reducing unintended overlaps. Another view is to reduce the number of choices. These proposals, however, are not likely to gain acceptance because they would have to overcome substantial inertia and would need a clear political constituency.

The new organizational forms clearly supplement rather than supplant the older ones. Some statutes and regulations, often written before the new entities became available, favor the more traditional forms. Familiarity among the business community and its counsel, as well as the more fully developed body of case law and document forms, assure that the older forms will not soon disappear. The corporation remains the predominant choice for new enterprises, due to such factors as inertia, well-developed rules, and a perception that formation of a corporation is


106. Professor Ribstein proposes a “limited liability sole proprietorship,” which would allow a single owner to gain limited liability without forming a single-shareholder corporation or one of the several entities requiring two or more owners. Larry E. Ribstein, The Loneliest Number: The Unincorporated Limited Liability Sole Proprietorship, 1. ASSET PROTECTION, May–June 1996, at 46.

107. See, e.g., Clark, supra note 40, at 173–75; Keatinge, supra note 24, at 31 (proposing business entity statutes comparable to traditional ones but “revised to eliminate the differences in the rules that are accidental rather than driven by policy differences”); Oesterle & Gazur, supra note 14, at 141–48. One basic approach would create a single business form, which might be called Universal Contractual Organization, Universal Business Organization or Limited Liability Entity. Another approach would create a “hub and spoke” statutory structure, with provisions applicable to all types of entities in one statute, and provisions specific to each type of entity in the individual organic acts governing those entities. Keatinge, supra note 24, at 81; see also Ribstein, supra note 24, at 334–35 (discussing several approaches).

108. One commentator argues that the “check-the-box” regulations removed the reasons for some of the distinctions and that some could be combined. For example, limited partnership could be absorbed by limited liability company. Clark, supra note 40, at 173–74.


110. Vestal, supra note 44, at 515 nn.124–25 (referring to some federal programs and franchise tax provisions in some states that favor limited partnership over LLP).
simpler and cheaper than formation of the alternatives. General partnership will continue by default, especially for the many businesses organized by persons who are less sophisticated in legal matters, because it is the only form that does not require a filing or other formality.

The proliferation of entities adds considerable complexity to the choice confronting business owners and their counsel, but also provides opportunities. One important consequence is that a particular business initially organized in one form may see advantages to converting to a different form, as its circumstances and its owners’ sophistication and needs develop over time. Also, the likelihood of dissimilar entities wishing to combine, particularly in a strong economy, increases as the number of entity forms increases.

III. CONVERSION AND MERGER UNDER TRADITIONAL LAW

Statutes authorizing changes in organizational form commonly treat the issue in a haphazard and inconsistent manner, without an apparent rationale. Although merger statutes provide direct and effective means of combining businesses, they apply only to certain combinations. Conversions have rarely been authorized. Where statutes do not explicitly provide for merger or conversion, they are not permitted by common law.

A. Transactions and Problems in the Absence of Statutes

Absence of statutory authority for changing an entity’s organizational form or for combining with another entity does not mean that such transactions will not occur. Instead, it means that businesses and their

---


112. MODEL BUS. CORP. ACT § 11.02, cmt. (1999). The United Kingdom provides an interesting illustration. In U.K. corporate law, the concept of merger does not exist. Nevertheless, corporate combinations and buyouts occur by use of more complicated techniques. One approach is for an acquiring company to first create a special-purpose vehicle—“Newco”—which borrows money and uses it to buy the stock of the target company, thereby becoming its holding company. Because the device of merger is not available to transfer the target’s assets to Newco, the target guarantees the obligations of Newco, thereby providing security to the lenders. Marc R. Paul & Charles A.A. Whitefoord, All Buyouts Are Not the Same: It Can Be Different in England, BUS. LAW TODAY, Mar.–Apr. 2000, at 18.
counsel are required to engage in procedures that are more creative, more cumbersome, and more expensive but accomplish the same result.

For example, a partnership can become a corporation or an LLC without benefit of statutory authority. At least three procedures achieve that result:

(1) The partnership dissolves, distributing its net assets in cash or in kind to the former partners. The former partners then transfer the assets to a newly organized corporation or LLC in exchange for stock or LLC membership interests.

(2) The partnership sells its assets to a newly organized corporation or LLC in exchange for the corporation’s stock or LLC’s membership interests. The partnership then dissolves, distributing its only asset, the stock, to the former shareholders.

(3) The partners transfer their partnership interests to a newly organized corporation or LLC in exchange for stock or membership interests. The corporation or LLC, which has become the sole “partner,” dissolves the partnership, acquiring the partnership’s assets.\footnote{113. Culpepper, supra note 10, § 39A.65; Lubin, supra note 9, at 362; Lion & Chacun, supra note 5, at 213. The IRS lists the same three methods for incorporating a partnership. Rev. Rul. 84-111, 1984-2 C.B. 88, cited in Culpepper, supra note 10, § 39A.65.}

In all three variations, the partnership dissolves, the assets and business continue in the hands of the corporation, and the former partners become shareholders of the corporation. A fourth approach, applicable when the jurisdiction authorizes cross-entity merger but not conversion, is for the partners to organize a corporation and then arrange for the partnership to merge into it.\footnote{114. Lubin, supra note 9, at 362.}

Though these processes work, they can be extremely burdensome and problematic. Under the tax code, distribution of assets to partners is a taxable event.\footnote{115. Culpepper, supra note 10, § 39A.65 (citing I.R.C. § 708 (2000); Priv. Ltr. Rul. 93-21-047 (Feb. 25, 1993); Priv. Ltr. Rul. 92-26-035 (Mar. 26, 1992)).} Under the Uniform Partnership Act, in a dissolution any partner may compel sale of the partnership property, with the proceeds used first to discharge its liabilities and then to pay in cash the amount owing to the partners.\footnote{116. OR. REV. STAT. § 68.600(1) (1999); UNIF. PARTNERSHIP ACT § 38(1) (1914).} Under the Revised Uniform Partnership Act, a partner can dissociate, compelling the remaining partners who wish to
continue the business to pay the value of the dissociated partner’s interest and indemnify against all partnership liabilities, both previously and subsequently incurred.117 Pragmatically, any dissenting partner can probably block a transaction intended to change the organizational form of the business by threatening to dissociate if it occurs.118

After dissolution, each partner continues to have power to bind the partnership not only for winding up119 but also for any other transaction if the other party had no knowledge or notice of the dissolution and the fact of dissolution had not been published.120 Thus, partners may be subjected to personal liability to third parties for post-dissolution debts.121 In addition, the corporation receiving an assignment of the partnership property may also be liable for these post-dissolution debts.122 If the entity continues the dissolved partnership’s business and assumes its debts, then creditors of the dissolved partnership are also creditors of the person or entity continuing the business.123

The dissolution of the partnership and sale of its assets may trigger or breach terms of existing contracts, requiring accelerated payment of debts or other changes.124 For example, mortgages commonly contain due-on-sale clauses requiring payment of the entire debt when the collateral is conveyed. Long-term commercial leases and franchise agreements commonly prohibit assignment without the express written consent of the landlord or franchiser.125 The mortgagee, landlord, or franchiser could either prevent the transfer, or impose burdensome terms or additional payments as the price for granting consent. Finally, some

---

118. Ribstein, supra note 86, at 325.
120. Or. Rev. Stat. § 68.570(1); Unif. Partnership Act § 35(1).
121. Ribstein, supra note 86, at 325.
122. Id.
123. Id. (citing Unif. Partnership Act § 41). The equivalent Oregon provision is Oregon Revised Statutes section 68.630(4).
124. See Ribstein, supra note 86, at 326.
real estate titles are subject to restraints on alienation that might be triggered by a transfer. 126

At a minimum, documentation will be necessary for transfer of assets, with the attendant expense and delay. Transfers of real estate require deeds and recording, may incur transfer taxes, and may trigger reassessment of property valuation. 127 Title insurance that protected the original insured party will likely not cover the transferee unless the title insurer consents to the transfer or provides an endorsement to the policy, which will probably be at additional cost. 128 For personal property, bills of sale and assignments must be prepared. Sales or transaction taxes may apply. 129

Any transaction involving a transfer of a security will trigger federal and state securities law, requiring either registration or an exemption. 130 Corporate stock is always a security, assuming it has the usual characteristics of stock. 131 Interests in a general 132 or limited partnership, 133 an

127. In California, for example, many counties and cities impose a documentary transfer tax when real estate is conveyed and the property assessment may be changed. See Lion & Chacon, supra note 5, at 172–74 (discussing when these results occur upon conversion of existing enterprise to LLC form).
128. Joyce Dickey Palomar, Limited Liability Companies, Corporations, General Partnerships, Limited Partnerships, Joint Ventures, Trusts—Who Does the Title Insurance Cover?, 31 REAL PROP., PROB. & TR. J. 605, 619 (1997). Title policies generally protect only the named insured and are not assignable; therefore, a transferee of real property seeking certainty that the policy will transfer should obtain consent from the title insurer or purchase a successor endorsement. Id. at 621. When a general partner leaves a general or a limited partnership, the result may be dissolution of the partnership and termination of the title policy, even if the remaining partners continue the business under the same name. Id. at 625–28. Similar problems occur if a general partnership changes its form to a limited partnership or corporation. Id. In an LLC, a member’s withdrawal may cause dissolution and lapse of the title policy. Id. at 635. Problems of this nature can be avoided or minimized if a statute makes clear that property of a business entity that engages in a conversion or merger passes by operation of law.
129. Real property transfer taxes apply in some jurisdictions when a deed is recorded, though policies of cities and counties differ when the deed merely reflects a change in the owner’s organizational form. Lubin, supra note 9, at 363.
131. United Hous. Found., Inc. v. Forman, 421 U.S. 837, 851 (1975) (noting that stock is characterized by free alienability, opportunity for dividends and appreciation, and voting rights proportional to number of shares owned).
132. Because partners in general partnerships typically have managerial power, the interests do not satisfy the Howey test and are not securities. See, e.g., SEC v. W.J. Howey Co., 328 U.S. 293,
LLC, or an LLP are likely to be deemed “investment contracts” and hence securities if the owner lacks meaningful control over management. Exemptions to registration are commonly available, but the applicability of securities law adds complexity and risk to the transaction.

B. Benefits of Conversion and Merger Statutes

Conversion and merger statutes greatly facilitate and simplify transactions to which they apply, in terms of property, transfer

301 (1946); Williamson v. Tucker, 645 F.2d 404, 425 (5th Cir. 1981); Casablanca Prods. v. Pace Int’l Research, 697 F. Supp. 1563 (D. Or. 1988). When a partner either does not have managerial power or does not have the experience and expertise to exercise formal power in a meaningful way, then the partnership interest can be a security. Koch v. Hankins, 928 F.2d 1471, 1477 (9th Cir. 1991).

133. Pratt v. Kross, 555 P.2d 765 (Or. 1976) (finding limited partnership interest to be a security); Kennedy & Owren, supra note 19, § 6.23.

134. In an LLC, a critical factor in determining whether a member’s interest is a security is whether the company is member-managed or manager-managed. If it is member-managed, the member does not rely on the essential managerial efforts of others, and so likely does not own a security. If it is manager-managed, a non-manager member likely does own a security. See generally LARRY E. RIBSTEIN & ROBERT R. KEATINGE, LIMITED LIABILITY COMPANIES §§ 14.02–.03 (1993); Culpepper, supra note 10, § 39A.71; Lion & Chacun, supra note 5, § 5.5, at 175–76; James J. Wheaton, Limited Liability Company Interests and the Securities Laws, in MARTIN I. LUBANOFF & BRIAN L. SCHOOR, FORMING AND USING LIMITED LIABILITY COMPANIES AND LIMITED LIABILITY PARTNERSHIPS 487, 500–07 (1994).


136. No case law was found discussing whether an LLP is a security, but no reason appears why an LLP should be treated differently for this purpose than other partnerships. LLPs are not merely like partnerships but by statute actually are partnerships. Ribstein, supra note 86, at 320. The special feature of limited liability is not one of the factors in the “economic realities” analysis of whether a form of investment is a security. The classic economic-realities test is that a particular instrument is within the statutory phrase “investment contract” if it represents an investment of money in a common enterprise with an expectation of profits from the efforts of others. Howey, 328 U.S. at 298–99.


138. For example, section 12(a)(1) of the Securities Act allows an investor to rescind a purchase of a security if the offering was neither registered nor exempt, even if no fraud was involved. 15 U.S.C. § 77l(a)(2). Other provisions such as section 12(a)(2) of the Securities Act and SEC Rule 10b-5 impose liability for fraud. 15 U.S.C. § 77l; 17 C.F.R. § 240.10b-5; see also Lion & Chacon, supra note 5, at 197 (discussing whether conversions or mergers constitute “sales” of securities).
restrictions, and transaction costs. Property of the converting or disappearing entity becomes property of the converted or surviving entity by operation of law, without the need for bills of sale, deeds, assignments, or other documents of conveyance.139 Title insurance continues in force in favor of the converted or surviving entity.140 Sales taxes and transfer taxes do not apply.141 Restrictions on transfer, such as due-on-sale clauses in mortgages and non-assignment clauses in leases, are not triggered unless specifically drafted to cover mergers.142

The treatment of conversion and merger under federal income tax, a critically important consideration, is complex and outside the scope of this Article. Change in the entity’s classification, from corporation (or “association”) to partnership, or the reverse, can be deemed a taxable liquidation, with consequences to both the entity and the owners.143 Certain other transactions are disregarded, producing no tax consequences, based on either the form or substance of the transaction.144

C. Limited Availability of Conversion and Merger: Oregon as an Example

Both conversion and merger have been recognized in the statutes for many years, but restricted to certain combinations. The development of the law in this field has been arbitrary and haphazard, explained only by history, not by policy rationale. The evolution of Oregon law provides an example.

Prior to 1993, the four chapters of Oregon statutes governing corporations, cooperatives, limited partnerships, and general partnerships varied considerably in rules for conversion and merger. Corporations and cooperatives were each authorized to convert to the

139. See infra Part IV.H.
140. Palomar, supra note 128, at 619.
141. See infra Part IV.H.
142. See infra Part IV.H.
143. See, e.g., Sheldon I. Banoff, Conversions of Legal Entities: Should Form or Substance Control? 416 PLI/Tax 101, 127 (1998) (finding “a frustrating absence of consistency as to the form versus substance issue” and repeated reversals of positions by IRS); Joseph H. Newberg & Stephen A. Evans, Entity Transfer Actions: Conversions of Existing Entities and Mergers; Dissolutions, in A PRACTICAL GUIDE TO MASSACHUSETTS LIMITED LIABILITY COMPANIES §§ 7.1–7.2 (1996) (stating that converting S or C corporation into LLC will trigger taxation at corporate and shareholder levels, but converting partnership into LLC is tax-free); Silverman et al., supra note 103; Philip B. Wright, Disregarded Entities Issues & Opportunities, 416 PLI/Tax 59 (1998).
144. Banoff, supra note 143, at 127.
other form—apparently the first example of cross-entity conversion. \(^{145}\) Each could merge with like entities (such as a corporation with another corporation), \(^{146}\) or could merge with the other forms (such as a corporation with a cooperative). \(^{147}\) Professional corporations were permitted to convert to or from business corporations by filing restated articles of incorporation. \(^{148}\)

Those combinations, however, were the only ones authorized. Corporations could not convert or merge with any type of partnership. The partnership law, based on the Uniform Partnership Act, contained no reference at all to mergers or conversions, so one partnership could not merge or convert even with another, much less a different type of entity. \(^{149}\) The limited partnership act had corresponding rules. \(^{150}\) The LLC act, adopted in 1993, allowed merger with another LLC, but not with any other entity. \(^{151}\)

The RUPA proposed by the ABA in 1994 and adopted in Oregon in 1997 \(^{152}\) made a number of major doctrinal innovations, laying the groundwork for more comprehensive reform. The RUPA authorized general partnerships, for the first time, to merge with other general partnerships and with limited liability partnerships. \(^{153}\) Further, the RUPA


\(146\). Prior to the 1999 amendments, two or more corporations could merge, id. § 60.481 (1997), and two or more cooperatives could merge, id. § 62.610 (1997).

\(147\). Prior to the 1999 amendments, a cooperative could merge with a corporation. Id. § 60.625 (1997).

\(148\). Id. § 58.125 (1997); Laird, supra note 30, § 31.44 (describing practice of Oregon Secretary of State’s Office, in absence of specific statutory provision for conversions).

\(149\). UNIF. PARTNERSHIP LAW, OR. REV. STAT. §§ 68.010–.650 (repealed 1997). Legislation in 1997 repealed the law, replaced it with a revised partnership act, and established a transition period ending January 1, 2003. 1997 Or. Laws 75, §§ 100, 84, 101 (codified at OR. REV. STAT. § 68.010 note (1999)).

\(150\). UNIF. LTD. PARTNERSHIP ACT (ULPA), OR. REV. STAT. §§ 70.005–.490. The Act was adopted in 1985. 1985 Or. Laws 677. Provisions regarding conversion and merger were not added until 1999. OR. REV. STAT. §§ 70.500–540. Neither ULPA nor RULPA provided for merging limited partnerships, either with like or unlike entities. Michael K. Pierce, Substantive Partnership Law: Special Problems of General and Limited Partnerships, SB 85 ALI-ABA 1, 86 (May 1, 1997).

\(151\). 1993 Or. Laws 173, § 90. The provision was not changed until the 1999 amendments, effective in 2000. OR. REV. STAT. § 63.481 (1999).

\(152\). OREGON REV. PARTNERSHIP ACT, 1997 Or. Laws 775, §§ 1–84 (codified at OR. REV. STAT. §§ 67.005–.815 (1999)).

introduced cross-entity merger to partnership law by authorizing a general or limited liability partnership to merge with a limited partnership\textsuperscript{154} subject to the limitation of LLPs to professionals,\textsuperscript{155} although curiously the Limited Partnership Act had no corresponding or reciprocal provision. Indeed the ULPA and RULPA did not even authorize merger of an LP into another LP.\textsuperscript{156}

In another innovative provision, the RUPA allows a general or limited liability partnership to convert into a limited partnership\textsuperscript{157} or the reverse.\textsuperscript{158} The primary purpose of the provision was to allow existing LPs to convert to the newly available LLP form, which is more attractive in some respects.\textsuperscript{159}

Finally, a general partnership, but not a limited partnership, can register as an LLP by filing a document with the state.\textsuperscript{160} In the opposite direction, an LLP can cancel its registration, thereby becoming a general partnership and subjecting its partners to personal liability for

---

Because an LLP is not a separate organizational form but rather a general partnership with special liability rules, this aspect by itself does not constitute cross-entity merger. The partnership act specifies that the term “partnership includes a limited liability partnership.” OR. REV. STAT. § 67.005(7) (1999). Thus a partnership that registers as an LLP, or cancels its registration, never changes its status as a partnership.

\textsuperscript{154} REV. UNIF. PARTNERSHIP ACT § 905; see also OR. REV. STAT. § 67.360 (1997) (repealed 1999).

\textsuperscript{155} OR. REV. STAT. § 67.400(1) (1999) (providing that partnership can register as LLP if it renders professional service or is affiliated with and provides services or facilities for LLP that renders professional service). A “professional” is defined for this purpose as an accountant, architect, attorney, chiropractor, dentist, landscape architect, naturopath, nurse practitioner, psychologist, physician, podiatrist, radiology technologist, real estate appraiser, or similar professionals. Id. § 67.005(12).

\textsuperscript{156} Pierce, supra note 150, at 86. Delaware and several other states, however, provide for LPs to merge with one another. Id.

\textsuperscript{157} REV. UNIF. PARTNERSHIP ACT § 902; see also OR. REV. STAT. § 67.345 (1997) (amended 1999).

\textsuperscript{158} REV. UNIF. PARTNERSHIP ACT § 903; see also OR. REV. STAT. § 67.350 (1997) (amended 1999). Oregon’s LP act, however, did not contain any corresponding provision, which was anomalous. The 1997 provision was replaced by more encompassing language in 1999. OR. REV. STAT. §§ 67.342, 70.505 (1999).

\textsuperscript{159} Primarily, the former LP’s general partner would no longer have personal liability and the former limited partners would no longer be precluded from participating in management. OR. REV. STAT. § 70.185 (1999) (subjecting general partners to liability); § 70.135 (protecting limited partners from liability unless they participate in management).

\textsuperscript{160} REV. UNIF. PARTNERSHIP ACT § 1001; see also OR. REV. STAT. § 67.500 (1999). However, in Oregon, LLP status is available only to partnerships that render professional services. OR. REV. STAT. § 67.500.
subsequent business debts. The business remains the same entity that it was before filing the registration, or before its registration ceased.

The RUPA dealt specifically with the situation of a general partner becoming a limited partner as a result of a conversion, without the knowledge of outside parties. As to obligations incurred before the conversion, the partner remained personally liable. As to obligations incurred within ninety days after the conversion to limited partnership, an outside party who reasonably believed that the limited partner was a general partner could impose liability.

It is particularly noteworthy that the RUPA authorized a general partnership to merge with, convert into, or register as an entity with limited liability for owners. Those provisions opened significant opportunities for businesspeople to exit from business forms imposing personal liability. This option likely will be used most frequently by persons who enter into business, without legal sophistication or counsel, as general partners, due to failure to affirmatively seek limited-liability status, and later decide on another organizational form.

Despite the many advances included in the RUPA, limitations remained. It affected general (including LLP) partnerships and limited partnerships, but no other form of business organization. The RUPA does not authorize merger of any type of partnership with a corporation or an LLC, or conversion of any type of partnership with any non-partnership entity.

The resulting array of merger and conversion rules was inconsistent and irrational in Oregon. Several types of cross-entity mergers and conversions were possible, while others were not. This disarray is likely to remain in any state that has not comprehensively examined and revised its laws.

161. OR. REV. STAT. § 67.595. Changes between general partnership and LLP are arguably not conversions to a different entity form, because the act treats LLP not as a separate type of entity but rather as a partnership with special characteristics, but that distinction seems more metaphysical than meaningful.

162. REV. UNIF. PARTNERSHIP ACT § 904; see also OR. REV. STAT. § 67.500.

163. REV. UNIF. PARTNERSHIP ACT § 902(e); see also OR. REV. STAT. § 67.345 (1997) (amended 1999).

164. REV. UNIF. PARTNERSHIP ACT § 902(e); see also OR. REV. STAT. § 67.345 (1997) (amended 1999).
D. The Issue of Personal Liability Entities Changing to Limited Liability Entities

A significant theoretical issue is whether an entity in which owners are subject to personal liability for business debts (general partners in general partnerships or in limited partnerships) should have the option of converting or merging into a business form protecting them from liability such as corporation, LLC, or LLP. A concern is that creditors of a partnership might be unfairly hindered or disadvantaged in pursuing their claims against the owners once it becomes, for example, a corporation. Some case law has held that partners who incorporate remain liable to creditors who do not receive notice of incorporation.165 Conversely, shareholders of a corporation that converts or merges into a partnership might unfairly or unwillingly be subjected to personal liability.

Consistent with those concerns, merger of corporations with partnerships is authorized by relatively few states, including Colorado,166 Maryland,167 Pennsylvania,168 and Texas.169 Delaware is an example of a state that generally offers great flexibility in cross-entity conversion and merger170 but excludes the possibility of entities such as corporations converting or merging with general partnerships.171

165. See, e.g., Burke Mach. Co. v. Copenhagen, 6 P.2d 886 (Or. 1932); Annotation, Liability of Former Partners as Such in Respect of Transactions Subsequent to Incorporation of Their Business, 89 A.L.R. 986 (1943).
166. COLO. REV. STAT. ANN. § 7-90-203(1), (2) (West Supp. 1998) (allowing domestic entity to merge with one or more domestic or foreign entities).
167. MD. CODE ANN., CORPS. & ASS’NS § 3-102(4) (allowing Maryland corporation to merge into Maryland or foreign partnership), § 4A-211(a) (allowing Maryland general or limited partnership to convert to LLC), § 4A-701 (allowing Maryland LLC to merge with LLC, general or limited partnership, corporation, or business trust), § 9A-901(a) (allowing Maryland partnership to merge with partnership, LLC, limited partnership, corporation, or business trust), § 10-208(b) (allowing Maryland limited partnership to merge with any other entity) (1999).
170. Pierce, supra note 150, at 86. In Delaware, limited partnerships may merge with other limited partnerships or with general partnerships, joint stock companies, corporations, and other entities if permitted by the laws governing those entities. DEL. CODE ANN. tit. 6, § 17-211 (LP act); tit. 8, § 263(a) (corporation act) (Supp. 2000).
171. A Delaware corporation may merge with a business or nonprofit corporation, joint stock association, and limited partnership. DEL. CODE ANN. tit. 8, § 251 (1974 & Supp. 2000). However, the term “joint stock association” specifically excludes partnership, making merger with a partnership unauthorized. Id. § 254(a); Huff, supra note 169, at 140 n.100.
In a broader perspective, however, conversion or merger of personal liability entities to or with limited liability entities is well established. Sole proprietorships and partnerships have long been allowed to organize corporations, for example, and to transfer their property to the new entity. Several states allow an LLC to merge with an LP, and an LP to merge with a general partnership, even though every LP includes at least one general partner with personal liability. Of greater importance, the RUPA allows a general partnership (with personal liability) to register as an LLP (with limited liability) merely by filing a statement of qualification and allows an LLP to abandon limited liability protection for its partners by canceling its registration.

E. A Situation Ripe for Reform

The overall pattern was difficult to defend. A general partnership, whose partners are subject to personal liability, could merge or convert into an LP or LLP, which provides limited liability, but not into an LLC or corporation, which also provides limited liability. An LLP, which provides limited liability, could convert or merge into a general partnership, but an LLC or corporation could not do likewise. Even entities with limited liability could not merge into other entities with limited liability, such as an LLC with a corporation.

The only explanation for that anomalous state of the law appeared to be that the new policy reflected in the RUPA had not yet been reflected in the other statutes. As a result, the opportunities available to newly organizing businesses to select among the recently expanded list of entity choices were not as readily available to existing businesses.

The logical course was to extend the opportunities to change business entity form to all possible combinations. Moreover, changes should be permitted for two entities, in the form of merger, or for a single entity, in the form of conversion.

172. E.g., COLO. REV. STAT. ANN. § 7-90-203(1), (2) (West Supp. 1998); DEL. CODE ANN. tit. 6, § 17-211 (Supp. 2000); MD. CODE ANN., CORPS. & ASS’NS § 4A-701.
174. Id. § 903.
175. For example, Oregon’s merger provision prior to 1999 allowed a corporation to merge only with another corporation but not with any other type of entity. OR. REV. STAT. § 60.481 (1997).
IV. THE TREND TOWARD CONVERSION AND MERGER OF DISPARATE ENTITIES

A number of states authorize conversions and mergers of disparate business entities, though in different forms. Since 1988, Delaware has permitted corporations to merge with partnerships, nonprofit corporations, joint-stock associations, and limited partnerships. In addition, Delaware limited partnerships can merge with multiple business entities. Texas amended its corporation act in 1989 to broadly authorize cross-entity conversion and merger. Colorado and Nevada have comprehensive approaches with modernized provisions. Other states that authorize at least some conversion and merger of disparate business entities include California, Georgia, Illinois, Kansas, Maryland, Oklahoma, Tennessee, and West Virginia.
In 1999, the Model Business Corporation Act was amended to authorize mergers of corporations with “other entities,” though the reform did not provide for conversions.\textsuperscript{192}

Flexibility in a state’s laws in this regard is mentioned among the reasons to select a particular state as the state of incorporation or organization.\textsuperscript{193} Business headquarters or operations need not be located in the state selected for organization. Moreover, the initiative taken by some states to permit conversion and merger of disparate entities tends to lead the remaining states in the same direction.

A. Use of Foreign Law To Accomplish What Domestic Law Fails To Provide

The differing policies among states, coupled with traditional full faith and credit principles, opens opportunities for business counsel to accomplish goals not directly authorized by the laws of the state of organization of the business entity. For example, an Oregon corporation seeking to convert to LLC form, prior to the 1999 legislative amendments, was precluded from doing so directly because Oregon statutes allowed mergers of like entities only.\textsuperscript{194} However, the Oregon corporation could organize a Texas or Delaware LLC and merge into it under Texas and Delaware law allowing cross-entity mergers.\textsuperscript{195} The enterprise could then operate in Oregon as a foreign LLC.\textsuperscript{196}

Similarly, if an Oregon corporation wished to merge into an Oregon limited partnership, it could not do so directly prior to the 1999

\begin{footnotes}
\item[191] W. VA. CODE ANN. § 31-1-38 (Michie 1999) (authorizing corporation merging with stock, nonstock, or nonprofit corporation).
\item[193] A commentator in a Texas bar journal, for example, notes that both Texas and Delaware permit corporations to merge with other types of entities, including nonprofit corporations and limited partnerships, but that Texas offers a wider range of entities and clearly permits some mergers on which Delaware statutes are ambiguous. Huff, \textit{supra} note 181, at 37.
\item[195] DEL. CODE ANN. tit. 8, § 264(a) (1974 & Supp. 2000) (authorizing merger of domestic corporation with foreign or domestic LLC); TEX. BUS. CORP. ACT art. 5.01, .03, .06 (Vernon 2001) (authorizing mergers of corporations with “other entities”).
\item[196] A foreign LLC (meaning one organized under the laws of another state, regardless of where the company actually is located or does business) need merely pay a fee and obtain a certificate of authorization from the Oregon Secretary of State. OR. REV. STAT. § 62.027 (1999).
\end{footnotes}
amendments, but could nevertheless accomplish its objective. First, the Oregon corporation could organize and merge into a Texas corporation, and the Oregon LLC could similarly organize and merge into a Texas LLC. The two Texas entities could then merge under Texas law, and the business could continue to operate in Oregon as a foreign entity. Accordingly, a state’s failure or refusal to explicitly authorize cross-entity mergers does not prevent those transactions. The state merely subjects its businesses to unnecessary complexity and expense for no apparent benefit.

When a state does decide to modify its statutes, at least two drafting approaches are available. One is the “junction box” model, with cross-entity provisions in a single statute. Another is the self-contained approach, repeating the cross-entity provisions in each of the statutory chapters governing one of the business-entity forms.

B. The “Junction Box” Model

One legislative drafting model, which has been characterized as a “junction box,” is used in Colorado and Nevada. In this model, provisions governing mergers of dissimilar forms are located in a statute separate from the organic statute such as the corporation act and the partnership act. The approach is said to assure consistency and require fewer provisions than if cross-entity provisions were inserted into each organic statute.

197. Pierce, supra note 150, at 85.
198. H. Gregory Austin, An Introduction to the Colorado Uniform Partnership Act (1997), 27 COLO. LAW. 5, 12 (1998); Keatinge, supra note 1, at 44. The junction box statute is formally known as the Colorado Corporations and Associations Act, enacted in 1997. Austin, supra, at 12–14. Its provisions for inter-entity merger and conversion overlap with provisions of Colorado’s Uniform Partnership Act authorizing merger and conversion of partnerships with other general or limited partnerships. Without reason, differences exist in the procedures for conversion and mergers in the junction box statute as compared to the partnership act. Id. at 12–13.
200. NEV. REV. STAT. ANN. 92A.005–.510 (Michie 1999). The statute allows merger of any “entities,” defined to include corporations, nonprofit corporations, LLC, limited partnerships, and business trusts. Id. 92A.045.
201. One author calculated that provisions to allow each of six types of entity to merge or convert into each of the other five types would result in twenty possible transactions per statute, or 120 provisions total. Keatinge, supra note 24, at 76 n.194. In fact, however, that unreasonable result can readily be avoided by the expedient of amending each of the six organic statutes to authorize conversion or merger of the form of entity it governs into any “business entity,” defined to include all six forms. Procedures and rights applicable to each form are specified in its own
In the Colorado statutes, general provisions applying to all types of business entities permit any domestic or foreign entity to convert to any other or to merge with any other.\textsuperscript{202} The terms “domestic entity” and “foreign entity” are defined to include corporation, general partnership (which includes LLP), cooperative, LLC, LP, limited partnership association, nonprofit association, and non-profit corporation.\textsuperscript{203} The entity resulting from a conversion or merger is for all purposes the same entity as the previous entity or entities. The statute describes the procedures and the effect on property, liability, and governing documents.\textsuperscript{204}

Following those provisions, however, is another that provides that any conversion or merger is subject to the rules of any organic statute or the common law that prohibit or restrict the transaction, grant dissenters rights, or impose other requirements.\textsuperscript{205} Hence, any business converting to a different form, and any business merging with a business of unlike form, must refer to three statutes. For a conversion, the entity will refer to the statute applicable to the entity before the transaction, the separate statute allowing conversion, and the statute applicable to the converted entity. For a cross-entity merger, the statutes are the ones applying to each of the entities, plus the one permitting merger of unlike entities.

\section*{C. The Self-Contained Model}

A more direct approach to drafting statutes authorizing conversion and merger of disparate business entities is used in states such as Maryland\textsuperscript{206} and now Oregon.\textsuperscript{207} Similarly, 1999 amendments to the organic statute, but in a manner that is coordinated with the others and uniform wherever possible. See the Appendix of this Article for a chart illustrating that approach in Oregon’s laws.

\begin{itemize}
  \item \textsuperscript{202} \textit{COLO. REV. STAT. ANN. §§ 7-90-201 to -206 (West Supp. 2000).} The provision is located in title 7 of the Colorado Revised Statutes, Corporations and Associations, which encompasses the nonprofit corporation, cooperative, partnership (including general, limited, and limited liability partnership), LLC, and corporation forms of business organization. The provision was labeled and positioned in the code in a manner making it difficult to find. The merger provision is in a category located between LLCs and corporations labeled merely “Corporations and Associations” (the same label as the entire Title).
  \item \textsuperscript{203} \textit{Id.} § 7-90-102.
  \item \textsuperscript{204} \textit{Id.} § 7-90-206.
  \item \textsuperscript{205} \textit{Id.} § 7-90-206.
  \item \textsuperscript{206} \textit{MD. CODE ANN., CORPS. & ASS’NS § 3-102 (1999).}
  \item \textsuperscript{207} \textit{OR. REV. STAT. §§ 60.470–.487 (corporation), 60.605–.623 (cooperative), 63.467–.497 (LLC), 67.340–.365 (partnership), 70.500–.540 (limited partnership) (1999).}
\end{itemize}
Cross-Entity Conversion and Merger

Model Business Corporation Act authorizing cross-entity merger but not conversion use the self-contained drafting approach. Each organic statute contains all of the provisions needed to convert or merge that type of entity, and is the only one that need be consulted if the transaction involves another entity of the same type. For cross-entity conversions or mergers, the statute governing each type of entity applies. For example, the partnership act specifies that when a partnership and corporation intend to merge, the corporation must follow the procedures specified by the corporation act.

As compared to Colorado’s “junction box” approach, this drafting technique entails somewhat more repetition and length, but produces a result easier for business executives and their counsel to utilize. Those seeking to convert an entity to a different form need consult only the chapters applicable to the entity before conversion and the chapter applicable after conversion, and not a third chapter for the transaction itself, as in Colorado. Similarly, an entity merging with an entity of unlike form need consult only the statutory chapters for the two entity forms, and not a third chapter for the transaction. The risk of inconsistency is avoided by consciously making the provisions identical except as necessary to reflect the distinctive characteristics of a particular form of organization.

V. OREGON’S BUSINESS ENTITY CONVERSION AND MERGER LEGISLATION

A. Legislative History

In Oregon, efforts to reform and expand statutory authorization for conversion and merger of different business entities resulted in formation of a Task Force on Multi-Entity Mergers of the state bar’s

---

208. The Model Act, as revised in 1999, allows a merger of a corporation with a corporation or “other entity,” defined as “any association or legal entity, other than a domestic or foreign corporation, organized to conduct business, including, without limitation, limited partnerships, general partnerships, limited liability partnerships, limited liability companies, joint ventures, joint stock companies, and business trusts.” MODEL BUS. CORP. ACT §§ 11.01(d), 12(a) (1999).

209. MD. CODE ANN., BUS. REG. §§ 3-102 (allowing corporation to merge with multiple other types of entities), 4A-701 (allowing merger with LLC), 10-208 (LP), 10-208(f) (general partnership) (1999).

210. Id. § 9-903(2)(ii); OR. REV. STAT. § 60.481.
Business Law Section in 1997\textsuperscript{211} and 1998.\textsuperscript{212} The task force included a law professor, representatives of the Corporation Division of the Secretary of State’s office and the title industry, and lawyers from around the state who advise businesses ranging from small, closely held enterprises to large, publicly traded corporations.

In 1998, the task force developed a proposal that received the approval of the business law section and the state bar. It was submitted to the 70th Oregon Legislative Assembly in its 1999 Regular Session, where it became Senate Bill 145.\textsuperscript{213} Hearings were held before the Senate Judiciary Committee and the House Committee on Business and Consumer Affairs. The task force chair Douglass Schmorr, and its principal drafter, the author of this Article, testified. Questions and commentary from the legislators were minimal, the only amendment was a technical correction requested by the task force,\textsuperscript{214} and enactment was not controversial.

\textsuperscript{211} The first committee, chaired by Eva Kripilani, studied the issue and prepared a table comparing the laws of many states.

\textsuperscript{212} The second committee was chaired by Douglass Schmorr of Brophy, Mills, Schmorr, Gerking & Brophy. Robert Art, law professor at Willamette University and author of this Article, served as principal drafter for the committee. Other members were Ernest Bootma of Dunn, Carney, Allen, Higgins & Tongue, David C. Culpepper of Miller Nash Wiener Hager & Carlsten, Jacob A. Heth of Hagen, Dye, Hirsch, & DiLorenzo, Margaret Kushner of Stoel Rives, and Jennifer Holt Mair of Louisiana Pacific Corporation. Representing the state agency that would be responsible for conversion and merger filings were Jan Sullivan, director of the Secretary of State Corporation Division, and Tom Wrosch of the Corporation Division. The task force consulted David R. Aldrich of Transnation Title Insurance Company on property transfer issues. Peter J. Bragdon of Stoel Rives was the business law section’s legislative committee chair. Susan Grabe of the Oregon State Bar was liaison to the state bar.

\textsuperscript{213} Senate Bill 145 was printed at the request of the Senate Interim Judiciary Committee.

\textsuperscript{214} The amendment was a change to a single line, repeated in each of the five chapters of the Oregon Revised Statutes, to eliminate a possible unintended inference from the original formulation that obligations of an entity could be imposed on business owners as a result of a merger. The corrected formulation is: "(c) All obligations of each of the business entities that were parties to the merger, including, without limitation, contractual, tort, statutory and administrative obligations, are obligations of the surviving business entity . . . ." House Amendments to Senate Bill 145 by Committee on Business and Consumer Affairs; 1999 Or. Laws 362 §§ 14(1)(c) (corporation), 27(1)(c) (cooperative), 39(1)(c) (LLC), 49(1)(c) (partnership), 64(1)(c) (LP). House Amendments by the Committee on Business and Consumer Affairs changed the earlier phraseology, which was: "(c) All obligations of each of the business entities or its owners, including, without limitation, contractual, tort, statutory and administrative obligations, are obligations of the surviving business entity," or other phraseology that was similar but by error not identical. SB 145 §§ 14(1)(c) (corporation), 27(1)(c) (cooperative), 39(1)(c) (LLC), 49(1)(c) (partnership), 64(1)(c) (LP). A second amendment on an unrelated matter was inserted because filing deadlines made it difficult to introduce that issue as a separate bill. 1999 Or. Laws 362, amending OR. REV. STAT. §§ 732.521, .538 (regarding health care service contractors).
B. The Central Premise: Continuity of Entity Existence

The existence of a business throughout a conversion or merger is vital for such purposes as title to property, contract rights, and claims of creditors. For example, if a limited partnership converts or merges into a LLP, it does not dissolve and then reorganize but rather continues on in the new LLP form. This distinction, which might appear metaphysical, has major substantive consequences.

Such clarity has long been present as to mergers of corporations. For example, when a corporation that owns real estate subject to a mortgage to a lender and protected by a title insurance policy merges with another corporation, the real estate transfers by operation of law.215 No deed is required.216 The succession in ownership does not trigger a due-on-sale clause in the mortgage or transfer taxes unless the contractual clause or tax legislation specifically applies to mergers, because a merger is not deemed to be a transfer of assets.217 The title policy that covered the first corporation now covers the surviving corporation.218 Statutes allowing conversion of general partnerships to LLPs, or limited partnerships to or from other partnership forms, have been similarly clear that the business remains the same entity.219

The same certainty is now provided for conversion or merger of any business entity to or with any other business entity. For conversions, the principle that the business entity remains the same one despite the

215. See, e.g., FRANKLIN A. GEVURTZ, CORPORATION LAW § 7.2(b) (2000) (“[T]he surviving corporation in a merger . . . succeeds by operation of law to all of the assets and liabilities of the disappearing corporations . . ..”).

216. See, e.g., MODEL BUS. CORP. ACT § 11.07(a) (1999). (“[A]ll property owned by, and every contract right possessed by, each corporation or other entity that merges into the survivor is vested in the survivor . . . .”). The Official Comment to section 11.07 further specifies that “the survivor automatically becomes the owner of all real and personal property.” MODEL BUS. CORP. ACT § 11.07 cmt.

217. The Official Comment to the Model Business Corporations Act emphasizes that “[a] merger is not a conveyance, transfer, or assignment,” and the vesting of property in the surviving corporation “does not give rise to a claim that a contract with a party to the merger is no longer in effect on the ground of nonassignability, unless the contract specifically provides that it does not survive a merger.” MODEL BUS. CORP. ACT § 11.07, cmt.

218. Palomar, supra note 128, at 619 (citing American Land Title Ass’n Owner’s Policy, Conditions & Stipulations para. 1(a) (Oct. 17, 1992)).

change in organizational form is stated unequivocally. The corresponding phraseology is that the “separate existence” of every business entity other than the survivor ceases, indicating that existence continued throughout the merger and that only separateness ceased.

C. Principles and Goals

The legislation provides opportunities to all types of business, without favoring one size or organizational form over another. Moreover, it carefully protects the interests of business owners, creditors, and others dealing with businesses. Key policy decisions, reflected throughout the legislation and explained in greater depth later in this Article, include the following:

(1) Utilization of existing procedures and rules applicable to each type of entity. Wherever possible, the conversion and merger provisions coordinate with the procedures, rules, and rights applicable to the business entity prior to the legislation. Provisions previously applicable to mergers between like entities also apply to newly authorized mergers between unlike entities and to conversions. If mergers were not previously authorized at all, procedures and rights were written to be consistent with those applicable to other organic decisions such as dissolution.

(2) Protection of owners’ rights. Conversions and mergers, like other organic changes, are critically important to owners of the enterprise, changing their ownership, control relationships, and exposure to liability. Moreover, such changes could provide an occasion for freeze-out or other serious detriment to certain owners. The law addressing these issues should be consistent with the law and agreements applicable before the transaction, which will vary among the different types of entities.

220. The “effects of conversion” provision in each statute clearly states the continuity principal. OR. REV. STAT. §§ 60.478(1)(a) (corporation), 62.613(1)(a) (cooperative), 63.479(1)(a) (LLC), 67.348(1)(a) (partnership), 70.520(1)(a) (LP) (1999).

221. The “effects of merger” provision in each statute has the same continuity language. Id. §§ 60.497(1)(a) (corporation), 62.623(1)(a) (cooperative), 63.497(1)(a) (LLC), 67.365(1)(a) (partnership), 70.540(1)(a) (LP).

222. See infra Part V.E.

223. See infra Part V.G.
(3) **Coordination among business entity statutes.** The procedures (including requirements for filing with the state) and rights of owners and creditors are identical among the statutory chapters governing each of the various entities, except as necessary to accommodate the distinctive characteristics of some of the entities. Business owners, counsel, and others should not be faced with inconsistent or incompatible provisions in different chapters.

(4) **Preservation of rights of creditors and litigants.** Scrupulous protection of the rights of creditors of the business and litigants is critical, lest conversion and merger be used as a means of defrauding or disadvantaging innocent parties. Rights and claims arising before a conversion or merger, by or against the business or its owners, are not to be changed by the transaction. Rights and claims arising after a conversion or merger are determined according to the law applicable to the converted or surviving entity. Special treatment is accorded, however, to creditors who deal with an entity in which the owners have personal liability (almost always general partners), and then continue to deal with it after a conversion to a limited liability form.

(5) **Efficient filings and effective tracking.** A public record is maintained of every conversion and merger so that, among other reasons, creditors of businesses can track the disposition and location of assets and obligees. The procedure is as efficient and direct as possible, requiring only a single filing to report the exit of a business from one form of entity and the entrance to a new form.

(6) **Coordination with foreign jurisdictions.** The flexibility and advantages that Oregon law provides to domestic business entities extend to entities organized under the laws of other jurisdictions, including states of the United States and foreign countries, to the extent permitted by the law of the foreign jurisdictions.

The rules and processes designed to accomplish those policies are uncomplicated, direct, and consistent. The merger provision in the

---

224. See infra Parts V.F., V.I.
225. See infra Part V.F.
226. See infra Part V.I.
Oregon Business Corporation Act, which had considerable history of interpretation, was used as the model for not only the new section on corporate conversions but also the new conversion and merger provisions in the statutes governing the other entities. The provisions are coordinated and parallel among the several chapters to eliminate the danger of incompatible rules. Nevertheless, including conversion and merger provisions in each chapter allowed the opportunity to tailor them as necessary to address the distinctive characteristics of each type of entity. Prior provisions allowing certain specified combinations or conversions were repealed and replaced by provisions that more broadly authorize any cross-entity conversion or merger.

D. Types of Business Entities Included

Cross-entity conversion and merger are allowed for any “business entity.” That key term is defined to include corporation, professional corporation, cooperative, partnership, limited partnership, and limited liability company. The definition is repeated verbatim in all of the organic acts to preclude any inconsistencies.

Both general and limited liability partnerships can convert or merge. However, for a transition period ending in 2003, general partnerships are eligible for this treatment only if they are governed by the most recent partnership act (based on the RUPA). The reason is that the 1999

230. For example, the old partnership act allowed conversion of a partnership to a limited partnership, and the reverse. OR. REV. STAT. §§ 67.345–350 (1997) (repealed in 1999). Those provisions were repealed in 1999, SB 145§ 65 (Feb. 25, 1999), and replaced by provisions allowing those or any other type of conversion. OR. REV. STAT. §§ 67.342 (partnership), 70.505 (limited partnership) (1999).
232. The definition of “business entity” for the conversion and merger provisions includes “[a] partnership organized in Oregon after January 1, 1998, or that is registered as a limited liability partnership, or that has elected to be governed by [OR. REV. STAT. ch. 67], and a partnership governed by law of another jurisdiction that expressly provides for conversions and mergers. . . .” OR. REV. STAT. §§ 60.470(1)(e) (corporation), 62.605(1)(e) (cooperative), 63.467(1)(e) (LLC), 67.340(1)(e) (partnership), 70.500(1)(e) (LP) (1999). Hence, a partnership governed by the older OREGON REVISED STATUTES chapter 68 (based on UPA) is not a “business entity” eligible to convert or merge under the new provisions. Starting January 1, 2003, all partnerships will be governed by the new act. 1997 Or. Laws 775, § 84 (codified at OR. REV. STAT. § 67.005 note, § 68.010 note (1999)).
Cross-Entity Conversion and Merger

legislation utilizes internal governance procedures specified in the new partnership act, for approving conversions and mergers. Such procedures are absent from the earlier partnership act, which was silent on conversions and mergers. There seemed to be no benefit in amending the old act that was already scheduled to be repealed in 2003. A partnership currently governed by the older act can easily amend its partnership agreement or register as an LLP to take advantage of the new legislation.

Foreign partnerships can take advantage of the proposed legislation if they are governed by a statute similar to the RUPA or any other statute that authorizes conversions and mergers. For professional corporations, matters of internal governance and organization are determined by the business corporation act. As a result, the 1999 amendment was limited to a simple cross-reference.

The list of business forms permitted to convert or merge does not include certain additional entities such as real estate investment trusts, although some states do extend the list to those forms. The 1999

233. OR. REV. STAT. §§ 67.005–.815 (based on RUPA).
234. Id. §§ 67.344 (action on plan of conversion), 67.361 (action on plan of merger).
235. Id. §§ 68.010–.650 (based on UPA).
236. 1997 Or. Laws 775, § 84(2)(b) (codified at OR. REV. STAT. § 67.005 note, § 68.010 note).
237. OR. REV. STAT. § 67.340(1)(e).
238. Id. § 58.045.
239. The Texas Business Corporation Act, for example, is exceptionally liberal in authorizing conversion or merger with a long list of unlike entities, extending to any entity, whether organized for profit or not, that is a corporation . . . , limited or general partnership, limited liability company, real estate investment trust, joint venture, joint stock company, cooperative, association, bank, trust, insurance company or other legal entity organized pursuant to the laws of this state or any other state or country.

TEX. BUS. CORP. ACT ANN. art. 1.02.A(20) (Vernon Supp. 2001). Delaware authorizes conversion among “a corporation, business trust or association, real estate investment trust, common law trust, or any other unincorporated business, including a partnership (whether general . . . or limited . . . ), and a foreign limited liability company . . . .” DEL. CODE ANN. tit. 6 § 18-209(a) (1974 & Supp. 2000), § 18-209(b) (domestic limited liability company), §§ 18-214(a)–(b) (merger among same entities). The Colorado Corporations and Associations Act authorizes conversion and merger among nine forms: corporations, nonprofit corporations, general partnerships, limited partnerships, LLCs, limited partnership associations, cooperatives, and nonprofit associations. COLO. REV. STAT. ANN. § 7-90 (West 1999). The Model Business Corporation Act, as revised in 1999, allows a merger (but not conversion) of a corporation with an “other entity,” which is defined as “any association or legal entity, other than a domestic or foreign corporation, organized to conduct business, including, without limitation, limited partnerships, general partnerships, limited liability partnerships, limited liability companies, joint ventures, joint stock companies, and business trusts.” MODEL BUS. CORP. ACT §§ 11.01(d), .02(a) (1999).
amendment makes no change to existing law on merger of nonprofit corporations, which are allowed to merge with business corporations under certain circumstances. Nothing prevents a nonprofit that wishes to merge with another entity, such as an LLC, from doing so in a two-step process: first the nonprofit corporation merges into a business corporation, and then the business corporation merges into an LLC.

Both domestic and foreign entities can use the provisions, provided that the statutes governing all affected entities permit the transaction. For conversion, the statutes governing the entity before and after the transaction must permit it. For merger, statutes governing all parties must permit it. The policy of respecting the authority of other states and countries to determine whether the business entities organized under their laws can convert or merge is consistent with prior Oregon law.

For Oregon corporations, cooperatives, limited liability companies, partnerships, and limited partnerships, this limitation will present no

240. The task force considered whether the 1999 amendments should be extended to cover nonprofits, and decided against it. Nonprofit organizations raise issues not involved with the other types of entities, including property tax exemptions, bingo for charitable organizations, protecting the charitable nature of its assets, and monitoring by the Oregon Department of Justice. The issue was deemed best left for another day, and the term selected to encompass the entities eligible for conversion or merger was “business entity.”

241. Nonprofit corporations may merge with other nonprofit or business corporations, Or. Rev. Stat. § 65.481 (1999), but special rules apply to public-benefit or religious corporations. Unless they obtain the prior consent of the Attorney General or approval of a circuit court, they can merge only with a similar corporation, or with a business corporation provided that the fair market value of the public-benefit or religious corporation’s assets are transferred to a similar corporation, and that no member receives an improper benefit. Id. § 65.484. If a business corporation is a party, it must comply with the provisions of the Business Corporation Act. Id. § 65.504. The Nonprofit Corporation Act makes no mention of conversion and no mention of merger with entities other than corporations.

The special provisions for public-benefit and religious corporations respond to the danger of diversion of charitable resources to non-charitable uses. A comparable approach in Colorado allows conversions or mergers of a nonprofit corporation with a business corporation only if the same result could be achieved without using the conversion or merger statute. For example, a nonprofit corporation can merge with a business only if it transfers the value of its assets to another nonprofit. Austin, supra note 198, at 12 (citing COLO. REV. STAT. § 7-90-206(2) (1998)).

242. The first step is permitted by Oregon Revised Statutes sections 65.481 and 60.481, and the second step is permitted by sections 60.481 and 63.481. Similar transactions are described supra note 10 and accompanying text.


244. Id. §§ 60.481(1) (corporation), 62.617(1) (cooperative), 63.481 (LLC), 67.360 (partnership), 70.525 (limited partnership).

Cross-Entity Conversion and Merger

problem; authorization is explicit and uniform. For entities organized in other states and countries, it will be necessary to investigate the law of the applicable jurisdiction, which will vary. Oregon does not assert power to affect the existence or form of an entity governed by law of a different jurisdiction without that jurisdiction’s concurrence.

Oregon entities also are authorized to convert or merge into entities governed by the law of other jurisdictions, if permitted by that law. For example, an Oregon limited partnership could convert into a Texas corporation using Texas law. Although certain provisions expressly allowing mergers based on law other than Oregon’s were repealed, that was not intended to and did not negate the possibility.

E. Procedures for Approving the Transaction

A plan of conversion or plan of merger must be prepared before any transaction, stating the key elements and consequences. Such plans identify the names and types of entities before and after the transaction, and the means of converting ownership rights to some other form of property. Plans supply whatever information is required to be in the organizational document of the surviving or converted entity. For example, a partnership converting to a corporation would include in its plan of conversion all the information required in articles of incorporation. Similarly, a corporation merging into an LLC would

246. OR. REV. STAT. §§ 60.472 (corporation conversion), 60.481 (corporation merger), 62.607 (cooperative conversion), 62.617 (cooperative merger), 63.470 (LLC conversion), 63.481 (LLC merger), 67.342 (partnership conversion), 67.360 (partnership merger), 70.505 (LP conversion), 70.525 (LP merger) (1999).

247. TEX. BUS. CORP. ACT ANN. art. 5.17.B (Vernon Supp. 2001) (allowing any foreign entity to convert to Texas corporation if conversion is permitted by or not inconsistent with laws of state or country under which it is organized).

248. E.g., OR. REV. STAT. § 67.370 (1997) (repealed 1999) (specifying that Oregon statute on merger is not exclusive, and that partnership or LP can convert or merge “in any other manner provided by law”).

249. The definition of each “business entity” includes an entity organized under Oregon statutes or “comparable law of another jurisdiction.” OR. REV. STAT. §§ 60.470(1) (corporation), 62.605(1) (cooperative), 63.467(1) (LLC), 67.340(1) (partnership), 70.500(1) (LP) (1999). Hence a foreign business entity can convert to or merge with a domestic one.

250. OR. REV. STAT. §§ 60.472 (corporation conversion), 60.481 (corporation merger), 62.607 (cooperative conversion), 62.617 (cooperative merger), 63.470 (LLC conversion), 63.481 (LLC merger), 67.342 (partnership conversion), 67.360 (partnership merger), 70.505 (LP conversion), 70.525 (LP merger) (1999).

251. Id. Owners of the pre-transaction entity will most often receive equity interests in the surviving or converted entity, but could also receive debt or other property.
include in its plan of merger any new or amended provision necessary in
the articles of formation.

The procedures for approving a plan are specified in the statute and
organizational document governing each business entity prior to the
transaction.\textsuperscript{252} When an LLC merges with a corporation, for example, the
members and managers of the LLC vote as provided in the LLC act and
operating agreement, while the shareholders and directors of the
corporation vote as provided in the corporation act and articles of
incorporation. For a partnership, procedures were modeled on previous
procedures for LLP action on plan of mergers.\textsuperscript{253}

However, private agreements may supercede or modify the statutory
norms, and in many cases negotiated clauses will be advisable. In a
limited partnership, for example, the statutory rule is that conversion or
merger must be approved by all the partners, unless the certificate of
limited partnership provides for a lesser vote.\textsuperscript{254} Such a unanimity
requirement provides a veto and hence the possibility of inordinate
bargaining power to a dissident owner.\textsuperscript{255} A well-counseled limited
partnership might well establish a lesser voting requirement in its
limited partnership agreement, superceding the statutory default rule.

The procedures that previously existed in the LLC act for approving
mergers\textsuperscript{256} were extended to conversions of LLCs and served as a model
for the corresponding provisions for the other entities. When statutes
were silent as to mergers, the procedures applicable to organic changes,
such as amending the organic document, served as a model for
approving conversion and mergers. For example, the Limited
Partnership Act had not previously allowed mergers, so its provisions
for amending the certificate of limited partnership were used instead.

After the plan is approved, the business files articles of conversion or
articles of merger with the Corporation Division of the Secretary of

\textsuperscript{252} One aspect of the Oregon statutory revisions was to move a couple provisions from one
business entity chapter of the statutes to another. \textit{OREGON REVISED STATUTES} sections 67.345 and
67.360 (1997), relating to conversion of partnership to or from limited partnership, were moved to
section 70.505 (1999) on limited partnerships.

\textsuperscript{253} \textit{OR. REV. STAT.} § 63.487 (1997) (amended 1999).

\textsuperscript{254} \textit{OR. REV. STAT.} §§ 70.510 (conversion), 70.530 (merger) (1999).

\textsuperscript{255} In corporation law, for example, the requirement has been a simple majority of all shares
entitled to vote, by voting group, unless the articles of incorporation provide otherwise. \textit{Id.}
§ 60.487(5),(6).

\textsuperscript{256} \textit{OR. REV. STAT.} § 63.481 (1997) (modified 1999).
Cross-Entity Conversion and Merger

State. This single document, accompanied by a copy of the plan, is all that is necessary. This aspect of the procedure was modeled on previous provisions in the Corporation Act and the LLC Act.

F. Creditors’ Rights and Owners’ Liabilities

Neither conversion nor merger allows an escape from debts or other forms of obligation to outside parties, such as those under building codes, labor law, environmental regulation, or other statutory, administrative, and regulatory regimes. Moreover, neither conversion nor merger causes a dissolution or other event requiring a winding up. Obligations incurred prior to the conversion or merger are treated differently from obligations incurred following the transaction.

1. Obligations Incurred Prior to Conversion or Merger

Because a business entity continues in existence without interruption by a conversion or merger, the post-transaction entity is liable for all obligations incurred by the entity in its earlier form. Owners of the entity who were previously liable for business debts remain liable for those pre-transaction debts, whether or not they are owners after the transaction. Thus, for example, general partners (in either a general partnership or a limited partnership) cannot evade their joint and several liabilities.

257. OR. REV. STAT. §§ 60.476(1) (corporate conversion), 60.494(1) (corporate merger), 62.611(1) (cooperative conversion), 62.621(1) (cooperative merger), 63.476(1) (LLC conversion), 63.494(1) (LLC merger), 67.346(1) (partnership conversion), 67.364(1) (partnership merger), 70.515(1) (LP conversion), 70.535(1) (LP merger) (1999).

258. OR. REV. STAT. § 60.494 (1997) (modified 1999) (addressing corporate articles of merger or share exchange).


260. Among the regulatory authorities consulted by the task force drafting Oregon’s cross-entity conversion and merger provisions were the State Department of Revenue, Department of Labor, and Workers’ Compensation Division.

261. For conversions, the relevant provisions are OREGON REVISED STATUTES sections 60.478(1)(c) (corporation), 62.613(1)(c) (cooperative), 63.479(1)(c) (LLC), 67.348(1)(c) (partnership), and 70.520(1)(c) (limited partnership) (1999). For mergers, the provisions are sections 60.497(1)(c) (corporation), 62.623(1)(c) (cooperative), 63.497(1)(c) (LLC), 67.365(1)(c) (partnership), and 70.540(1)(c) (limited partnership).

262. For conversions, the relevant provisions are OREGON REVISED STATUTES sections 60.478(1)(f)(A) (corporation), 62.613(1)(f)(A) (cooperative), 63.479(1)(f)(A) (LLC), 67.348(1)(f)(A) (partnership), and 70.520(1)(f)(A) (limited partnership). For mergers, the provisions are sections 60.497(1)(g)(A) (corporation), 62.623(1)(g)(A) (cooperative), 63.497(1)(g)(A) (LLC), 67.365(1)(g)(A) (partnership), and 70.540(1)(g)(A) (limited partnership).
liability for existing obligations of the business by arranging for the partnership to convert or merge into an entity providing limited liability, such as a corporation or LLC.

A further issue is responsibility of owners to make capital contributions. The Partnership Act and Limited Partnership Act include provisions not found in the acts governing the other business organizational forms. When a party to a conversion or merger is a partnership other than an LLP and does not satisfy the obligations it incurred prior to the conversion or merger, then the persons who previously were partners must contribute the amount necessary to satisfy the obligations as if the party was dissolved. The same rule applies in a limited partnership, but only to the general partners. The rule is designed to prevent an owner who had undertaken an obligation to contribute capital, but had not yet completed the contribution, from escaping the obligation in a conversion or merger in a manner that might damage the interest of outside parties.

Any litigation or proceeding pending against an entity or its owners at the time of a conversion or merger may be continued in the original name as if the transaction had not occurred. Alternately, the party bringing the action may substitute the new name but without any loss of rights.

2. Obligations Incurred After Conversion or Merger: the General Rule

For obligations incurred after the transaction, the law applicable to the converted or surviving entity and its owners controls in almost all cases. This rule is fair in most instances to creditors, who extend credit to the business with knowledge or at least the opportunity to learn in

263. Id. §§ 67.348(1)(g) (conversion), .365(1)(h) (merger).
264. Id. §§ 70.520(1)(g) (conversion), .540(1)(h) (merger).
265. For conversions, the relevant provisions are OREGON REVISED STATUTES sections 60.478(1)(d) (corporation), 62.613(1)(d) (cooperative), 63.479(1)(d) (LLC), 67.348(1)(d) (partnership), and 70.520(1)(d) (limited partnership). For mergers, the relevant provisions are sections 60.497(1)(d) (corporation), 62.623(1)(d) (cooperative), 63.497(1)(d) (LLC), 67.365(1)(d) (partnership), and 70.540(1)(d) (limited partnership).
266. For conversions, the relevant provisions are OREGON REVISED STATUTES sections 60.478(1)(f)(B) (corporation), 62.613(1)(f)(B) (cooperative), 63.479(1)(f)(B) (LLC), 67.348(1)(f)(B) (partnership), and 70.520(1)(f)(B) (limited partnership). For mergers, the provisions are sections 60.497(1)(g)(B) (corporation), 62.623(1)(g)(B) (cooperative), 63.497(1)(g)(B) (LLC), 67.365(1)(g)(B) (partnership), and 70.540(1)(g)(B) (limited partnership).
advance of the credit-worthiness and extent of liability of the entity and its owners. For example, a supplier who furnishes goods to a corporation relies on the credit of the corporation and not of the shareholders, whose protection from liability is well established. Conversion of the corporation into a limited liability company does not significantly change the risk to the creditor. The creditor will still have rights against the business entity and no rights (except in extraordinary circumstances, such as “piercing the veil”) against its owners (now members rather than shareholders).

A merger of a corporation into an LLC or LLP with a less solid financial status admittedly might increase the risk to creditors who extend credit after the merger. This type of risk, however, is no different from the risk following like-entity mergers (such as a corporation merging into a second corporation with less solid finances), which the law has traditionally countenanced. The risk is also comparable to the risk that a known customer will suffer business reverses making it less able or likely to pay bills—the sort of risk inherent in every credit decision.

In general, the law relies on the concept of constructive notice to creditors and potential creditors. A conversion or merger always requires a filing with the state, which is a matter of public record.267 Those who deal or consider dealing with the entity after the filing are presumed to know the contents of the filing. The same principle applies to real estate conveyances and liens, Uniform Commercial Code security interests268 and, notably, business-entity documents such as articles of incorporation.269 The legal fiction—or, more charitably, legal principle—of constructive notice admittedly can surprise morally innocent but unsophisticated parties, who are not aware of or proficient with the public filing system, but this is accepted as a necessary element of a workable commercial system.

267. Articles of conversion are required by Oregon Revised Statutes sections 60.476(1) (corporation), 62.611(1) (cooperative), 63.476 (LLC), 67.346 (partnership), and 70.515 (LP) (1999). Articles of merger are required by sections 60.494(1) (corporation), 62.621(1) (cooperative), 63.494 (LLC), 67.364 (partnership), and 70.535 (LP).

268. Or. Rev. Stat. § 79.3010 (1999) (providing priority for secured transactions that are properly filed and otherwise perfected, whether or not other claimants have actual knowledge of filing).

269. See, e.g., id. § 60.051(1) (specifying that corporate existence begins when articles of incorporation are filed by Secretary of State). With corporate existence comes the shield for shareholders from liability for acts and debts of the corporation, id. § 60.151(2), which applies whether or not an outside party has actual knowledge of the incorporation.
3. Special Provisions for Entry into Limited Liability Status

Special issues of fairness and notice are raised when an entity without limited liability transacts business with a creditor, converts or merges into an entity form providing limited liability for owners, and then engages in further transactions with the creditor. The creditor might well have no actual knowledge of the change and be significantly disadvantaged when the information becomes known. Although the concept of constructive notice could be applied, realists recognize the improbability of creditors actually gaining information from the public records in this circumstance.

For example, “A&B Plumbing,” a general partnership, may have purchased pipe and fittings for years as needed from a plumbing-supply company on account, paying at the end of each month. The supplier relied on the reputation and creditworthiness not only of A&B Plumbing but also of A and B individually, who were jointly and severally liable for partnership debts. A&B Plumbing then converts to “A&B Plumbing, LLC,” and continues to purchase pipe on account from its supplier, who does not know of the change. Realistically, such a conversion is most probable when the owners (but not creditors) recognize that business is declining and insolvency is a real danger. If the LLC later defaults, the supplier will then discover, to its surprise and detriment, that A and B are no longer personally liable.

The constructive notice concept would deny relief, on the premise that the supplier could have avoided surprise by checking with the state every time A or B ordered a new faucet or pipe fitting on credit, to assure that the company that had been a general partnership for years is still a general partnership. Plainly, however, the burden to continually search is too expensive, wasteful, and impractical for most businesses to even consider. Constructive notice has reached its limit of plausibility.

An alternative might be to impose on the converting or merging entity the obligation to provide actual notice to parties with whom it has done business in the past. Placing the duty on the party seeking to benefit from the conversion or merger is more practical, economical, and fair than imposing a duty of search on outside parties. Yet such a rule is also subject to objection. No general rule applies in other contexts for businesses to notify those with whom they have transacted business in the past of mergers or other developments which might negatively affect the likelihood of payment of new obligations. Also, such a rule of actual notice could generate litigation by creditors asserting that they did not
receive actual notice and hence are entitled to pursue former partners despite a conversion or merger long ago.

Provisions responding to issues such as these were included in RUPA and Oregon’s partnership act prior to the 1999 amendments. A partner in a general partnership who became a limited partner in a limited partnership remained liable as a general partner for obligations incurred before the conversion,270 which is not surprising. However, the partner also was liable as a general partner for obligations incurred within ninety days after the conversion to an outside party who reasonably believed when entering the transaction that the person was a general partner.271 The Task Force drafting the 1999 Oregon legislation was concerned that the “reasonable belief” standard did not provide the partner a conclusive means of terminating exposure to liability early and that ninety days seemed unrealistically brief.

A separate RUPA section relating to dissociation reflected a similar concern for post-conversion creditors without actual knowledge of a conversion, but a somewhat different approach. If the dissociation had not been advertised in a newspaper, a dissociated partner retained apparent agency authority for a period after dissociation to bind a partnership to a third party who reasonably believed that the person was still a partner and did not have notice.272 The period was two years in the Uniform Act, and six months in the Oregon statutes.273

The Oregon cross-entity conversion and merger legislation in 1999 established a one-year window of exposure to liability,274 seeking to balance the competing interests and perspectives of the business entity owners and the third parties who deal with the entity without actual knowledge of the change. When a general or limited partnership converts or merges into a form of business organization that normally shields owners from liability, a former general partner continues to be

---

271. OR. REV. STAT. § 67.345(5) (1997) (repealed 1999); REV. UNIF. PARTNERSHIP ACT § 902(e). Because this provision applies only to limited partnerships, it belongs more logically in the Limited Partnership Act, not the Partnership Act. Oregon repealed the provision in 1999, replacing it with more encompassing legislation, parallel in the two chapters.
272. OR. REV. STAT. § 67.255 (1999); REV. UNIF. PARTNERSHIP ACT § 702.
273. OR. REV. STAT. § 67.255; REV. UNIF. PARTNERSHIP ACT § 702.
274. OR. REV. STAT. §§ 60.478(1)(h) (corporation conversion), 60.497(1)(h) (corporation merger), 62.613(1)(g) (cooperative conversion), 62.623(1)(g) (cooperative merger), 63.479(1)(g) (LLC conversion), 63.497(1)(h) (LLC merger), 67.348(1)(h) (partnership conversion), 67.365(1)(h) (partnership merger), 70.520(1)(h) (LP conversion), 70.540(1)(i) (LP merger) (1999).
personally liable on obligations incurred during the one year following
the conversion or merger, if the outside party reasonably believes that
the owner would be personally liable and had not received notice of the
conversion or merger.\footnote{275}

This provision was drafted by the Task Force, not drawn directly
from any uniform act or other state’s statute. It provides the former
general partner a conclusive means of assuring freedom from personal
liability for post-conversion transactions by sending notice to outside
parties.\footnote{276} It imposes a burden of notice (or, perhaps more accurately, the
risk of failure to notify) on the party with the information, in contrast to
the constructive notice principle, which imposes a burden of searching
on the party without the information. The reasonable expectations of the
outside party, when entering into post-merger or post-conversion trans-
actions, are protected, but only for one year. That period was selected
arbitrarily but was considered sufficient for creditors to become aware of
the current form of business entity. After the year, the constructive
notice principle operates, conclusively presuming notice based on the
public filings.\footnote{277}

The form of actual notice is not specified in the statute, though it must
be more than merely filing with the state.\footnote{278} Adding the new entity type
to the business name, by appending such terms as “corporation,”
“incorporated,” “limited liability company,” “limited partnership,”
“limited liability partnership,” or an abbreviation, should be sufficient if

\footnote{275. For conversion, the provisions on liability of former partners are OREGON REVISED
STATUTES sections 60.478(1)(g) (corporation), 62.613(1)(g) (cooperative), 63.479(1)(g) (LLC),
67.348(1)(h) (partnership), and 70.520(1)(h) (LP). For merger, the provisions are sections
60.497(1)(h) (corporation), 62.623(1)(h) (cooperative), 63.497(1)(h) (LLC), 67.365(1)(i) (partner-
ship), and 70.540(1)(i) (LP).

276. OR. REV. STAT. §§ 60.478(1)(g) (corporation conversion), 62.613(1)(g) (cooperative
conversion), 63.479(1)(g) (LLC conversion), 67.348(1)(h) (partnership conversion), 70.520(1)(h)
(LP conversion); id. §§ 60.497(1)(h) (corporation merger), 62.623(1)(h) (cooperative merger),
63.497(1)(h) (LLC merger), 67.365(1)(i) (partnership merger), 70.540(1)(i) (LP merger).

277. OR. REV. STAT. §§ 60.478(1)(g) (corporation conversion), 62.613(1)(g) (cooperative
conversion), 63.479(1)(g) (LLC conversion), 67.348(1)(h) (partnership conversion), 70.520(1)(h)
(LP conversion); id. §§ 60.497(1)(h) (corporation merger), 62.623(1)(h) (cooperative merger),
63.497(1)(h) (LLC merger), 67.365(1)(i) (partnership merger), 70.540(1)(i) (LP merger).

278. Each provision imposes liability unless “the other party” to a business transaction “received
notice” of the conversion or merger. OR. REV. STAT. §§ 60.478(1)(g) (corporation conversion),
62.613(1)(g) (cooperative conversion), 63.479(1)(g) (LLC conversion), 67.348(1)(h) (partnership
conversion), 70.520(1)(h) (LP conversion); id. §§ 60.497(1)(h) (corporation merger), 62.623(1)(h)
(cooperative merger), 63.497(1)(h) (LLC merger), 67.365(1)(i) (partnership merger), 70.540(1)(i)
(LP merger). The filing of a document with the state by a converting or merging entity does not,
without more, constitute receipt of notice by a party to a business transaction.
the creditor would likely see it. Factors would include the size and prominence of the new name on business cards, order forms, vehicles, or other contexts to determine when the creditor should be charged with notice. To avoid these issues, the principals of the converting or merging entity are best advised to provide the most direct, most specific notice possible: a letter addressed to the creditor explicitly stating the transaction.

In the A&B Plumbing example, A and B will be individually liable as general partners for orders placed with the supplier before the conversion, whether or not notice is given. They will also be individually liable for orders placed within one year after conversion to a limited liability company, unless they give adequate actual notice of the conversion to supplier. As to orders placed more than a year after conversion, the LLC statute applies, whether or not actual notice is given, protecting members from personal liability for business debts.

G. Owners’ Rights

Rights of owners are determined in the plan of conversion or merger and by the statutes, common law, and private agreements (including the organizational documents) governing the entity before the transaction.279 The results vary, reflecting the variations in the business forms.

Shareholders have dissenters’ and appraisal rights, except in corporations with publicly traded shares.280 The same substantive rights and procedures that traditionally applied to mergers between or among corporations281 have been extended to mergers between or among a corporation and a non-corporate entity, and to conversions of a

---

279. For conversion, the provisions applicable to owners’ rights are OREGON REVISED STATUTES sections 60.478(2) (corporation), 62.613(2) (cooperative), 63.479(2) (LLC), 67.348(2) (partnership), and 70.520(2) (LP). For merger, the provisions are sections 60.497(2) (corporation), 62.623(2) (cooperative), 63.497(2) (LLC), 67.365(2) (partnership), and 70.540(2) (LP).

280. Id. § 60.554; REV. MODEL BUS. CORP. ACT § 3.02 (1996). Examples of state corporation acts that deny dissenter and appraisal rights to shareholders of publicly traded corporations include DELAWARE CODE ANNOTATED title 8, section 262(b)(1) (Supp. 2000) and 15 PENNSYLVANIA CONSOLIDATED STATUTES ANNOTATED section 1571(b)(1) (West 1995). One view, not yet adopted by any state, is that this difference between public and nonpublic corporations will be resolved by eliminating dissenter and appraisal rights for all corporations. Clark, supra note 40, at 177.

281. OR. REV. STAT. §§ 60.551–594 (regarding dissenters rights and judicial appraisal of shares).
corporation to another entity.\textsuperscript{282} All such transactions significantly change the nature of the shareholders’ investment.

In cooperatives, shareholders have only what rights may be provided in the articles of conversion and merger. Dissenters’ or appraisal rights are not provided,\textsuperscript{283} consistent with the general pattern in cooperative law that the entity is for the mutual benefit of members, not primarily to generate financial return to investors.\textsuperscript{284}

Partners who disapprove (when a decision may be effectively made without unanimity) are deemed to have dissociated effective immediately before the conversion or merger, unless they notify the partnership within sixty days of a contrary intent, and the dissociation is not deemed wrongful.\textsuperscript{285} In this situation, as in other situations in which partners dissociates rightfully and the partnership continues the business, the entity must purchase those persons’ interests and indemnify them against all liabilities of the partnership.\textsuperscript{286}

The right to dissociate and to be bought out is granted because a conversion drastically changes the relationship of the partner to the entity and to creditors in ways that are not entirely favorable. For example, a partner who becomes a shareholder gains the shield of limited liability from creditors but loses managerial powers and the protection of some fiduciary duties owed to partners but not to shareholders. A partner confronted with such a fundamental change should be allowed to cash out.

In a limited partnership, the limited partners are treated in a manner comparable to partners in a general partnership. A limited partner who did not vote in favor of a conversion or merger is deemed to have withdrawn immediately upon conversion, unless the partner gives notice to the contrary and the withdrawal is not deemed wrongful.\textsuperscript{287} Unless

\textsuperscript{282} Id. §§ 60.478(2)(a) (establishing rights of shareholders in conversion), .554(1)(f) (listing conversion as one event triggering right to dissent).

\textsuperscript{283} Id. § 62.623(2) (providing that owners have only rights provided in plan of merger).

\textsuperscript{284} Simon, supra note 34, § 30.1. Dissenters rights in cooperatives are discussed in David C. Crago, Cooperative Dissent: Dissenting Shareholder Rights in Agricultural Cooperatives, 27 IND. L. REV. 495 (1994).


\textsuperscript{286} OR. REV. STAT. § 67.250 (1999); REV. UNIF. PARTNERSHIP ACT § 701 (1994).

\textsuperscript{287} OR. REV. STAT. §§ 70.520(2)(a) (conversion), .540(2)(a) (merger) (1999). In addition, a limited partner may withdraw at the time or occurrence of an event specified in the partnership
otherwise agreed, withdrawing partners are entitled to receive the fair value of their interests. 288

As to general partners in limited partnerships, no statutory sections specifically address rights in the conversion or merger situation. General partners have only whatever rights are provided in the partnership agreement and plan of conversion or merger, plus the general statutory provisions. 289 Usually, general partners may withdraw at any time by giving written notice, subject to liability for damages if withdrawal violates the partnership agreement, and then be paid the fair value of their interest. 290 No statutory provision indicates that withdrawal of a general partner in a conversion or merger situation is not wrongful.

Members of limited liability companies, under the laws of Oregon and many other states, have only the rights specified in the LLC agreement, plus a very restricted power to withdraw. 291 A member may voluntarily withdraw as provided in the articles of organization or upon six month’s notice unless the articles expressly deny that power, but the member is exposed to liability if the withdrawal breaches any provision of the articles of organization or an operating agreement. 292 Dissenters’ rights are not provided. 293

agreement, or if the agreement is silent, upon six month’s prior written notice. Or. Rev. Stat. § 70.255(2) (1999); RULPA § 603 (amended 1985), 6A U.L.A. 217–18 (1995). Some statutes have modified the rules to authorize partnership agreement clauses preventing withdrawal of a limited partner prior to dissolution and winding up. See Keatinge, supra note 24, at 58 (citing Delaware, Maine, Nebraska, Pennsylvania, Rhode Island, and Tennessee statutes).


In addition to the rights formally granted by statute or private agreement, owners of businesses are typically protected from oppression or other breach of fiduciary duty by the controlling owners. The extent and definition of fiduciary duty varies among the different organizational forms, and are undergoing development within many of the forms. For example, partners were traditionally protected and subjected to a strict fiduciary duty of loyalty, but the trend is toward freedom of contract with fiduciary duties that are more narrowly defined and often subject to waiver by contract. In an LLC, fiduciary duties usually apply at least to the managers in manager-managed companies, but may be modified by agreement.

Whether the disparities in protection of owners of the different entities is justified is open to debate, but they are a matter of preexisting law. The 1999 Oregon legislation extends the provisions in preexisting law regarding like-entity mergers to conversions and cross-entity mergers, but does not otherwise change the rules.

H. Property

The converted or surviving business entity becomes the owner, by operation of law, of all of the property (real or personal, tangible or intangible) of the converted or disappearing entity. This principle that the owner of the property changes without a transfer of the property may provide to the contrary. Id. (citing VA. CODE ANN. §§ 13.1–1032 (Michie Supp. 1997); WASH. REV. CODE § 25.15.130(3) (2000)).

293. Part of the explanation may be that the uncertainty and delay occasioned by dissenters rights were not necessary or desirable in light of fiduciary-duty protection and the members’ right to withdraw or dissolve the LLC, thereby obtaining the fair value of their interests. Clark, supra note 40, at 168 (quoting and discussing PROTOTYPE LTD. LIAB. CO. ACT §1202 cmt. (1992)). The rule may also be based on federal tax considerations no longer applicable in light of the “check-the-box” regulations. Id.


295. OREGON REVISED STATUTES section 63.155 (1999) establishes duties of care and loyalty, with provisions for private agreements reducing the duties or defining activities that do not violate them. Section 63.160 provides for indemnification or exculpation of a member from liability.

296. For conversion, the provisions applicable to change of ownership of property are OREGON REVISED STATUTES sections 60.478(1)(b) (corporation), 62.613(1)(b) (cooperative), 63.479(1)(b) (LLC), 67.348(1)(b) (partnership), and 70.520(1)(b) (LP). For merger, the provisions are sections 60.497(1)(b) (corporation), 62.623(1)(b) (cooperative), 63.497(1)(b) (LLC), 67.365(1)(b) (partnership), and 70.540(1)(b) (LP) (1999).
have metaphysical connotations, but also has substantial practical import.

One major consequence is that no deeds, bills of sale, assignments, or other documents of conveyance are necessary or appropriate, thus significantly reducing the cost of documentation. Taxes on transfer of property should not apply, because the conversion or merger of the owner of the property is not deemed to be a transfer.

Nevertheless, for real estate, some documentation is advisable because the lack of a deed in the name of the converted or surviving entity may create difficulties in the future, especially when the entity seeks to sell the property to a third party.297 Accordingly, although the statute does not mention or require it, the business is well advised to prepare a memorandum of the conversion and merger, establishing the identity of the owner before the transaction and the identity afterwards, perhaps attaching the plan of conversion or merger. The memorandum should be filed with the recorder of deeds of every county in which the business owned an interest in real estate, serving to prevent misunderstanding by others as to ownership and expediting future conveyances and title insurance.

Title insurance on real property typically applies only to the purchaser, not to any transferee of the real estate unless the title company specifically agrees to such an extension of coverage.298 Nevertheless, the original title policy will operate in favor of a converted or surviving entity because of statutory provisions specifying that entity existence continues and all property, which includes rights under the insurance policy, vests in the converted or surviving entity.299 The same principles also apply to other forms of insurance.300

297. For example, if a corporation owning a building converts to an LLC or merges with an LLC, the LLC becomes the owner of the building. Recorded title to the building remains in the corporate name, however. A potential purchaser from the LLC, or a potential purchaser’s title insurance company, might demand proof that the LLC is the owner.

298. Palomar, supra note 128, at 621.

299. For conversion, the provisions are Oregon Revised Statutes sections 60.478(1)(a)–(b) (corporation), 62.613(1)(a)–(b) (cooperative), 63.479(1)(a)–(b) (LLC), 67.348(1)(a)–(b) (partnership), and 70.520(1)(a)–(b) (LP). For merger, the provisions are sections 60.497(1)(a)–(b) (corporation), 62.623(1)(a)–(b) (cooperative), 63.497(1)(a)–(b) (LLC), 67.365(1)(a)–(b) (partnership), and 70.540(1)(a)–(b) (LP).

300. See, e.g., Imperial Enters. Inc. v. Fireman’s Fund Ins. Co., 535 F.2d 287, 292–93 (5th Cir. 1976) (finding that when insured party merged, surviving entity became insured by operation of law, despite non-assignment clause in insurance policy).
Long-term commercial leases and franchise agreements routinely have non-assignment clauses, and mortgages have due-on-sale clauses, all prohibiting transfer without consent. The landlord, franchiser, or mortgagee sometimes denies consent to a transfer, or conditions it on a payment or a renegotiation of terms. However, because a conversion or merger of the business entity owning the interest of the tenant, franchisee or mortgagor is by operation of law, the clauses are normally not triggered. To overcome this rule requires careful drafting of contractual clauses restricting transfer, specifying that merger or conversion is an event of default.

The statute specifies that title vests in the converted or surviving business entity “without reversion or impairment.” The application of that phrase includes title in a defeasible estate: fee simple determinable, fee simple subject to condition subsequent, or fee simple subject to executory interest. For example, the fee simple determinable resulting from a grant “to Alpha Corporation for so long as it operates a railroad on the land” does not terminate if Alpha Corporation converts to Alpha LLC or merges with Beta LP, because the converted or surviving entity is deemed the same entity as Alpha Corporation.

301. See, e.g., Dodier Realty & Inv. Co. v. St. Louis Nat’l Baseball Club, Inc., 238 S.W.2d 321, 325 (Mo. 1951) (finding that merger of lessee corporation with another corporation did not breach contractual prohibition on assignments because transfer was by operation of law); accord Segal v. Greater Valley Terminal Corp., 199 A.2d 48, 51 (N.J. 1964). See generally Zitter, supra note 125. Patents, however, may not be treated similarly. Some cases find that the merger of patent licenses with other corporations violates anti-assignment clauses. PPG Indus., Inc. v. Guardian Indus. Corp., 597 F.2d 1090, 1095 (6th Cir. 1979); Unarco Indus., Inc. v. Kelley Co., 465 F.2d 1303, 1306 (7th Cir. 1972).

302. Contractual provisions specifying that certain assets such as licenses and permits are non-assignable and nontransferable, by merger or otherwise, are valid. Citizens Bank & Trust Co. v. Barlow Corp. of Maryland, 456 A.2d 1285 (Md. 1983); Pac. First Bank v. New Morgan Park, 876 P.2d 761 (Or. 1994); O.K. Delivery Sys., Inc. v. Haley, 487 P.2d 1391 (Or. App. 1971); Culpepper, supra note 10, § 39A-64 (discussing LLC Act provisions that preceded 1999 amendments but were not significantly changed by them).

303. For conversion, the provisions applicable to change of ownership of property are OREGON REVISED STATUTES sections 60.478(1)(b) (corporation), 62.613(1)(b) (cooperative), 63.479(1)(b) (LLC), 67.348(1)(b) (partnership), and 70.520(1)(b) (LP). For merger, the provisions are sections 60.497(1)(b) (corporation), 62.623(1)(b) (cooperative), 63.497(1)(b) (LLC), 67.365(1)(b) (partnership), and 70.540(1)(b) (LP).


305. The estate is a fee simple determinable because it is capable of lasting forever, but subject to automatic early termination upon occurrence of a stated condition. See generally id. § 2.4.
Cross-Entity Conversion and Merger

Similarly, if title is held subject to a valid restraint on alienation, a conversion or merger will not be a violation because it is not deemed a transfer of property.306 For example, if a grantor conveyed “the family farm to Smith Limited Partnership, which may never convey to anyone else,” a subsequent conversion of the limited partnership to a corporation or a merger with another business entity is not a problem. The question of whether the restraint is valid should not even arise, because conversion or merger is, by statute, deemed not to be a conveyance.

I. Documentation and Public Filings

Each transaction requires only one filing with the state—articles of conversion or articles of merger.307 The plan that must accompany that document includes any additional information that is required to organize the converted or surviving entity.308 Consequently, there is no need to file separate documents, such as articles of incorporation, articles of organization, or amended articles.

For example, a limited partnership that opts to be a corporation files articles of conversion with a plan of conversion that contains all the items statutorily required in all articles of incorporation, but no separate articles of incorporation.309 Assumed business names of the converting or merging entities automatically become assumed business names of the converted or surviving entities.310

Consistent with this approach, the definitions of the organic documents were amended. For example, the definition of “articles of

306. Because of the public policy against restraints on alienation, courts construe anti-assignment clauses narrowly and hold that transfers by operation of law do not violate them. See, e.g., Segal, 199 A.2d at 51.

307. Articles of conversion are required by OREGON REVISED STATUTES sections 60.476(1) (corporation), 62.611(1) (cooperative), 63.476 (LLC), 67.346 (partnership), and 70.515 (LP) (1999). Articles of merger are required by sections 60.494(1) (corporation), 62.621(1) (cooperative), 63.494 (LLC), 67.364 (partnership), and 70.535 (LP).

308. Id. §§ 60.472(2)(e) (corporate conversion), 60.481 (corporation merger), 62.607(2)(e) (cooperative conversion), 62.617(2)(e) (cooperative merger), 63.470(2)(e) (LLC conversion), 63.481(2)(e) (LLC merger), 67.342(2)(e) (partnership conversion), 67.360(2)(e) (partnership merger), 70.505(2)(e) (LP conversion), 70.525(2)(e) (LP merger).

309. OR. REV. STAT. § 60.051.

310. Id. §§ 60.478(1)(h) (corporate conversion), 60.497(1)(i) (corporate merger), 62.613(1)(h) (cooperative conversion), 62.623(1)(i) (cooperative merger), 63.479(1)(h) (LLC conversion), 63.497(1)(i) (LLC merger), 67.348(1)(i) (partnership conversion), 67.365(3) (partnership merger), 70.520(3) (LP conversion), 70.540(1)(j) (LP merger).
incorporation” was changed to include articles of conversion and articles of merger,\textsuperscript{311} to coordinate with the preexisting provision that “corporate existence begins when the articles of incorporation are filed by the Secretary of State.”\textsuperscript{312} Comparable changes were made to the definitions of articles for a cooperative,\textsuperscript{313} articles of organization of an LLC,\textsuperscript{314} and certificate of limited partnership.\textsuperscript{315} No comparable provision applies to partnership, which can be created without a filing with the state.

The system permits tracking of entities that change form, an essential element for the protection of creditors, claimants, and parties considering transacting business with a business.\textsuperscript{316} A creditor of an LP, for example, will be able to determine from the Secretary of State’s records that the LP converted to an LLC, as well as the current address and registered agent of the LP. Moreover, the public can search the Secretary of State’s records for earlier organizational documents.

The filing requirement is somewhat different for general partnerships which, unlike all other business organization forms, can be created without a filing (or even a writing), unless it registers as an LLP. When a non-LLP partnership merges with another non-LLP partnership, no filing is required.\textsuperscript{317} In all other merger or conversion transactions, however, filings are required to update the previous filing or to establish the new organizational form.\textsuperscript{318}

\begin{itemize}
\item \textsuperscript{311} Id. § 60.001(2).
\item \textsuperscript{312} Id. § 60.051(1).
\item \textsuperscript{313} Id. § 62.015(2).
\item \textsuperscript{314} Id. § 63.001(2).
\item \textsuperscript{315} Id. § 70.005(1). The 1999 legislation repealed Oregon Revised Statutes section 67.365(5) (1997) (repealed 1999), which had provided for cancellation of an LLP’s registration upon merger, followed thirty days later by filing of a cancellation notice or, if the LLP survived, filing of an amendment to registration thirty days after the merger. This two-step process was unnecessary and problematic for the Corporation Division, and therefore was replaced by the single filing of articles of merger. Also repealed was Oregon Revised Statutes section 67.365(6) (1997) (repealed 1999), which required an LP merging out of existence to cancel its certificate of LP. Matters relating to LPs were moved to the LP act (and, again, the rule only requires filing articles of merger).
\item \textsuperscript{316} If tracking were not possible, a business entity could convert or merge into another form, disappearing from the Corporation Division’s records under the original name and making it difficult or impossible for a creditor to find the entity and its owners.
\item \textsuperscript{317} After approval of a plan of merger, the surviving entity must deliver articles of merger to the Secretary of State, “except that no filing is required if all of the parties to the merger are partnerships that have not registered as limited liability partnerships.” Or. Rev. Stat. § 67.364(1) (1999).
\item \textsuperscript{318} For example, an LLC converting to or merging with a general partnership will have previously filed articles of organization under Oregon Revised Statutes section 63.044. In this case,
Cross-Entity Conversion and Merger

Unlike other conversion situations, the Secretary of State will not be able to cross-reference a partnership’s articles of conversion to a previous filing. For that reason, a partnership must include in its articles of conversion the names of at least two of the partners, providing creditors information needed to locate the partnership. Comparable disclosures are not present or necessary in parallel provisions in the corporation, LLC, and LP acts because those business entities file annual reports disclosing addresses. The accompanying plan of conversion will contain all of the information required in the organizational document, such as articles of incorporation, for the resulting entity. In the opposite direction of an entity converting to a general partnership, which does not normally require a filing, the articles of conversion will serve as public notice of exit from the previous form of doing business, which did require a filing.

J. Income Tax Consequences

The income tax effects of a conversion or merger are separate from the business organization aspects discussed above, and beyond the scope of this Article. The tax consequences depend largely on whether the entities before and after the transaction are classified as corporations or as partnerships for tax purposes.

In general, merger or conversion of a partnership into or with another partnership, or into or with a corporation, does not result in tax. Merger or conversion of a corporation into or with a partnership does result in tax. However, exceptions to these generalizations exist, and secondary consequences such as changes in basis can be important. The IRS can recharacterize a conversion or merger as being a liquidation, generating gain or loss recognition, but has indicated that it will generally not do so.

the filing should be updated. Conversely, a general partnership converting to an LLC will need articles of organization—a purpose served by the articles of conversion. \textit{Id.} §§ 67.342(1)(e), 63.044.

319. \textit{Id.} § 67.346(1).


321. STEVEN L. CHRISTENSEN, MERGERS AND CONVERSIONS OF OREGON BUSINESS ENTITIES UNDER SB 1452 (Multnomah Bar Association Continuing Legal Education Series, June 15, 2000).


323. CHRISTENSEN, supra note 321, at 9 (citing Rev. Rul. 84-111, 1984-2 C.B. 88). A number of sources discuss the tax treatment of mergers and conversions in detail. See, \textit{e.g.}, Kevin D.
K. Early Experience with Conversion and Merger of Disparate Business Entities

In the first year after the cross-entity amendments became effective on January 1, 2000, eighty-two conversion documents were filed. The Business Registry database of the Corporation Division of the Oregon Secretary of State reported the following conversions among domestic entities:

- 26 General partnerships converted to LLCs
- 21 LLCs converted to corporations
- 18 Corporations converted to LLCs
- 3 LLCs converted to PCs
- 3 PCs converted to corporations
- 2 LLCs converted to LPs
- 2 LLPs converted to corporations
- 1 LP converted to an LLC
- 1 Cooperative converted to a corporation

In addition, one foreign corporation (from Washington) converted to an Oregon corporation and four Oregon LLCs converted to foreign LLCs. These statistics suggest that the greatest use of the conversion option is for general partnerships to become LLCs, and for corporations and LLCs to change to the other form. Mergers among disparate entities were far fewer and displayed no clear trend.

Anderson, Slicing, Dicing and Combining Partnerships: A Look at the Proposed Regulations on Partnership Mergers and Divisions, Tax Management Memorandum (Mar. 27, 2000); Sheldon Banoff, Mr. Popeil Gets “Reel” About Conversions of Legal Entities: The Pocket Fisherman Flycasts for “Form” but Snags on Substance, 75 TAXES 887 (1997); Christensen, supra note 321, at 14; Steven Frost, The Federal Tax Consequences of Business Entity Conversions, 26 J. REAL EST. TAX’N 83 (1999).

324. E-mail from Twila K. Coakley, Oregon Secretary of State’s Office, to Janet M. Sullivan, director of the Corporation Division, and to the author (Mar. 13, 2000) (on file with author).
325. Id.
326. Id.
327. Id.
328. E-mail from Twila K. Coakley, Oregon Secretary of State’s Office, to the author (Jan. 29, 2001) (on file with author). Records of five cross-entity mergers included two corporations merging with LLCs, and one merger each of a corporation with an LLC, an LLC with an LLP, and an LP with an LLC. Id. However, the records were not complete, omitting several additional cross-entity mergers. Id. Still, the number of cross-entity mergers appears far below the number of cross-entity conversions.
VI. CONCLUSION

Legislation allowing business entities to merge with entities of disparate organizational form or to convert into an alternative business form, is logical, workable and, in most states, overdue. As has been demonstrated, such legislation provides many benefits and violates no policy.

The development of new forms of business entities in recent years, especially the limited liability company and limited liability partnership, presents new opportunities to businesses and their counsel. Many existing businesses, as well as newly organizing firms, will see advantages in converting to the new forms or in merging with other firms that may well have a different organization form.

The absence of legislation does not prevent cross-entity conversion or merger, but merely forces businesses to engage in burdensome, expensive, and complex transactions to accomplish the equivalent goals. Moreover, given the federal system, inadequacy of laws in one state leads businesses to organize or reorganize in whichever state offers the most attractive legal regime, and then to conduct business as a foreign entity in whichever state is most economically attractive.

Current legislation in many states, especially those that have adopted the Revised Uniform Partnership Act, already authorizes cross-entity conversion and merger, thereby conceding the legitimacy of the concept. In particular, the RUPA allows entities that impose personal liability on some or all of the owners (general partners in general partnerships or limited partnerships) to transform themselves into entities shielding all owners from liability. That those same states do not freely allow any entity to convert or merge into any other entity is merely anomalous.

Drafting issues are numerous and involved, but not insuperable. Oregon presents an example of comprehensive, parallel, and tailored legislation. The statutory provisions provide flexibility and simplicity, protect the interests of creditors and owners, and maintain a public record of the transactions. Other states should emulate Oregon’s step forward in this field.
### APPENDIX

OREGON REVISED STATUTES (1999) on Business Entity Conversion and Merger

<table>
<thead>
<tr>
<th>Subject</th>
<th>Corp. Ch. 60</th>
<th>Co-op. Ch. 62</th>
<th>LLC Ch. 63</th>
<th>Part. Ch. 67</th>
<th>LP Ch. 70</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Definitions</strong></td>
<td>60.470</td>
<td>62.605</td>
<td>63.467</td>
<td>67.340</td>
<td>70.500</td>
</tr>
<tr>
<td><strong>Conversion</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) both ways</td>
<td>60.472</td>
<td>62.607</td>
<td>63.470</td>
<td>67.342</td>
<td>70.505</td>
</tr>
<tr>
<td>(2) plan must set forth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3) plan may set forth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Action on plan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) approval</td>
<td>60.474</td>
<td>62.609</td>
<td>63.473</td>
<td>67.344</td>
<td>70.510</td>
</tr>
<tr>
<td>(2) abandonment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Articles of conversion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) filing</td>
<td>60.476</td>
<td>62.611</td>
<td>63.476</td>
<td>67.346</td>
<td>70.515</td>
</tr>
<tr>
<td>(2) effective date</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effect of conversion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) existence, property, obligations, actions, liabilities of owners, amendments, ABN</td>
<td>60.478</td>
<td>62.613</td>
<td>63.479</td>
<td>67.348</td>
<td>70.520</td>
</tr>
<tr>
<td>(2) owners’ rights</td>
<td>60.554</td>
<td></td>
<td></td>
<td>67.255</td>
<td>67.260</td>
</tr>
<tr>
<td><strong>Merger</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) both ways</td>
<td>60.481</td>
<td>62.617</td>
<td>63.481</td>
<td>67.360</td>
<td>70.525</td>
</tr>
<tr>
<td>(2) plan must set forth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3) plan may set forth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Action on plan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1) approval</td>
<td>60.487</td>
<td>62.619</td>
<td>63.487</td>
<td>67.362</td>
<td>70.530</td>
</tr>
<tr>
<td>(2) abandonment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Cross-Entity Conversion and Merger

<table>
<thead>
<tr>
<th>Subject</th>
<th>Corp. Ch. 60</th>
<th>Co-op. Ch. 62</th>
<th>LLC Ch. 63</th>
<th>Part. Ch. 67</th>
<th>LP Ch. 70</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect of merger</td>
<td>60.497</td>
<td>62.623</td>
<td>63.497</td>
<td>67.365</td>
<td>70.540</td>
</tr>
<tr>
<td>(1) existence, property, obligations, actions, liabilities of owners, amendments, ABN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2) owners’ rights</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

411
DNA TYPING: EMERGING OR NEGLECTED ISSUES

Edward J. Imwinkelried* & D.H. Kaye**

Abstract: DNA typing has had a major impact on the criminal justice system. There are hundreds of opinions and thousands of cases dealing with DNA evidence. Yet, at virtually every stage of the process, there are important issues that are just emerging or that have been neglected.

At the investigative stage, courts have barely begun to focus on the legal limitations on the power of the police to obtain samples directly from suspects and to use the data from DNA samples in various ways. Issues such as the propriety of “DNA dragnets” (in which large numbers of individuals in a geographic area are asked to provide samples voluntarily), the validity of court orders for samples based on a lesser standard than probable cause, and the permissibility of collecting DNA “abandoned” in public places are being litigated for the first time. Using crime-scene samples to infer racial or ethnic characteristics is emerging as a distinct possibility. Then there are the more than 282 million specimens of human biological material stored by private and public agencies in the United States; in some situations, police may well turn to some of these repositories to obtain samples. There is little or no case law analyzing the constitutional restrictions on these investigative practices.

After the filing of charges, an accused sometimes moves to dismiss on the ground of the expiration of the statute of limitations. However, there is a movement to carve out a DNA exception to the statute of limitations in cases in which DNA evidence permits the identification of the perpetrator after the expiration of the normal period of limitations. The argument is that the legislative purpose of the statute is to prevent the maintenance of prosecutions based on stale, unreliable evidence but that DNA evidence is so reliable that its availability should lift the bar of the statute. However, little attention has been given to the difficulties inherent in drafting such a legislative exception that will not be overinclusive.

At the trial stage, in a growing number of cases, after the defense attacks the weight of the government’s DNA evidence, prosecutors are commenting to the jury that the defense has requested an opportunity to retest the DNA. Do such comments run afoul of the Fifth or Sixth Amendment?

The purpose of this Article is to identify and analyze such emerging issues. If the criminal justice system is to realize the full potential of DNA technology while maintaining the essential fairness of the system, it must come to grips with these issues in short order.

Fifteen years ago, deoxyribonucleic acid (DNA) analysis began to be applied to law enforcement.1 Before long, its suitability for the

---

*Professor, University of California at Davis School of Law; J.D., B.A., University of San Francisco.

**Regents’ Professor, Arizona State University College of Law; Fellow, Center for the Study of Law, Science, and Technology; J.D., Yale Law School; M.A., Harvard University; B.S., Massachusetts Institute of Technology.

1. The earliest instance of DNA analysis for legal purposes is reported in Alec J. Jeffreys et al., Positive Identification of an Immigration Test-Case Using Human DNA Fingerprints, 317 NATURE 818 (1985) (applying the multilocus probes described in Alec J. Jeffreys et al., Individual-Specific
courtroom was bitterly contested. Significant questions were raised in the popular press, books from scientific publishers, law reviews, and, of course, in trial and appellate courts. Today, the controversy over the scientific validity of forensic DNA testing has largely dissipated, but more subtle issues of criminal procedure and evidence often arise when DNA is employed in the investigations and trials. Unlike the question of the scientific validity of the usual methods of forensic DNA analysis, many of these new matters have yet to be extensively litigated. This Article identifies some of these emerging or neglected issues.


2. See, e.g., Gina Kolata, Some Scientists Doubt the Value of “Genetic Fingerprint” Evidence, N.Y. TIMES, Jan. 29, 1990, at A1 (“Leading molecular biologists say a technique promoted by the nation’s top law enforcement agency for identifying suspects in criminal trials through the analysis of genetic material is too unreliable to be used in court.”). Several of the biologists referred to in this story complained that their views were misrepresented, but the New York Times declined to print their letters to the editor. See Andre A. Moenssens, DNA Evidence and Its Critics—How Valid Are the Challenges?, 31 JURIMETRICS J. 87, 99–100 (1990).


5. For reviews of the challenges to admissibility, see Paul C. Giannelli, The DNA Story: An Alternative View, 88 J. CRIM. L. & CRIMINOLOGY 380 (1997) (concluding that courts were too willing to admit an untested technology); David H. Kaye, DNA Evidence: Probability, Population Genetics, and the Courts, 7 HARV. J.L. & TECH. 101 (1993) (suggesting that the principal objection to the computations of random-match probabilities was exaggerated); William C. Thompson, Evaluating the Admissibility of New Genetic Identification Tests: Lessons from the “DNA War,” 84 J. CRIM. L. & CRIMINOLOGY 22 (1993) (reviewing the debate on population structure but not discussing studies indicating that the effect is generally minor). Some of the leading court opinions are reproduced in D.H. KAYE, SCIENCE IN EVIDENCE 167–206 (1997).

6. See infra Part II.A.

7. This Article builds on a report prepared by the authors for the Legal Issues Working Group of the National Commission on the Future of DNA Evidence, entitled FORENSIC DNA TYING: SELECTED LEGAL ISSUES (2000). The authors are grateful to Paul Bender, Susan Ehrlich, Rockne Harmon, Dorothy Nelkin, Barry Scheck, Michael Smith, Ralph Spritzer, Jeffrey Thoma, William Thompson, James Weinstein, and Richard Wilting for information, comments, discussions, or arguments about topics discussed in this Article.
Part I discusses constitutional problems that arise when the government uses DNA evidence in investigating a crime. It focuses primarily on methods of acquiring DNA from an individual suspected of committing a crime and considers the constitutionality of compelling suspects to submit to DNA sampling and of acquiring stored samples of a suspect’s DNA or medical records relating to these samples from private medical providers or laboratories. It also considers the constitutionality of gathering DNA from large numbers of people to see whether any have genotypes that match those seen in the trace evidence. It shows that although the Fourth Amendment usually requires the police to have probable cause and a warrant to compel a person to provide a DNA sample, there are many situations in which police may be able to secure DNA samples without these safeguards.

Part I also considers a second investigative use of DNA: deducing physical or other characteristics of an individual whose DNA is found at the scene of crime. Genetic typing will permit inferences as to ancestry, physiognomy, or inherited disorders. Part I concludes that investigators can use genetic data to make valid inferences without infringing any constitutional rights.

Part II addresses legal issues that arise at a later stage in the justice system, when DNA analysis is used as evidence in the prosecution of persons charged with crimes. It discusses the admissibility of new DNA tests and the results of proficiency tests at particular laboratories and suggests that the rules of character evidence create a largely unrecognized, and probably undesirable, obstacle to admissibility. It also considers proposals that the durability of DNA evidence justifies extending the statute of limitations for prosecutions for certain crimes. It calls for more legislative sensitivity to the policies that underlie

---

8. This Article does not consider legal issues associated with collecting or using DNA from individuals who are not specifically suspected of the crime that is under investigation. These issues are central to building and administering many DNA databases for law enforcement and will be addressed in a separate article that is in progress. The most comprehensive treatment published to date is Michelle Hibbert, DNA Databanks: Law Enforcement’s Greatest Surveillance Tool?, 34 WAKE FOREST L. REV. 767 (1999). For rejoinders to some of the criticisms advanced there, see D.H. Kaye, Bioethics, Bench, and Bar: Selected Arguments in Landry v. Attorney General, 40 JURIMETRICS J. 193 (2000); David H. Kaye & Edward J. Imwinkelried, FORENSIC DNA TYPING: SELECTED LEGAL ISSUES (2000). For discussion of the acquisition of DNA for the purpose of proving that a convict is innocent, see Cynthia Bryant, When One Man’s DNA Is Another Man’s Exonerating Evidence: Compelling Consensual Sexual Partners of Rape Victims to Provide DNA Samples to Postconviction Petitioners, 33 COLUM. J.L. & SOC. PROBS. 113 (2000).
limitations on the period of criminal liability and a better understanding of how the durability of DNA evidence interacts with these policies.

For some of these emerging issues, it is possible to make a reasonably confident prediction of the courts’ ultimate resolution of the questions posed by DNA technology. However, other questions seem much closer and require more attention to the underlying policies. The hope is that this Article will stimulate judicial attention and scholarly commentary by identifying the full range of issues and analyzing some of the unresolved questions surrounding the use of DNA evidence.

I. DNA ANALYSIS IN CRIMINAL INVESTIGATIONS

Traditionally, DNA has been employed to link a suspect to a crime. Finding that a suspect’s DNA matches the DNA left at a crime scene, for example, tends to incriminate the suspect.9 Inversely, when the DNA does not match, the suspect usually can be excluded as the source of the crime-scene DNA.10 If trace evidence is to be used in these ways, the police must secure samples of DNA from individuals who might have committed the crime under investigation. Officials can secure such samples in many ways. They can seek a court order to compel an individual to submit to sampling; they can turn to a preexisting collection of DNA samples; they can take a sample with the consent of the individual; or they can try to locate a sample that the suspect has abandoned.

As a matter of constitutional law, the principal constraint11 on such government action is the Search and Seizure Clause of the Fourth Amendment to the U.S. Constitution,12 which states:

9. It does so by suggesting that the suspect is the source of the crime-scene DNA. Of course, other explanations may exist. The match might be a laboratory artifact, it might be coincidental in that an unrelated person is the source of the trace evidence, or the match could be the result of kinship in that a close relative of the defendant is the source. Suitable investigative and testing procedures often can eliminate such alternative hypotheses. See David H. Kaye & George F. Sensabaugh, Jr., Reference Guide on DNA Evidence, in Reference Manual on Scientific Evidence 485 (Federal Judicial Center ed., 2d ed. 2000).

10. As with an inclusion, there can be other explanations for a reported exclusion. See Kaye & Sensabaugh, Jr., supra note 9.

11. The Self Incrimination Clause of the Fifth Amendment and the Due Process Clauses of the Fifth and Fourteenth Amendments have less force in this context. See D.H. Kaye, The Constitutionality of DNA Sampling on Arrest, 10 CORNELL J.L. & PUB. POL’Y (forthcoming 2001); cf., e.g., Schmerber v. California, 384 U.S. 757, 760–65 (1966) (rejecting such claims with regard to involuntary taking blood from an individual suspected of driving while intoxicated for the
The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no Warrants shall issue, but upon probable cause, supported by Oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized.  

However, what makes a search reasonable and when a warrant supported by probable cause is required are not always apparent. The following sections analyze the application of these requirements to the various methods of acquiring DNA from suspects.

A. Legality of Compelling Suspects To Submit to DNA Extraction

Because the Search and Seizure Clause protects “[t]he right of the people to be secure in their persons,” it applies both to restraining a person and to searching the inside or outside of a person’s body. The Supreme Court made this clear in Schmerber v. California. Schmerber was arrested at a hospital while receiving treatment for injuries suffered in an automobile accident. He had allegedly been driving while intoxicated, and a police officer directed a physician at the hospital to withdraw a blood sample. Chemical analysis revealed a high concentration of alcohol in his blood at the time of the offense, and the report of this analysis was admitted in evidence at trial in Los Angeles Municipal Court even though the defendant objected that he never consented to the test. He was convicted of driving an automobile while under the influence of intoxicating liquor, and the state appellate court affirmed the conviction. The U.S. Supreme Court affirmed, but only


13. U.S. CONST. amend. IV.


15. Id. at 758.

16. Id. at 758–59.

17. Id. at 759.
after applying the Search and Seizure Clause to the act of removing the biological sample. The majority began its analysis from the premise that “such testing procedures plainly constitute searches of ‘persons,’ and depend antecedently upon seizures of ‘persons’ . . . within the meaning of that Amendment.”

The Court emphasized that “[t]he integrity of an individual’s person is a cherished value of our society.”

Because Schmerber established that the Fourth Amendment applies to removing material from a suspect’s body, as a general rule, police must persuade a judge or magistrate that there is probable cause to believe that the desired DNA sample will produce evidence linking the suspect to the crime. With judicial authorization, police can use necessary force to extract the biological material. Furthermore, once the authorities legally have acquired a suspect’s profile, they are permitted to compare it to profiles from unrelated, unsolved crime-scene stains. The current state of the law appears to allow evidence legitimately acquired for one purpose to be used for another purpose, at least if the additional use entails no further search or seizure of the person.

18. Id. at 766–72.
19. Id. at 767.
20. Id. at 772.
23. See Bickley v. State, 489 S.E.2d 167, 170 (Ga. Ct. App. 1997) (rejecting the claim that investigators violated Fourth Amendment by using a DNA sample taken pursuant to a search warrant in a 1994 rape investigation to convict a man of two earlier rapes as well as the 1994 rape); Smith v. State, 734 N.E.2d 706, 709–10 (Ind. Ct. App. 2000) (involving DNA obtained by court order in rape case in which defendant was acquitted on a consent defense; police found a match between this DNA sample and a profile in the state’s database of DNA from unsolved crimes; the court held the database check was constitutional because “police conduct in comparing Smith’s court-ordered DNA sample with the DNA obtained from the V.O. rape is not a Fourth Amendment search or seizure”); Wilson v. State, 752 A.2d 1250, 1268–72 (Md. Ct. Spec. App. 2000) (stating that because “[n]o new Fourth Amendment intrusion is involved,” use of a previously legally obtained sample to link defendant to present crime obviated need for new warrant even though police had obtained one); People v. King, 663 N.Y.S.2d 610, 614 (App. Div. 1997) (stating that police could use the profile from a sample obtained under a warrant with probable cause in a second rape investigation even if they lacked probable cause to acquire the sample for that investigation because “once a person’s blood sample has been obtained lawfully, he can no longer assert either privacy claims or unreasonable search and seizure arguments with respect to the use of that sample”).
24. Because the invasion of privacy was justified (by probable cause and a warrant for seizing the DNA and searching its structure), the conventional additional-use theory would allow the later
DNA Typing Issues

In some circumstances, however, either a warrant or probable cause might not be essential to obtain the sample in the first place. For instance, if a person is legitimately under arrest, the seizure of the person is justified, and routine, non-invasive DNA sampling of all arrestees solely for the purpose of creating a record of the true identity of the individual is probably constitutional. Furthermore, according to conventional wisdom, once the government has acquired the sample further comparisons. See supra note 23. An analogy can be drawn to the situation in which police searching a dwelling for specific stolen items record the serial number of an item not enumerated in the warrant and check this number against a list of serial numbers of other stolen items. Recording the serial number would not be considered a search if the number was in plain view. Cf. Arizona v. Hicks, 480 U.S. 321, 324–25 (1987) (“[T]he mere recording of the serial numbers did not constitute a seizure, [but] taking action, unrelated to the objectives of the authorized intrusion, which exposed to view concealed portions of the apartment or its contents, did produce a new invasion of respondent’s privacy unjustified by the exigent circumstance that validated the entry.”). Under the additional-use theory, checking the number against a list is not an additional search of the person or the property and therefore would be allowed. Cf. State v. Wamre, 599 N.W.2d 268, 274–77 (N.D. 1999) (holding that the Fourth Amendment was satisfied when the police used a serial number that was in plain view during a search to secure a warrant by telephone that allowed them to seize the additional items).

The logic of the existing Fourth Amendment doctrine has been sharply questioned. See Harold J. Krent, Of Diaries and Data Banks: Use Restrictions Under the Fourth Amendment, 74 Tex. L. Rev. 49, 53 n.25 (1995) (arguing that the current understanding of the Fourth Amendment should be altered so that “each governmental use of information about an individual constitutes a separate seizure of that person’s effects” and hence “all such uses must satisfy the reasonableness requirement”).

A valid arrest requires probable cause to believe that the individual has committed an offense. Alternatively, the presence of exigent circumstances may justify a seizure in the absence of a warrant. This was the basis for upholding the search in Schmerber v. California, 384 U.S. 757 (1966). The Court reasoned that there was an exigency because blood-alcohol concentrations decline rapidly. Id. at 770–72. However, DNA cases are readily distinguishable. In the typical case, the police desire a DNA sample to test for permanent, identifying markers—markers that will not evaporate or disappear with the mere passage of time. Consequently, it will be more difficult for the authorities to justify a warrantless seizure of a DNA sample than it would be to justify a similar acquisition of a blood sample for intoxication testing. See In re J.W.K., 583 N.W.2d 752, 757 (Minn. 1998). In the absence of exigent circumstances, the police would be obliged to obtain the functional equivalent of a warrant, that is, a court order that the suspect furnish a DNA sample. See Thurman v. State, 861 S.W.2d 96, 100 (Tex. Ct. App. 1993).

26. See Kaye, supra note 11 (describing the types of sampling and the protections on disclosure of the information that might be required to satisfy the Due Process and the Search and Seizure Clauses of the Bill of Rights). Of course, a state may adopt a more restrictive approach under its own constitution. See generally DARIEN A. McWHIRTER & JOHN D. BIBLE, PRIVACY AS A CONSTITUTIONAL RIGHT 174, 178 (1992); RICHARD C. TURKINGTON & ANITA L. ALLEN, PRIVACY LAW 122–25 (1999) (listing state constitutional provisions).
consistent with Fourth Amendment protections, the Search and Seizure Clause does not bar its use for another purpose. 27

It also is likely that an order compelling a person to give a sample could be issued on something less than probable cause. In Davis v. Mississippi, 28 the Supreme Court suggested in dictum such a procedure. A woman in Meridian, Mississippi, reported that “a Negro youth” broke into her home and raped her. Police, “without warrants, took at least 24 Negro youths,” including Davis, “to police headquarters where they were questioned briefly, fingerprinted, and released without charge.” 29 After Davis’s fingerprints were discovered to match those lifted from the windowsill, he was indicted, tried, and convicted. 30 His objection to the admission of the fingerprint evidence was overruled, and the Mississippi Supreme Court affirmed the conviction on the theory that fingerprint evidence is so reliable that the Fourth Amendment exclusionary rule does not apply to this evidence. 31 The U.S. Supreme Court reversed. The Court held that the Fourth Amendment requires the exclusion of evidence that is the fruit of an unreasonable search or seizure, regardless of how reliable that evidence may be. 32 Reasoning that Davis was detained without a warrant and without probable cause, and that he was not merely fingerprinted but interrogated, the Court concluded that the resulting fingerprints were inadmissible. 33 However, the Court’s response to the state’s argument that an arrest made solely for the purpose of obtaining fingerprints should be allowed without probable cause was less definitive. Although Justice Brennan, writing for the

27. See supra note 23. Under City of Indianapolis v. Edmond, 121 S. Ct. 447 (2000), and Ferguson v. City of Charleston, 121 S. Ct. 1281 (2001), the reuse would not be allowed if the samples from arrestees were obtained as part of a program that had as its primary goal the acquisition of DNA samples for later database searches. See Kaye, supra note 11. In addition this reuse would not be acceptable under the theory advanced in Krent, supra note 24, at 53 n.25. A narrower version of this view holds that when the police rely on a special justification to deviate from the normal Fourth Amendment requirements of probable cause or a warrant, their utilization of the evidence seized must be limited to uses that promote that special justification. Thus, if as in Schmerber v. California, 384 U.S. 757 (1966), the police justify a warrantless seizure of evidence on the theory that blood alcohol testing requires a sample to be taken without further delay, they could not use the blood sample for DNA testing. This narrower version has merit, although no published opinion has embraced it as a basis for excluding evidence.


29. Id. at 722.

30. Id.

31. Id. at 723–24.

32. Id. at 724.

33. Id. at 726–28.
majority of the Court, emphasized that “[d]etentions for the sole purpose of obtaining fingerprints are . . . subject to the constraints of the Fourth Amendment,” he added that:

It is arguable, however, that, because of the unique nature of the fingerprinting process, such detentions might, under narrowly defined circumstances, be found to comply with the Fourth Amendment even though there is no probable cause in the traditional sense . . . . Detention for fingerprinting may constitute a much less serious intrusion upon personal security than other types of police searches and detentions. Fingerprinting involves none of the probing into an individual’s private life and thoughts that marks an interrogation or search. Nor can fingerprint detention be employed repeatedly to harass an individual, since the police need only one set of each person’s prints. Furthermore, fingerprinting is an inherently more reliable and effective crime solving tool than eyewitness identifications or confessions . . . . Finally, because there is no danger of destruction of fingerprints, the limited detention need not come unexpectedly or at an inconvenient time.

The Court opened the door to the possibility that “the requirements of the Fourth Amendment could be met by narrowly circumscribed procedures for obtaining, during the course of a criminal investigation, the fingerprints of individuals for whom there is no probable cause to arrest.” The Court virtually invited states to devise procedures to obtain evidence of identifying characteristics on the basis of something less than probable cause.

Many states seized on this invitation by adopting statutes or court rules permitting the police to obtain evidence of identifying physical characteristics after a showing of founded or reasonable suspicion. For instance, Arizona authorizes magistrates to issue “an order authorizing . . . temporary detention, for the purpose of obtaining evidence of identifying physical characteristics” on a showing of “[r]easonable cause
for belief that a felony has been committed and proof that the “physical characteristics . . . may contribute to the identification of the individual who committed such offense.” As in this instance, the language of many of these statutes and court rules is broad enough to apply to DNA samples.

One might argue that these statutes or rules are too broad—that unlike the fingerprints in *Davis*, blood, urine, or hair samples should be treated differently because they have the potential to reveal information that is more significant than the pattern of whorls and ridges in a fingerprint. Some support for this distinction can be found in *Skinner v. Railway Labor Executives’ Ass’n*, which involved drug testing of railway employees. The Court observed that “chemical analysis of urine, like that of blood, can reveal a host of private medical facts about an employee, including whether he or she is epileptic, pregnant, or diabetic.” The same concern with “private medical facts” arises with

---

38. ARIZ. REV. STAT. ANN. § 13-3905(A) (West 1999).

39. The Arizona statute specifies that “‘identifying physical characteristics’ includes, but is not limited to, the fingerprints, palm prints, footprints, measurements, handwriting, handprinting, sound of voice, blood samples, urine samples, saliva samples, hair samples, comparative personal appearance or photographs of an individual.” Id. § 13-3905(G).

40. Clyde M. Tande, Note, *DNA Typing: A New Investigatory Tool*, 1989 DUKE L.J. 474 (describing the breadth of the language of the various state statutes and court rules). For example, Alaska Rule of Criminal Procedure 16(c) allows a court to order detention to “[p]ermit the taking of samples of blood, hair and other materials of the person’s body which involve no unreasonable intrusion thereof” on the basis of an affidavit or testimony establishing probable cause to believe that: (i) An offense has been committed by one of several persons comprising a narrow focal group that includes the subject person; (ii) The evidence sought may be of material aid in identifying who committed the offense; and (iii) The evidence sought cannot practicably be obtained from other sources. ALASKA R. CRIM. P. 16(c)(2)(vii) & 16(c)(1).

Although the Supreme Court has not ruled authoritatively on the constitutionality of these procedures, in *Hayes v. Florida*, 470 U.S. 811 (1985), the Court referred approvingly to its previous statement in *Davis*. Id. at 817.


42. Id. at 617. Read in context, however, this language does not necessarily support imposing a requirement of probable cause. In *Skinner*, the Supreme Court was reviewing a judgment of the Ninth Circuit Court of Appeals striking down the drug testing regulations because they did not require any showing of individualized suspicion—not even reasonable suspicion, let alone probable cause. Id. at 612–13. The majority merely mentioned “private medical facts” in establishing that urinalysis constitutes a search. Id. at 617. It does not follow from the fact that a search is involved that probable cause is required. That is precisely the point made in *Davis*, where the Court indicated its willingness to relax the probable cause requirement for the undeniable searches or seizures involved in compelling a suspect to provide fingerprints. See *Davis v. Mississippi*, 394 U.S. 721, 727 (1969).
DNA Typing Issues

any samples that can be subjected to DNA analysis. To this extent, it would be facile to say that DNA typing, like the fingerprinting in *Davis*, "involves none of the probing into an individual’s private life and thoughts that marks an interrogation or search."**43** Certain parts of one’s genome—those that are related to otherwise nonobvious disease states or behavioral characteristics—are as much, if not more, a part of “an individual’s private life” as are the hormones or other chemicals found in one’s urine.

However, all the other factors listed in the *Davis* dictum apply to DNA sampling. Detention to obtain the sample cannot “be employed repeatedly to harass an individual, since the police need only one set of each person’s [DNA types].”**44** DNA analysis “is an inherently more reliable and effective crime-solving tool than eyewitness identifications or confessions.”**45** And, “the limited detention need not come unexpectedly or at an inconvenient time.”**46** Moreover, in describing these features of fingerprinting, the *Davis* Court recognized the possibility that the police might abuse even fingerprinting to harass or inconvenience a suspect.**47** The suggestion of relaxing the probable cause requirement presupposes the police will conform to the court order and the judiciary will issue orders that avoid these problems. This premise applies as well to the informational privacy concern voiced in *Skinner*. Just as there is no need to detain an individual repeatedly or to detain a person in the middle of the night, there is no reason for the police to probe parts of the genome that conceivably could be used to indicate disease states, susceptibilities, or the like.**48** Because the judicial order can limit the search to loci that are of strictly biometric interest, the analogy to *Davis* is apt. Detention for DNA typing, as much as detention for fingerprinting, “may constitute a much less serious intrusion upon personal security than other types of police searches and detentions.”**49** If a person can be compelled to submit to fingerprinting on reasonable

---

43. *Davis*, 394 U.S. at 727.
44. *Id.* at 727–28.
45. *Id.*
46. *Id.*
47. *Id.* at 726–27.
49. *Davis*, 394 U.S. at 727.
suspicion rather than probable cause, he or she can be required to submit to DNA sampling on the same showing. 50

B. Legality of Acquiring Samples or Records from Medical Providers or Laboratories

Rather than compel a person to submit to DNA sampling, police might obtain DNA data on a suspect from preexisting samples or databases. As of 1998, it was estimated that there were more than 282 million specimens of human biological material stored in the United States, with samples from another 20 million individuals accumulating each year. 51 These samples are stored in academically based repositories of scientists studying genetic disorders, commercially based repositories that offer DNA banking as a service to researchers and individuals, teaching and other hospitals that have acquired samples in the course of clinical diagnostic or surgical procedures, laboratories that screen blood samples of newborns for metabolic or other diseases pursuant to state public health laws, and armed forces repositories of pathology specimens and samples collected to permit identification of human remains. 52 Although the Fourth Amendment plainly regulates police

50. See In re Non-Testimonial Identification Order Directed to R.H., No. 99-353, 2000 WL 1234251, at *12–13 (Vt. Sept. 1, 2000) (upholding the constitutionality of a Vermont rule as applied to an order for a saliva sample on the basis of reasonable suspicion); cf. Doe v. Senechal, 725 N.E.2d 225, 231 (Mass. 2000) (holding that even if the Fourth Amendment applies to a private action for assault and battery and other torts, a court-ordered buccal swab to test whether a member of the staff of a residential treatment facility for mentally ill adolescents fathered the child of a patient is a reasonable search and seizure).

The conclusion that a court order based on probable cause (or perhaps reasonable suspicion) normally is required applies even if the police do not themselves demand or collect the DNA sample, but direct or request private citizens to acquire the sample. Suppose that shortly after an incident, a suspect goes to or is taken to a private hospital. While the suspect is still at the hospital, the police learn of the suspect’s location. The police contact the hospital staff and request them to obtain a DNA sample from the suspect for law enforcement use. The private hospital would be acting as a government agent in making the intrusion and the Fourth Amendment would apply. See Skinner v. Ry. Labor Executives’ Ass’n, 489 U.S. 602, 614–16 (1989). There is sufficient state action to trigger the Fourth Amendment if “the drawing of blood is instigated by the government,” State v. Hardy, 963 S.W.2d 516, 523 (Tex. Crim. App. 1997), by, for instance, requesting that an emergency room doctor take a sample. Commonwealth v. Kohl, 615 A.2d 308, 310 (Pa. 1992). When the authorities intervene and request a sample even before the private entity takes one, the otherwise private entity is acting as a government agent in seizing the sample.


52. See id. at 13–15; Lawrence O. Gostin, Health Information Privacy, 80 CORNELL L. REV. 451, 464, 467–68 (1995); M. Therese Lysaught et al., A Pilot Test of DNA-based Analysis Using
efforts to obtain samples directly from suspects,\textsuperscript{53} the prohibition against unreasonable searches applies only to government action.\textsuperscript{54} As the Supreme Court commented in \textit{United States v. Jacobsen}:\textsuperscript{55} “This Court has . . . consistently construed this protection as proscribing only governmental action; it is wholly inapplicable ‘to a search or seizure, even an unreasonable one, effected by a private individual not acting as an agent of the Government or with the participation or knowledge of any governmental official.’”\textsuperscript{56}

This state action doctrine raises the possibility that police may be able to acquire preexisting information from cooperative private hospitals or laboratories without a court order and without probable cause or reasonable suspicion.\textsuperscript{57} This practice would fall outside the constraints of the Fourth Amendment on two conditions: (1) the government did not instigate the original acquisition of the data, and (2) in acquiring the data that the suspect already has provided private entities, the state is not engaging in any search or seizure.

\begin{footnotesize}

\textsuperscript{53} When the police obtain a sample directly from a private citizen, there is undeniably sufficient state action to bring the Fourth Amendment into play. \textit{E.g.}, \textit{In re J.W.K.}, 583 N.W.2d 752, 754–56 (Minn. 1998); State v. Binner, 886 P.2d 1056, 1057–58 (Or. Ct. App. 1994). The extent of the government involvement is the same whether the motivation of the police is to obtain the sample to add to a database or to acquire an evidential sample to be compared to a database.

\textsuperscript{54} United States v. Pervaz, 118 F.3d 1, 6 (1st Cir. 1997); United States v. Reed, 15 F.3d 928, 931 (9th Cir. 1994); State v. Grant, 620 N.E.2d 50, 60 (Ohio 1993); State v. Maxfield, 125 Wash. 2d 378, 384, 886 P.2d 123, 127 (1994), \textit{rev’d on other grounds}, 133 Wash. 2d 332, 945 P.2d 196 (1997).


\textsuperscript{57} If the medical provider or researcher were uncooperative, the authorities could resort to compulsory process such as a subpoena. \textit{See, e.g.}, United States v. Dionisio, 410 U.S. 1, 10–11 (1973) (stating that “a grand jury subpoena to testify is not that kind of governmental intrusion on privacy against which the Fourth Amendment affords protection, once the Fifth Amendment is satisfied,” but recognizing that “[t]he Fourth Amendment provides protection against a grand jury subpoena duces tecum too sweeping in its terms ‘to be regarded as reasonable’”); State v. Fears, 659 S.W.2d 370, 375–76 (Tenn. Crim. App. 1983) (requiring only a showing that the object of the subpoena is logically relevant to the subject matter of a legitimate criminal investigation).

\end{footnotesize}
The first condition is relatively straightforward and often will be satisfied for medical records and tissue samples. 58 Pathology specimens at private hospitals, for example, would fall into this category, but newborn screening samples compelled under state law would not.

The second condition is more subtle, for it depends on the meaning given to the phrase “search or seizure.” The basic framework for determining whether a form of data collection amounts to a search or seizure for Fourth Amendment purposes is found in Katz v. United States. 59 In Katz, the government acquired key evidence to convict the defendant of interstate gambling by attaching an electronic listening-and-recording device to the outside of a public telephone booth. The government argued that the interception was not a search because there was no physical trespass and the telephone booth was a public place. 60 The Supreme Court held that neither entry onto private property nor inspection of tangible items is an essential feature of a search, for “the Fourth Amendment protects people, not places.” 61 The Fourth Amendment protected the defendant, the Court explained, because “a person in a telephone booth . . . who occupies it, shuts the door behind him, and pays the toll that permits him to place a call is surely entitled to assume that the words he utters into the mouthpiece will not be broadcast to the world.” 62 Because the federal agents had no warrant authorizing the interception, the Court held that the search violated the Fourth Amendment. 63 In a concurring opinion, Justice Harlan elaborated on the majority’s remarks. In perhaps the most famous passage in the Katz opinions, he wrote: “[T]here is a twofold requirement, first that a person have exhibited an actual (subjective) expectation of privacy and, second, that the expectation be one that society is prepared to recognize as ‘reasonable.’” 64 Applying this standard, he explained that “[t]he point is not that the booth is ‘accessible to the public’ at other times, but that it is

58. The condition is fulfilled if, before the government’s request, the private parties acted on their “own initiative” and out of an “independent” medical or research “motivation.” State v. Comeaux, 818 S.W.2d 46, 50 (Tex. Crim. App. 1991).
60. Id. at 352–53.
61. Id. at 351.
62. Id. at 352.
63. Id. at 354–59.
64. Id. at 361 (Harlan, J., concurring).
a temporarily private place whose momentary occupants’ expectations of freedom from intrusion are recognized as reasonable.”

Although the courts have applied the *Katz* “test” in many contexts, for present purposes the most important is the Supreme Court’s 1976 decision in *United States v. Miller*. In *Miller*, the accused was charged with possessing an unregistered still, carrying on a distillery business without paying the whiskey tax, and possessing whiskey on which no taxes had been paid. Prior to trial, the government served subpoenas on two banks at which the defendant had accounts. The banks surrendered copies of the defendant’s checks and deposit slips as well as the bank’s own records of the defendant’s accounts. The defendant moved to suppress the documents, but the trial judge denied the motion and admitted the evidence at trial. The defendant was convicted and later appealed. The First Circuit reversed, holding that the banks’ surrender of the records violated the defendant’s Fourth Amendment rights.

On appeal, the Supreme Court reversed the First Circuit’s decision and reinstated the defendant’s conviction. Writing for the majority, Justice Powell relied heavily on *Katz*. The Court upheld the denial of the suppression motion because “there was no intrusion into any area in which [the defendant] had a protected Fourth Amendment interest.” Because the defendant had transferred the checks and deposit slips to the bank, he could not assert “ownership [or possession] as to any of the subpoenaed records,” and “[a]ll of the documents obtained, including financial statements and deposit slips, contain only information voluntarily conveyed to the banks and exposed to their employees in the ordinary course of business.”

The defendant had argued that he retained an expectation of privacy because he had made the information available to the bank only “for a limited purpose.” However, Justice Powell made short shrift of that argument:

---

65. *Id.* (Harlan, J., concurring).
67. *Id.* at 436.
68. *Id.* at 438–39.
69. *Id.* at 439.
70. *Id.* at 440.
71. *Id.*
72. *Id.* at 442.
73. *Id.*
The depositor takes the risk, in revealing his affairs to another, that the information will be conveyed by that person to the Government . . . . This Court has held repeatedly that the Fourth Amendment does not prohibit the obtaining of information revealed to a third party, even if the information is revealed on the assumption that it will be used only for a limited purpose and the confidence placed in the third party will not be betrayed.\textsuperscript{74}

Justice Powell concluded that even if the banks were “acting solely as Government agents in” transcribing and surrendering the information, there was “no intrusion upon the [defendant’s] Fourth Amendment rights.”\textsuperscript{75}

The logic of \textit{Miller} has been applied to medical samples or records. As in the case of financial information, the patient or research subject neither owns nor possesses tissue samples or data that have been “voluntarily conveyed” to health care providers or medical researchers. Thus, in many of the cases challenging police requests for medical samples and records, the lower courts invoke \textit{Miller} to defeat the defendant’s claim.\textsuperscript{76}

\textsuperscript{74} Id. at 443.
\textsuperscript{75} Id.
\textsuperscript{76} See, e.g., Tims v. State, 711 So. 2d 1118, 1123–24 (Ala. Crim. App. 1997); People v. Perlos, 462 N.W.2d 310, 316 (Mich. 1990); State v. Guido, 698 A.2d 729, 734 (R.I. 1997); State v. Fears, 659 S.W.2d 370, 375–76 (Tenn. Crim. App. 1983); Thurman v. State, 861 S.W.2d 96, 98 (Tex. Ct. App. 1993); Robert M. Gellman, \textit{Prescribing Privacy: The Uncertain Role of the Physician in the Protection of Patient Privacy}, 62 N.C. L. REV. 255, 290–91 (1984) (“As a practical matter, in the absence of a statute or a definitive court decision, the \textit{Miller} decision is effectively being applied when medical records are subpoenaed.”). However, few cases deal with the specific question of obtaining DNA samples from private entities. See \textit{In re J.W.K.}, 583 N.W.2d 752, 754–57 (Minn. 1998). The vast majority concern blood samples used for intoxication testing. E.g., People v. Dolan, 408 N.Y.S.2d 249, 250–51 (Sup. Ct. 1978); State v. Enoch, 536 P.2d 460, 460–61 (Or. Ct. App. 1975); Commonwealth v. Kohl, 615 A.2d 308, 310–11 (Pa. 1992); Guido, 698 A.2d at 731–32; State v. Comeaux, 818 S.W.2d 46, 48–49 (Tex. Crim. App. 1991). To an extent, that area of law is sui generis. In the United States, drunk driving exacts a huge toll in death and economic loss. For that reason, every state has enacted an implied-consent law. ANDRE A. MOENSSENS ET AL., \textit{SCIENTIFIC EVIDENCE IN CIVIL AND CRIMINAL CASES} 176–225 (4th ed. 1995). By the terms of the implied-consent statute, when a citizen obtains a driver’s license, he or she impliedly consents to later tests of his or her blood, breath, or urine alcohol concentration. The existence of these statutes in every jurisdiction reflects legislative recognition of an important public interest in the prevention of drunk driving, and the provisions of the statutes also weaken the argument that the citizen has a protected expectation of privacy in either the sample or a medical record reflecting the results of a sample’s test. In the drunk-driving area, the implied-consent laws enable the government to contend both that the statutes reduce the citizen’s privacy expectations and that any expectations are ones that society is unprepared to recognize as reasonable.
Nevertheless, *Miller* might be distinguished on at least three bases. First, the *Miller* Court stated that the bank records in question concerned “commercial transactions.” In contrast, a medical record can relate to far more intimate aspects of a person’s life, and the law should not create significant disincentives to persons seeking necessary medical advice and treatment. Although a patient has a greater privacy interest in medical records than a business or individual has in bank records, it is not clear that the interest is so much deeper that the two types of records are distinguishable when the prosecution seeks the records to advance a criminal investigation. Although recent polls indicate that the vast majority of Americans are concerned about the privacy of their medical records, there are countervailing indications of the relevant, societally recognized expectations of privacy that are firmly settled. Most jurisdictions have enacted a general physician-patient privilege, which typically encompasses not only communications in a conventional sense, but also any information that the physician gains by virtue of the examination. Yet, the majority of jurisdictions do not recognize the privilege in criminal cases. Thus, several courts have pointed to that

---

77. 425 U.S. at 442.

78. See *Perlos*, 462 N.W.2d at 324 (Levin, J., dissenting) (“A person’s medical records are an intensely personal matter. Few persons would willingly share their medical records with the state.”); *State v. Binner*, 886 P.2d 1056, 1059 (Or. Ct. App. 1994) (“Such testing could reveal the most personal of the medical details of our private lives that would not be known to the public in general.”); Peter H. W. van der Goes, Jr., Comment, *Opportunity Lost: Why and How to Improve the HHS-Proposed Legislation Governing Law Enforcement Access to Medical Records*, 147 U. PA. L. REV. 1009, 1037–40 (1999).

79. See *Thurman*, 861 S.W.2d at 98.

80. *Fears*, 659 S.W.2d at 376 (asserting, after discussing *Miller*, “[w]e think that the same principle applies to medical records as to bank records and that the above holding of the *Miller* case is applicable”).

81. See van der Goes, *supra* note 78, at 1011 n.10 (“A 1996 CNN/Time poll indicates that 87% of Americans want to be asked permission every time their medical records are accessed for any reason.”).


limitation as evidence that society is unwilling to recognize a constitutionally enforceable privacy expectation in medical records or samples. In *Thurman v. State*, supra 85 for instance, the Texas Court of Appeals pointed out that the medical privilege is inapplicable to criminal cases, supra 86 hence, even if the citizen has a subjective expectation of privacy, society does not recognize that expectation as reasonable in a criminal setting, supra 87 For that matter, many jurisdictions not only carve out an exception to the privilege for criminal proceedings, but also require physicians to report certain types of events such as violent injuries and child abuse to the public authorities. supra 88 One jurisdiction has enacted even more sweeping legislation, requiring hospitals to furnish blood-test results to the prosecutor on request in a criminal investigation. supra 89 The reporting requirements and limited scope of the medical privilege strongly suggest that the person, who is the subject of the medical record, lacks a “societally recognized,” constitutionally protected privacy expectation. supra 90

Second, *Miller* might be distinguished by arguing that even if it permits the government to obtain medical records from a private source such as a hospital, it does not apply to a tissue sample. Certainly, the analogy between a DNA sample and the records in *Miller* is, at the very least, debatable, for there may be a more intense privacy expectation in the sample than in the records reflecting the results of a particular test on the sample. The sample represents a greater threat to privacy in that its existence would permit further testing and the revelation of additional information. Although noting that “[t]he precise question as to who owns the blood upon extraction from an individual raises a novel point

---


86. See id. at 100 (“Society can afford the physician-patient privilege in certain civil cases in order to protect personal privacy, but the need to protect the public from crime requires disclosure of the same information in criminal cases.”).

87. See id. at 98–101.

88. See State v. Fears, 659 S.W.2d 370, 376 (Tenn. Crim. App. 1983); Gellman, supra note 76, at 274.

89. See People v. Perllos, 462 N.W.2d 310 (Mich. 1990) (discussing MICH. COMP. LAWS. ANN. § 257.625(a)(9) (West 1990)).

without apparent judicial precedent,” a New York trial court suggested that the “defendant had a [property] interest in the blood specimen in . . . vial containers” retained by a private hospital.91 Also, in upholding subpoenas, courts occasionally stress that the records in question were “made [and] kept . . . by the hospital”92 or that the records were not “personal papers created or kept by the defendant,”93 suggesting that the outcome might be different if the defendant personally had produced the subpoenaed items.

Yet, most courts construe Miller to apply to biological samples as well as to mere records.94 In Miller, the prosecution sought not only financial statements that the bank had generated, but also checks and deposit slips from the defendant. Miller expressly rejected the argument that there was a significant difference between the documents generated by the bank and those prepared by the depositor.95 Financial statements prepared by a bank are like medical records prepared by a hospital or laboratory, and checks and deposit slips from the depositor can be analogized to samples from the defendant. To this extent, Miller appears pertinent whether the government seeks the original genetic samples or merely records documenting the results of tests on those samples.

Finally, Miller could be distinguished in that the defendant “voluntarily conveyed” the information to the banks96 while the voluntariness of providing tissue samples might be questionable. For instance, in People v. Perlos,97 a leading case98 involving blood-alcohol testing, a dissenting justice of the Michigan Supreme Court argued that

91. People v. Dolan, 408 N.Y.S.2d 249, 252 (Sup. Ct. 1978). The usual view is that body parts and tissues are not property, but that they are subject to the legal protections afforded to a person’s body. See Moore v. Regents of the Univ. of Cal., 793 P.2d 479, 487–97 (Cal. 1990) (holding, in response to a claim for conversion, that a commercially valuable cell line is not the property of the patient from which the original cells were taken); Moe M. Litman, The Legal Status of Genetic Material, in HUMAN DNA: LAW AND POLICY, INTERNATIONAL AND COMPARATIVE PERSPECTIVES 17, 25–27 (Bartha Maria Knoppers ed., 1997) (reviewing cases and advocating a sui generis classification in which genetic material is regarded as a hybrid of private property, common property, person, and information).
94. See Perlos, 462 N.W.2d at 317 (stating, immediately after discussing Miller, that this line of authority deals with “privacy rights in medical records [and] blood samples”).
96. Id.
98. 1 WAYNE R. LAFAVE, SEARCH AND SEIZURE: A TREATISE ON THE FOURTH AMENDMENT § 2.7(d), at 638 (3d ed. 1996).
although under the implied consent statute a driver agrees in advance to a
government test of his or her alcohol concentration, the driver does not
consent to a search of medical records reflecting a test conducted by a
private entity. As the dissent observed, “[i]n today’s society, a person has
little choice but to undergo medical treatment at a medical facility,
generally licensed by and authorized to operate by the state. Few persons
have the ability to obtain medical treatment in their homes . . . .”99
Nevertheless, the majority relied on \textit{Miller} to uphold a state statute
mandating that hospital personnel disclose to the prosecution the results
of any blood-alcohol test of a driver involved in an accident.100

When a conscious person in need of medical treatment is admitted to
a hospital, the patient consents to treatment\textsuperscript{101} and explicitly or implicitly
agrees to medical testing incident to the treatment. The patient therefore
voluntarily conveys the data disclosed by the test results to the
institution’s staff in the same manner that the depositor in \textit{Miller}
“voluntarily conveyed [information] to the banks [to be] exposed to their
employees in the ordinary course of business.”\textsuperscript{102} Given the rules of
medical ethics,\textsuperscript{103} the patient might have a stronger subjective expec-
tation that the hospital will keep the information in question confidential. However, \textit{Miller} states that when a person voluntarily
reveals information to a third party, for Fourth Amendment purposes the
person “takes the risk” that the third party will disclose the information
“to Government authorities, even if the information is revealed . . . only

\textsuperscript{99} \textit{Perlos}, 462 N.W.2d at 324 & n.7 (Levin, J., dissenting); \textit{see also} \textit{State v. Hardy}, 963 S.W.2d
516, 524 (Tex. Crim. App. 1997) (“\textit{R}elease of medical information to hospitals is less optional
than the release of financial information to banks. A person can choose not to maintain a bank
account, but it hardly seems reasonable to expect someone to forego medical attention.”).
\textsuperscript{100} \textit{See Perlos}, 462 N.W. 2d at 316.
\textsuperscript{101} If the person in need of emergency medical treatment is unconscious, the hospital will still
administer treatment on the legal theory underlying the implied-intent statutes: the hospital
personnel may ordinarily assume that if the person were conscious and understood his or her
condition, he or she would want to be treated. \textit{See People v. Woodson}, 630 N.Y.S.2d 670, 671
(Sup. Ct. 1995) (involving defendant who was unconscious and comatose when he was originally
admitted to the hospital); \textit{State v. Guido}, 698 A.2d 729, 732 (R.I. 1997) (involving defendant who
was unconscious when he was found, and the hospital staff withdrew a blood sample during
emergency-room treatment); \textit{Hardy}, 963 S.W.2d at 526–27 (“Many statutes also permit officers to
conduct a chemical test (without a warrant) on an unconscious person . . . . Some statutes even
permit obtaining a sample for chemical testing (without a warrant) from an unconscious person
even when that person is not under arrest.”); \textit{State v. Jenkins}, 259 N.W.2d 109, 110 (Wis. 1977)
(involving defendant who was in and out of consciousness after an accident).
\textsuperscript{103} \textit{See SUZANNE B. MCNICOL, LAW OF PRIVILEGE} 342 (1992) (citing the American Medical
Association and the World Medical Association’s codes).
for a limited purpose and the [person expects that] confidence placed in the third party will not be betrayed."104

In sum, the argument that Miller governs both medical records and samples is strong, but not conclusive. The statutory patterns, restricting the medical privilege and imposing reporting duties on physicians, lend powerful support to the view that there is no reasonable expectation of privacy in medical records or samples. Admittedly, a minority of cases reject this conclusion105—sometimes on dubious grounds106—and Miller itself has been the target of intense criticism.107 However, the majority view is that if the private hospital or laboratory obtains a biological sample on its own initiative for lawful medical reasons, its subsequent surrender of the sample to the authorities does not violate any constitu-

104. Miller, 425 U.S. at 443.
105. See Hardy, 963 S.W.2d at 525 (“No consensus exists in court decisions on whether an expectation of privacy exists in medical records.”); id. at 524 (acknowledging a division of sentiment in the Texas cases, then stating that “[f]our out of five [Texas courts] held that society does not recognize as reasonable an expectation of privacy in those kinds of records”); see also State v. Nelson, 941 P.2d 441, 448 (Mont. 1997) (“[M]edical records have not been historically protected by the Fourth Amendment’s prohibition against unreasonable searches and seizures.”); 1 LAFave, supra note 98, § 2.7(d), at 640.

The minority view that the suspect retains a constitutionally protected privacy expectation in the sample would not necessarily prevent the state from obtaining the medical information or sample. The state usually has to establish founded suspicion or even probable cause to obtain a court order that the hospital or laboratory deliver the record or the sample. See Nelson, 941 P.2d at 449 (finding “a compelling state interest” in the form of “probable cause”).

106. For example, in People v. Woodson, 630 N.Y.S.2d 670 (Sup. Ct. 1995), an unconscious patient was taken to a private hospital. The hospital personnel prepared vials of blood and urine specimens. The grand jury subpoenaed the samples, the hospital turned over both vials, and the authorities conducted toxicological tests. The defendant argued that the authorities’ acquisition of the two vials violated the Fourth Amendment and moved to suppress the vials and testimony as to their contents. Id. at 671. A New York Supreme Court granted the motion and declared that “[b]y now, it is settled law that a hospitalized defendant retains a Fourth Amendment right in blood and urine samples.” Id. The court in effect ruled the suspect had an objectively reasonable expectation of privacy in his hospital records. See id. at 671–74. However, the reasoning in Woodson is dubious. The judge cited People v. Natal, 553 N.E.2d 239 (N.Y. 1990), to support the exclusion of the evidence, but Natal has nothing to do with blood or urine samples. Woodson, 630 N.Y.S.2d at 787. Natal involved clothing and personal effects taken from a defendant when he was placed in pretrial confinement. Natal, 553 N.E.2d at 240–41. Although the Natal court held that the district attorney had abused the subpoena power, the court ruled that “inspection of personal effects previously exposed to police view . . . does not invade any substantial privacy interest.” Id. at 241. Moreover, the Woodson court does not so much as mention, much less evaluate, the possibility that Miller is pertinent. Similarly ignoring Miller, the Pennsylvania Supreme Court recognized a privacy interest in medical records in Commonwealth v. Riedel, 651 A.2d 135, 138–39 (Pa. 1994).

107. See, e.g., 1 LAFAVE, supra note 98, § 2.7(d), at 639–40.
tionally protected expectation of privacy. The same is true, even more clearly, of laboratory findings or medical records involving the samples. Although it would be desirable to have additional statutory and regulatory protection for medical records, it is difficult to escape the conclusion that under Miller and its progeny, the Fourth Amendment does not confer that protection.

However, the Court’s recent decision in Ferguson v. City of Charleston complicates the analysis. Ferguson involved a challenge to a set of policies and procedures developed by a public hospital, the Medical University of South Carolina (MUSC), and local law enforcement authorities. MUSC was administering diagnostic tests, including screens for cocaine use, to patients receiving prenatal treatment. If a patient tested positive, the MUSC staff referred her for

108. The courts have employed a variety of rationales to justify this view. Some reason that the suspect had bailed the sample with the private hospital and that as bailee, the hospital “may consent to . . . permit the warrantless search and seizure.” People v. Dolan, 408 N.Y.S.2d 249, 252 (Sup. Ct. 1978). Other courts simply assert that if a private hospital properly obtains a sample for medical reasons, as “an independent private citizen” it may “turn[] it over to the police.” State v. Enoch, 536 P.2d 460, 461 (Or. Ct. App. 1975). But still other courts invoke Miller. See, e.g., Thurman v. State, 861 S.W.2d 96, 98–101 (Tex. Ct. App. 1993) Although Thurman involved a subpoena for medical records, the court used broad language, expansive enough to apply to samples as well, and noted that “[c]ourts in other states have reached the same conclusion.” Id.

109. See State v. Gonzalez, 852 P.2d 851 (Or. Ct. App. 1993); State v. Fears, 659 S.W.2d 370 (Tenn. Crim. App. 1983); Conrad v. Texas, No. 14-96-00167-CR, 1997 WL 764527 (Tex. Ct. App. 1997), cert. denied, 526 U.S. 1068 (1999); Thurman, 861 S.W.2d 96. Many of these courts not only cite Miller but also treat it as dispositive. See People v. Perlos, 462 N.W.2d 310, 316–17 (Mich. 1990); Fears, 659 S.W.2d at 375–76; Gellman, supra note 76, at 290–91. In Nelson, the Montana Supreme Court found a protected expectation of privacy in medical records under its state constitution. 941 P.2d at 447–49. However, the Montana court pointed out that its cases construing its constitution protect privacy “more strict[ly] than” the cases interpreting the Fourth Amendment to the U.S. Constitution. Id. at 447. Indeed, the Montana court noted that “medical records have not been historically protected by the Fourth Amendment’s prohibition against reasonable searches and seizures.” Id. at 448.


111. Of course, it would not be unthinkable to extend the Fourth Amendment in that manner. However, that extension would necessitate a doctrinal reformulation and the rethinking of several leading modern precedents, notably Miller and its progeny.


113. Id. at 1284–86.
counseling and treatment. Despite the referrals, the incidence of cocaine use among the patients remained the same. To give patients a stronger incentive to refrain from drug use, MUSC contacted the local law enforcement authorities. Together, they developed the set of policies and procedures in question. Under these procedures, patients meeting certain criteria were tested for cocaine, and a chain of custody for the sample was maintained. In their final form, the procedures provided that if a woman tested positive for cocaine either during pregnancy or after labor, she would be given an opportunity to avoid arrest by consenting to substance-abuse treatment. The document codifying the policy specified the range of possible criminal charges and the logistics of police notification and arrest.

Several women who were arrested challenged the constitutionality of the practice under the Fourth Amendment. A six-justice majority invalidated the program. Six justices agreed that MUSC’s surrender of the test results to the police constituted a separate Fourth Amendment intrusion. In response to the public hospital’s practice of regularly submitting copies of team documents discussing the women’s progress, Justice Stevens wrote that “[t]he reasonable expectation of privacy enjoyed by the typical patient undergoing diagnostic tests in a hospital is that the results of those tests will not be shared with nonmedical personnel without her consent.”

Although, on its face, this statement is quite broad, Ferguson deals only with a continuing program developed by the police and a public hospital requiring the systematic disclosure of patient records for the “primary purpose” of advancing “the general interest in crime

114. Id. at 1284.
115. Id.
116. Id.
117. Id. at 1285.
118. Id. at 1291.
119. Id. at 1285.
120. Id. at 1285, 1291.
121. Id. at 1284.
122. Justice Stevens authored the majority opinion, Justice Kennedy filed a concurrence in the judgment, and Justice Scalia filed a dissent in which Chief Justice Rehnquist and Justice Thomas joined. Id. at 1284, 1293, 1296.
123. Id. at 1284.
124. Id. at 1288.
control.” The involvement of law enforcement was so “extensive . . . at every stage of the policy” that the hospital was acting “as an institutional arm of law enforcement.”

As such, *Ferguson* is distinguishable from the typical fact situation discussed in Part I.A, namely, a case in which medical personnel in the private sector surrender information that they previously obtained for legitimate, health-related reasons. Indeed, in a footnote, Justice Stevens suggested the distinction between the institutional program in *Ferguson* and the problem discussed here. After noting the existence of laws requiring medical personnel to report certain types of criminal activity to the police, Justice Stevens stated:

> While the existence of such laws might lead a patient to expect that members of the hospital staff might turn over evidence acquired in the course of treatment to which the patient has consented, they surely would not lead a patient to anticipate that hospital staff would *intentionally set out to obtain incriminating evidence from their patients for law enforcement purposes*. If the medical personnel previously obtained the data from the patient during the course of a regular hospital procedure, they could not be said to have “intentionally set out to obtain incriminating evidence . . . for law enforcement purposes.” In that situation, where private medical facilities later surrender information to the authorities, *Miller*—not *Ferguson*—presumably would still govern.

C. Acquiring DNA from Inadvertently Abandoned Samples

The police also might obtain a suspect’s DNA sample surreptitiously, without detaining the person. Saliva deposited on a coffee cup at a restaurant, for example, can be collected and analyzed. Police

---

125. *Id.* at 1290.
126. *Id.* at 1283.
127. *Id.* at 1294 (Kennedy, J., concurring). For his part, Justice Kennedy stressed the “substantial law enforcement involvement in the policy from its inception.” *Id.* (Kennedy, J., concurring). He emphasized that “the active use of law enforcement . . . [was] an integral part of [the] program.” *Id.* (Kennedy, J., concurring).
128. *Id.* at 1288 n.13.
129. *Id.* (emphasis added).
130. *Id.*
131. In New York City, police used this tactic to acquire DNA from a suspected serial killer and rapist. Richard Willing, *As Police Rely More on DNA, States Take a Closer Look*, USA TODAY,
unsuccessfully chasing a wounded felon might find sufficient blood has dripped onto the sidewalk for DNA profiling to be conducted. It could be argued that such activity is not a search (and hence requires neither probable cause nor a warrant) because the individual, having abandoned the material in a public place, retains no reasonable expectation of privacy in it. The Supreme Court used this reasoning in *California v. Greenwood*\(^\text{132}\) in holding that the Fourth Amendment does not prohibit “the warrantless search and seizure of garbage left for collection outside the curtilage of a home.”\(^\text{133}\) The Court commented:

> It is common knowledge that plastic garbage bags left on or at the side of a public street are readily accessible to animals, children, scavengers, snoops, and other members of the public . . . . Moreover, respondents placed their refuse at the curb for the express purpose of conveying it to a third party, the trash collector, who might himself have sorted through respondents’ trash or permitted others, such as the police, to do so.\(^\text{134}\)

However, depositing DNA in the ordinary course of life when drinking, sneezing, or shedding hair,\(^\text{135}\) dandruff,\(^\text{136}\) or other cells\(^\text{137}\) differs from placing private papers in a container on the street to be collected as garbage. Depositing paper in the trash is generally a voluntary act. Someone intent on preserving the secrecy of the papers can shred the papers or dispose of them in other ways that would defeat normal police surveillance. Leaving a trail of DNA, however, is not a conscious activity. The deposition of DNA in public places cannot be avoided unless one is a hermit or is fanatical in using extraordinary


\(^{133}\) Id. at 37.

\(^{134}\) Id. at 40.

\(^{135}\) See R. Higuchi et al., *DNA Typing from Single Hairs*, 332 NATURE 543 (1988).


containment measures. In this setting, the inference of intent to abandon is markedly weaker.

If the police collection of inadvertently deposited DNA cannot be justified solely on an abandonment theory, under Katz the question becomes whether society does or should recognize as reasonable the expectation that government agents will not follow one about to obtain and analyze DNA that almost inevitably is left in public places.138 A case can be constructed that such an expectation exists. The public is extremely concerned with preserving genetic privacy. Many states have enacted legislation to preserve the confidentiality of genetic information, and a few have even labeled a person’s genotypes as the property of the individual.139 Furthermore, Skinner v. Railway Labor Executives’ Ass’n140 lends support to this argument. In Skinner, the Federal Railroad Administration had promulgated regulations mandating blood and urine tests of employees involved in certain train accidents and authorizing the railroads to administer breath and urine tests on employees who violated particular safety rules. Some provisions authorized breath and urine tests based on a “reasonable suspicion” of drug or alcohol impairment, but others did not require any showing of individualized suspicion. Railway employees alleged that this system violated their Fourth Amendment rights. The Court of Appeals for the Ninth Circuit invalidated the regulations, holding that the drug testing required reasonable suspicion.141 The U.S. Supreme Court reversed the court of appeals, reasoning that the compelled collection of breath and urine samples was a search but the practice was reasonable because the government had a special need to ensure that railway personnel were not using substances that might cloud their judgment and impair their performance.142 The majority recognized

138. It might seem that police have no more obligation to secure a warrant before seizing DNA in a place they have right to be than they do to secure a warrant or other order before dusting a crime scene for fingerprints, viewing the site of a rape for semen, or looking for traces of blood on or near the corpse of person who obviously struggled with an assailant. However, these efforts to locate physical or biological evidence that later can be traced to an individual are quite different from the practice of following an individual already suspected of a crime to pick up that person’s DNA. In the former situation, the involvement of a magistrate would be pointless, for there is nothing for a magistrate to consider. In the latter, the magistrate could screen the information the police already have to ensure that there is sufficient reason to invade the target’s genetic privacy.


141. Id. at 612–13.

142. Id. at 620.
that “[u]nlike the blood-testing procedure at issue in Schmerber, the procedures prescribed by the . . . regulations for collecting and testing urine samples do not entail a surgical intrusion into the body.”

Nonetheless, the opinion concluded that urine sampling followed by urinalysis was a search, in part because “chemical analysis of urine, like that of blood, can reveal a host of private medical facts about an employee, including whether he or she is epileptic, pregnant, or diabetic.” Similarly, DNA testing can reveal “private medical facts” about the individual.

However, Skinner is distinguishable in that urinalysis involves a much more extensive intrusion into privacy: the possible revelation of private information, compelled excretion of bodily fluid, and monitoring the normally private act of excretion. Unlike the collection of blood from the suspect in Schmerber, collecting DNA left in public places entails neither a bodily invasion nor a seizure of the person. It seems clear that, in a public restaurant after a suspect departed, the police could pick up a coffee cup used by the suspect and, consistent with the Fourth Amendment, examine it for fingerprints. Courts may find it a small step to conclude that the warrantless collection of inadvertently abandoned DNA does not violate the Fourth Amendment.

Before taking this step, however, courts should consider the extent to which meaningful, personal information that would not be available to private citizens will fall into the hands of government agents interested in accessing this information. When society enters an era in which DNA analyzers are as accessible as home pregnancy-test kits, the argument for an expectation of privacy will be weak. But in a world still at the

143. Id. at 617.
144. Id.
145. See United States v. Nicolosi, 885 F. Supp. 50, 56 (E.D.N.Y. 1995) (relying on this facet of informational privacy to conclude that compelling a saliva sample for DNA testing is a search or seizure).
146. See id. at 55–56.
147. Research revealed no case holding, or even implying, that the collection of a fingerprint in a public area amounts to a Fourth Amendment intrusion in fact. After a person has been arrested, the collection of fingerprints directly from the person with no showing beyond the validity of the arrest itself is routinely upheld. See Napolitano v. United States, 340 F.2d 313, 314 (1st Cir. 1964); Smith v. United States, 324 F.2d 879, 882 (D.C. Cir. 1963); State v. Inman, 301 A.2d 348, 354 (Me. 1973); Commonwealth v. Young, 572 A.2d 1217, 1224 (Pa. 1990). For cases applying the abandonment theory to other materials that are at least as problematic as fingerprints, see United States v. Cox, 428 F.2d 683, 687–88 (7th Cir. 1970) (hair cut by prison barber); Venner v. State, 367 A.2d 949, 955–56 (Md. 1977) (contents of hospital bedpan).
threshold of an age of molecular biology, what expectation is reasonable is less obvious. Privacy expectations should turn on the incentives and disincentives for the government to acquire DNA information that is truly sensitive as well as the risk that this information will be used to harm individuals. For law enforcement purposes, there is little incentive to probe areas of the genome that would determine characteristics not discernible to individuals acquainted with a suspect. Identification rarely will be aided by disease prediction, for example, and the risk that police or laboratory personnel will be curious about and well positioned to collect the latter sort of information does not seem large. Rather, there are disincentives to such investigations. The existence of numerous laws restricting the use and release of genetic information on individuals by insurers, employers, and law enforcement personnel is pertinent here.\(^{148}\) Although further experience with DNA samples in the criminal justice system could lead to a reassessment, for the present the better course is to treat human cells left in public places like fingerprints in deciding what expectation of privacy is reasonable.

D. Securing the Consent of Suspects or Others

I. Voluntariness Under the Fourth Amendment

In addition to compelling individuals to submit DNA samples or acquiring samples indirectly from medical-care providers, researchers, or from other locations, the authorities might simply ask a suspect to provide a sample. Even if the acquisition of the sample constitutes a seizure under the Fourth Amendment and the authorities do not obtain a warrant or court order, a suspect’s consent is an adequate justification for a search or seizure.\(^{149}\) What, then, must authorities do to obtain legally effective consent? When that question arose with respect to waivers of the privilege against self-incrimination embodied in the Fifth Amendment, early Supreme Court cases applied a general voluntariness test.\(^{150}\) However, in 1966, the Court mandated in *Miranda v. Arizona*\(^{151}\)

---

148. *See*, e.g., MD. ANN. CODE, art. 88B, § 12A(p) (2000) (creating criminal penalties for unauthorized disclosure or receipt of “individually identifiable DNA information contained in the statewide DNA data base system or … repository”); Kaye, *supra* note 139, at 1–2 (collecting references to genetic privacy and discrimination legislation).


that the police administer specific warnings to a suspect in custody to ensure that any consent to interrogation was voluntary in the specific sense that the suspect’s waiver was intelligent and knowing.\textsuperscript{152} In particular, police must advise a suspect of the right to remain silent.\textsuperscript{153}

The Court has taken a different approach to waivers of one’s Fourth Amendment right to be free from unreasonable searches and seizures. In \textit{Schneckloth v. Bustamonte},\textsuperscript{154} the Court confirmed the continued applicability of a general voluntariness test and expressly held that police need not warn a suspect of the right to refuse to consent to a search.\textsuperscript{155} The administration of such a warning is simply one factor in the totality of the circumstances that must be considered in determining the voluntariness of the consent.\textsuperscript{156} \textit{Schneckloth} sharply distinguished between the Fourth and Fifth Amendment settings. The Court emphasized that while \textit{Miranda} warnings helped to ensure the reliability of any confession by a suspect, the Fourth Amendment exclusionary rule has little or nothing to do with the reliability of the seized evidence.\textsuperscript{157} According to the Court, concerns about the reliability of the evidence and the integrity of the fact-finding process justify a more rigorous standard under the Fifth Amendment than under the Fourth.\textsuperscript{158}

Although the \textit{Schneckloth} standard is more lax than the Fifth Amendment test,\textsuperscript{159} even \textit{Schneckloth} has teeth. In some cases, Fourth Amendment consent has been found involuntary. For example, in \textit{Bumper v. North Carolina},\textsuperscript{160} four police officers went to the house of “a 66-year-old Negro widow . . . located in a rural area at the end of an isolated mile-long dirt road.”\textsuperscript{161} She met the officers at the front door.


\textsuperscript{152} 384 U.S. 436 (1966).

\textsuperscript{153} \textit{Id.} at 445–58, 475–76.

\textsuperscript{154} 412 U.S. 218 (1973).

\textsuperscript{155} \textit{Id.} at 231–32.

\textsuperscript{156} \textit{Id.} at 226–27.

\textsuperscript{157} \textit{Id.} at 241–48.

\textsuperscript{158} \textit{Id.} at 246–49.

\textsuperscript{159} Under a state’s constitution, a state court can adopt a higher standard and require warnings about the search to be administered for consent to be valid. See Cooper v. California, 386 U.S. 58, 62 (1967); Commonwealth v. Walsh, 460 A.2d 767, 771 (Pa. Super. Ct. 1983).

\textsuperscript{160} 391 U.S. 543 (1968).

\textsuperscript{161} \textit{Id.} at 546.
One of them announced, “I have a search warrant to search your house” she responded, “Go ahead,” and opened the door. In the kitchen, the officers found the rifle that was later introduced in evidence at the trial of her grandson for rape. The Court reversed the conviction because the officers had no search warrant. The Court explained: “When a law enforcement officer claims authority to search a home under a warrant, he announces in effect that the occupant has no right to resist the search. The situation is instinct with coercion—albeit colorably lawful coercion. Where there is coercion there cannot be consent.”

Under Bumper, consent to submitting a DNA sample would be involuntary if, for instance, the police gave the suspect the impression that there was no alternative other than to provide the sample. However, on balance, consent should be deemed sufficient when the police make it clear that they seek a sample “for criminal investigation purposes” and avoid statements that could mislead the suspect into believing that there is a legal duty to furnish the sample when there is none.

Furthermore, under Scheckloth, the suspect can protect his or her privacy by limiting the scope of the consent. In Florida v. Jimeno, the Supreme Court acknowledged that consent to an intrusion may be limited in scope. For example, a suspect might authorize the warrantless search of an automobile but refuse to consent to a search of the person. By the same token, a suspect could consent to furnishing a DNA sample only for identification purposes in connection with a specific investigation. It would be helpful if police obtained written or recorded

162 Id.
163 Id. at 548–50.
164 Id. at 550.
165 Commonwealth v. Walsh, 460 A.2d 767, 773 (Pa. Super. Ct. 1983); see also State v. Bailey, 417 A.2d 915, 919 (R.I. 1980) (“The notion of a free and voluntary consent necessarily implies that the person knows what it is he is allowing the police to do.”). It has been held that it is not necessary to inform a suspect of the specific investigations for which the sample is desired.
166 People v. Cardenas, 604 N.E.2d 953, 956 (Ill. App. Ct. 1992) (holding consent to be involuntary when the officer created the false impression that there could be a search without the suspect’s consent).
168 Id. at 251.
169 See Strong v. United States, 46 F.2d 257, 259 (1st Cir. 1931); Witt v. Commonwealth, 293 S.W. 1072, 1073 (Ky. 1927); State v. Johnson, 71 Wash. 2d 239, 427 P.2d 705 (1967).
170 Where the individual does not state that the only use of the sample is to be in connection with a specific investigation, police may be tempted to reuse the information in a subsequent investigation. Indeed, they may even acquire a sample ostensibly for one investigation when their
consent for DNA sampling, but there does not appear to be any constitutional requirement to do so.  

2. Geographic Screening

Some European countries have solved difficult murder cases by appealing to all local residents to submit to DNA testing. In perhaps the largest such geographic screening, 16,400 men in western Germany were tested in the hunt for an eleven-year-old girl’s killer. A thirty-year-old man arrested after his DNA was found to match confessed to raping, stabbing, and killing the girl, as well as to raping another eleven-year-old girl. In the United Kingdom alone, police have conducted 118 such “mass screens,” resulting in forty-eight hits and seven convictions.

actual intent is to use it in another. For instance, in Washington v. State, 653 So. 2d 362, 363 (Fla. 1994), a ninety-three-year-old woman was murdered in her bedroom. She had been badly beaten and vaginally and anally raped. Id. Anthony Washington was imprisoned at a work-release center two miles from the woman’s home. Id. He did not show up at his job during the time of the rape, and he sold the woman’s gold watch to a coworker. Id. The detective investigating the murder did not tell Washington that he suspected him of this murder. Id. Instead, he asked Washington for blood and hair samples to use in an unrelated sexual battery case. Id. at 364. Washington provided these samples. Id. When the state sought to use the samples in the murder case, Washington moved to suppress them. Id. The trial court denied the motion, and Washington was convicted of the murder, burglary, and sexual battery. Id. The Supreme Court of Florida affirmed the conviction, reasoning: “Washington stated he understood his rights, orally waived them, and freely and voluntarily provided [the detective] with hair and blood samples. [O]nce the samples were validly obtained, albeit in an unrelated case, the police were not restrained from using the samples as evidence in the murder case.” Id. Applied to situations in which it is reasonably clear that the DNA donor meant to consent to only one use, additional uses of the sample may be problematic. The better rule here might be to presume that consent does not extend beyond the specific investigation called to the donor’s attention. Examples of such consent forms are included in CECELIA CROUSE & D.H. KAYE, THE RETENTION AND SUBSEQUENT USE OF SUSPECT, ELIMINATION, AND VICTIM DNA SAMPLES OR RECORDS: A REPORT TO THE NATIONAL COMMISSION ON THE FUTURE OF DNA EVIDENCE (2000).

171. This practice not only would safeguard the suspect’s rights, but it also would simplify the task of a judge ruling on a motion to suppress evidence acquired as a result of the search.


174. Id.

175. David Werrett, The Strategic Use of DNA Profiling, Address Before the 18th International Congress on Forensic Haemogenetics (Aug. 19, 1999). In these investigations, explicit permission
Geographic screening has been used on a smaller scale in Ann Arbor, Miami, and San Diego. Nothing in the Fourth Amendment prevents the police from approaching everyone in a community and asking for their cooperation. The “dragnet” nature of the inquiry is no obstacle. Consensual contact between a police officer and a citizen is was obtained from the individuals who provided DNA samples, and the samples from people who are eliminated as possible suspects were destroyed. Id. In late 1999, however, section 64 of the Police and Criminal Evidence Act was amended to permit the retention and use of DNA samples and data with a volunteer’s written permission. Home Office, Proposals for Revising Legislative Measures on Fingerprint, Footprints and DNA Samples 11 (July 1999).


177. See Philip P. Pan, Pr. George’s Chief Has Used Serial Testing Before; Farrell Oversaw DNA Sampling of 2,300 in Fla., WASH. POST, Jan. 31, 1998, at B1 (reporting that for “over two months, [Miami] police” searching for the Tamiami Strangler, a man who had been murdering prostitutes, “stopped about 2,300 men driving through an area known for prostitution [along the Tamiami Trail], asked them to voluntarily provide saliva samples,” and “followed up on everybody who was a preliminary match and everybody who refused to consent voluntarily”). This effort did not unearth the murderer. He was found after he left a thirty-year-old prostitute in his apartment, naked, bound with duct tape, and screaming for help. DNA testing of a sample from this woman matched that of five of the previous murder-rapes. See Lynn Carrillo, Suspect Fit Serial Killer Portrait: DNA Link to Prostitutes Helps Bolster Case, But Man’s Neighbors Stunned, SUN-SENTINEL (Ft. Lauderdale), June 28, 1995, at 1B.

178. The San Diego police department “tested about 800 men during its search for a serial killer who stabbed six women to death in their homes between January and September 1990.” As described in Pan, supra note 177, at B1:


179. Cf. United States v. Dionisio, 410 U.S. 1, 9 (1973) (finding no infirmity in issuing, without probable cause, grand jury subpoenas to twenty men to secure voice exemplars from them); Davis v. Mississippi, 394 U.S. 721 (1969) (containing no suggestion that the compulsory fingerprinting of twenty-four young men was unconstitutional merely because there was no reason other than their fitting the very general description provided by the victim).
neither a search nor a seizure under the Fourth Amendment. As a legal matter, police may ask anyone to give DNA and, as long as they do not engage in coercion or misrepresentation, the police may collect voluntary samples for analysis. The practice seems less likely to be effective in this country, however, because the number of residents who could choose not to cooperate might be larger. In addition, the drain on police resources in creating what is, in effect, an ad hoc database would be excessive.

E. Inferring Physical Characteristics from Crime-Scene Samples

To help trace the flow of human populations, geneticists and anthropologists have located genetic markers that help distinguish among ancestral populations, and various genes or other DNA sequences are known to have alleles that occur predominantly in certain racial or ethnic groups. In addition, genetic typing permits inferences as to

180. Cf. Davis, 394 U.S. at 727 n.6 (referring to "the settled principle that while the police have the right to request citizens to answer voluntarily questions concerning unsolved crimes they have no right to compel them to answer").

181. See supra Part I.D.1. For a description of tactics that may have crossed the line from consent to coercion, see Fred W. Drobner, Comment, DNA Dragnets: Constitutional Aspects of Mass DNA Identification Testing, 28 CAP. U. L. REV. 479, 505–06 (2000).

182. It would be possible to retain the samples of volunteers for possible later use, given adequate consent. The constitutionality of establishing more permanent, population-wide databases is addressed in Kaye & Imwinkelried, supra note 7, at III.C.2.

183. Human DNA is packaged into twenty-three pairs of chromosomes in cell nuclei. The DNA molecules in the chromosomes are composed of four types of units called "nucleotide base pairs." On average, a chromosome is on the order of 100 million base-pairs long. A "locus" is a location on a chromosome. Some loci involve sequences that are, in effect, instructions for manufacturing specific proteins. These are known as genes. Other loci are not involved in the production of any proteins, and hence, do not produce any observable traits (such as blood types). The distinct variations at the loci that code for proteins—the genes—are known as alleles of those genes. For example, the gene for red blood cell types has four major alleles. By extension, the variations in the order or number of base pairs at the noncoding loci are called DNA alleles. See, e.g., Kaye & Sensabaugh, supra note 9, at 491.

Although some 99.9% of the sequence of base-pairs are identical among individuals, some do exhibit substantial variation from one person to another. Id. Among these alleles is a class known as "short tandem repeats," or STRs. These noncoding regions vary in length. Id. at 494. The shortest have only one or two sets of a specific DNA sequence that is itself between two to seven base pairs long, but the typical STR allele has enough repetitions of the core sequence so that it is 50 to 350 base pairs long. Id. at 494 n.38. STRs probably are the most widely used system of alleles for forensic DNA identification. See id. at 494.

inherited disorders and may offer clues to facial or other bodily features. Learning that the person whose DNA was found at a crime scene might have such physical characteristics could well be useful in some criminal investigations.

DNA analysis to conduct such “physical profiling” poses few constitutional problems. The principal issue arises under the Equal Protection Clause of the Fourteenth Amendment. Normally, the government is free to draw whatever reasonable lines it wishes in adopting and enforcing the law. The Internal Revenue Service, for instance, can choose to concentrate its investigations of tax evasion on higher-income taxpayers. Some classifications, however, are suspect. Race is the prototypical example. Imposing the death sentence on the killers of whites but not blacks would be impermissible; likewise, a police officer who adopted a policy of arresting only African-Americans would be depriving those citizens of the equal protection of the law. Does the fact that race is a suspect classification prohibit the government

---

History, 286 SCIENCE 451 (1999); M. Klintschar et al., Is It Possible to Determine the Ethnic Origin of Caucasian Individuals Using Short Tandem Repeat Loci?, in ABSTRACTS OF THE EIGHTEENTH INTERNATIONAL CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC HAEMOGENETICS 63 (1999) (“It is possible to differentiate between major ethnic groups using forensically relevant short tandem repeat loci.”).

185. See, e.g., NCFDNA, supra note 184, at 61 (“Determining that a DNA sample was left by a person with red hair, dark skin pigment, straight hair, baldness, or color blindness may be practical soon, if not already.”; 1 M. McCulley et al., Genes and Faces: Classification of Midline Features, in ABSTRACTS OF THE EIGHTEENTH INTERNATIONAL CONGRESS OF THE INTERNATIONAL SOCIETY FOR FORENSIC HAEMOGENETICS 109 (1999).

186. See, e.g., Korematsu v. United States, 323 U.S. 214, 216 (1944). Laws or other state actions that distribute burdens or benefits according to “suspect” classifications are usually, but not invariably, unconstitutional. Id. (finding a compelling government interest in confining Japanese-Americans during World War II). The government bears an extremely heavy burden of showing that the classification is necessary to achieve a compelling state interest. See, e.g., In re Griffiths, 413 U.S. 717, 721 (1973). Certain other classifications, such as gender, are “quasi-suspect.” Craig v. Boren, 429 U.S. 190 (1976). Their use can be justified by a showing of a close enough connection to an important state interest. Id. The sex of the source of a biological sample from a crime scene can be ascertained with far greater confidence than can the race of the source. Although this Article does not discuss the use of genetic tests for inferring gender in criminal investigations, the analysis of the constitutionality of testing for race is applicable to testing for sex as well.

188. See McCleskey v. Kemp, 481 U.S. 279, 291 n.8 (1987) (“It would violate the Equal Protection Clause for a State to base enforcement of its criminal laws on an unjustifiable standard such as race, religion, or other arbitrary classification.”) (internal quotation marks and citation omitted).

189. Cf. Hall v. Pa. State Police, 570 F.2d 86, 90–91 (3d Cir. 1978) (reasoning that allegations that the state police directed banks to photograph only blacks entering the banks stated a violation of equal protection).
from conducting or funding research to develop or refine genetic markers for racial identification? Or, when it appears from such markers that the source of the crime-scene DNA is likely to belong to a particular racial or ethnic group, does the Equal Protection Clause prohibit the police from using that fact as an investigative lead and focusing on members of that group?

These questions require an examination of the purpose and impact of the racial classification. That race is a suspect classification does not mean that the government never can inquire into race.190 To the contrary, the collection and analysis of information about race are commonplace in enforcing the law and in criminological research undertaken or funded by the government.191 More generally, a great deal of social science and medical research supported or conducted by the government involves the collection of data on race and the analysis of race as a variable of interest.192 Likewise, if using physical evidence of race to focus an investigation were impermissible, police could not rely on an eyewitness’s report that a person fleeing the scene of a crime was Hispanic, on a victim’s report that a rapist was white, or on a linguist’s analysis of accent or word choice in a recorded death threat that suggested that the caller was African-American.193 These reports could be in error in any given case, but if they are generally accurate, paying attention to them is not unconstitutional.194


193. Testimony as to ethnic and racial identifications routinely is admitted in criminal trials. See, e.g., United States v. Card, 86 F. Supp. 2d 1115, 116–19 (D. Utah 2000) (reviewing cases on admissibility of lay testimony of race or ethnicity based on speech patterns).

194. There are few cases on the constitutionality of using race as an identifying feature in criminal investigations, presumably because the validity of the practice always has been taken for granted. See, e.g., Waldron v. United States, 206 F.3d 597, 604 (6th Cir. 2000) (holding that an investigative stop of a bank robber that relied in part on witnesses’ statements as to the robber’s race was not “illegal ‘racial targeting’ or ‘racial profiling’ because ‘[c]ommon sense dictates that, when determining whom to approach as a suspect of criminal wrongdoing, a police officer may legitimately consider race as a factor if descriptions of the perpetrator known to the officer include race’”). In Hall, the Third Circuit held that the Fourteenth Amendment prohibited the police from asking banks to photograph suspicious-looking blacks who entered their establishments. Hall, 570
The government can rely on racial information in a criminal investigation because the practice does not unfairly burden any racial group.\footnote{Cf. Strauder v. West Virginia, 100 U.S. 303, 308 (1879) (stating that the Fourteenth Amendment confers “the right to exemption from unfriendly legislation against [a group] distinctively” on account of race). The continuing vitality of \textit{Strauder} cannot be doubted. \textit{J.E.B. v. T.V.}, 511 U.S. 127, 128 (1994); \textit{Georgia v. McCollum}, 505 U.S. 42, 44 (1992); \textit{Powers v. Ohio}, 499 U.S. 400, 402 (1991). Consider also the use of racial information by government doctors on the staff of Veterans Administration hospitals. Suppose that a physician who is considering administering the drug isoniazid to an Asian patient who is either has pulmonary sarcoidosis or pulmonary tuberculosis. \textit{See Richard I. Kopelman et al., Clinical Problem Solving: A Little Math Helps the Medicine Go Down}, 341 NEW ENG. J. MED. 435 (1999). Although this therapy improves the chance of survival for tuberculosis (but not sarcoidosis), it can lead to hepatitis. \textit{Id.} at 437. The risk of this complication is about twice as large in Asian men as in other men. \textit{Id.} In weighing the therapeutic benefits against the risks it would be medically inadvisable for the physician to disregard the fact that the patient is Asian. \textit{See id.} at 437–38. Therefore, it is legal for the physician to use race as a factor in diagnosis, prognosis, or treatment.} In cases in which racial classifications have been struck down, the explicit purpose or actual use of the racial classification was to burden or stigmatize a racial group. \textit{Yick Wo v. Hopkins}\footnote{118 U.S. 356 (1886).} offers an early illustration. In 1880, San Francisco passed an ordinance requiring that persons obtain a permit before operating laundries in wooden structures. Yick Wo was convicted of operating such a laundry without a permit. The Supreme Court set aside the conviction because it concluded that city officials had issued permits with “an evil eye and an unequal hand.”\footnote{Id. at 373–74.} Almost without exception, permits were denied to Chinese and granted to non-Chinese.\footnote{Id. at 90–91.} Thus, the permit requirement, although not explicitly racial, was used to exclude the Chinese from the laundry business. It burdened this group for no legitimate reason.

A different question is presented when racial information is collected and used evenhandedly to advance legitimate state interests and in ways that are not designed to disadvantage any individual because of race. In these situations, courts have held the government can record the racial
DNA Typing Issues

information. For instance, *Hamm v. Virginia State Board of Elections* involved an equal protection challenge to a Virginia law that required every decree of divorce to recite the race of the spouses. A three-judge district court upheld this record-keeping provision because the racial information served the valid purpose of collecting social statistics and did not single out or burden any racial group. The Supreme Court affirmed without discussion.

Under these principles, governmental sponsorship of research on the variations of particular alleles across races and the investigative use of alleles that are reasonably accurate indicators of race should pass constitutional muster. Two factors are crucial to such constitutionality: No group is singled out for special treatment, and no one is penalized because of hostility toward race. If the police make investigative use


200. The court wrote:

[T]he designation of race, just as sex or religious denomination, may in certain records serve a useful purpose, and the procurement and compilation of such information by State authorities cannot be outlawed per se. For example, the securing and chronicling of racial data for identification or statistical use violates no constitutional privilege. If the purpose is legitimate, the reason justifiable, then no infringement results . . . .

Vital statistics, obviously, are aided by denotation in the divorce decrees of the race of the parties. This provision . . . is not objectionable in law. Of course, the advertence must be made in every case, not just in suits involving Negroes.

Id. at 158. Other laws required lists of qualified voters and property owners to be maintained so as to reveal the race of the individuals. The court struck down these requirements because “they serve no other purpose than to classify and distinguish official records on the basis of race or color.”


202. Where the government publicizes the race of individuals in a way calculated to disadvantage members of certain racial groups, it offends the Fourteenth Amendment. Thus, in *Anderson v. Martin*, 375 U.S. 399 (1964), individuals wishing to become candidates in an election to the county school board of East Baton, Louisiana, brought an action to enjoin the state from enforcing a statute that required nomination papers and ballots to designate the race of candidates for elective office. *Id.* at 401. On its face, the law applied to all races and was applied evenhandedly. Nonetheless, its only plausible purpose was to invite racial cohesion in elections. As the Court explained:

[By directing the citizen’s attention to the single consideration of race or color, the State indicates that a candidate’s race or color is an important—perhaps paramount—consideration in the citizen’s choice, which may decisively influence the citizen to cast his ballot along racial lines. Hence in a State or voting district where Negroes predominate, that race is likely to be favored by a racial designation on the ballot, while in those communities where other races are in the majority, they may be preferred. The vice lies not in the resulting injury but in the placing of the power of the State behind a racial classification that induces racial prejudice at the polls.}
of racial information whenever that information is useful, then all racial
groups are treated alike; none is stigmatized or disadvantaged in the
enforcement of laws that apply with equal force to members of every
race.

It is true that the information could have a disparate impact and lead to
the apprehension of more criminals from one race than another—but
not because of official (or even unofficial) hostility toward particular
races or individual prejudices about those races. Recent years have
witnessed outcries against “racial profiling” in policing.203 The
constitutional defect in this practice is that authorities unfairly target
minorities for traffic stops or arrests. As in Yick Wo, laws that are neutral
on their face—that do not explicitly classify people by race—can be
applied disproportionately to racial minorities.204 With DNA samples
from crime scenes, however, statistically valid inferences as to race

Id. at 402.

The use of race as an identifying feature in criminal investigations is quite different. It is not
designed to induce or capitalize on racial prejudice, but rather serves the legitimate purpose of
enabling authorities to identify and apprehend the perpetrators of crimes. Moreover, as discussed in
the text below, the use of DNA markers could serve to reduce the operation of racial prejudice by
providing an objective indication of the race of the perpetrator.

These benign characteristics are present even if the DNA genotypes associated with race
distinguish among but a handful of races. For example, suppose (unrealistically) that only two
groups could be distinguished with reasonable accuracy—individuals of African descent and those
of European descent. Deciding that the crime samples in one set of cases probably originated from
one group, while those in another set of cases probably came from the other group, would represent
an evenhanded application of the available information to all people. It is comparable to using black-
and-white surveillance videos of bank robberies. As long as the police consistently focus
investigations on individuals with the complexion of those depicted on film, they do not
purposefully disadvantage any one group, and the practice would not constitute the invidious
discrimination forbidden by the Equal Protection Clause.

203 See Jeffrey Goldberg, The Color of Suspicion, N.Y. TIMES MAG., June 20, 1999, at 51; Jerry
Gray, Unwelcome Addition to a Legacy, N.Y. TIMES, May 1, 1999, at A5; Steven A. Holmes,
Clinton Orders Investigation on Possible Racial Profiling, N.Y. TIMES, June 10, 1999, at A22; Steve
Miller, Ashcroft Demands Profiling Study: Gives Congress Six Months or Says He Will Do It

204 This risk can be intolerably high when race is itself the factor on which officers rely.
Consequently, the Equal Protection Clause may require the use of factors—gang membership, for
example—that are not explicitly racial, even though these factors can be highly correlated with
race. It is well settled that the use of valid decision-making factors does not violate the Equal
Protection Clause even when they have a disparate impact on one race. Hunt v. Crowmartie, 526
U.S. 541, 546 (1999) (“A facially neutral law, on the other hand, warrants strict scrutiny only if it
can be proved that the law was motivated by a racial purpose or object, or if it is unexplainable on
grounds other than race.”) (internal quotation marks and citations omitted); Washington v. Davis,
426 U.S. 229 (1976). When the underlying, nonracial factors are available, they are more accurate
indicators than the racial surrogate for them.
cannot lead the authorities to target minorities because of subjective racial stereotypes or prejudices. If anything, by focusing the investigation on the pertinent physical characteristics—whatever they may be—reliance on genetic information in crime-scene samples could correct any tendency to pursue one racial group exclusively or disproportionately. For example, if DNA analysis indicated that the source of a sample was more likely to be Caucasian than African-American, it might help overcome a stereotypical assumption that only blacks need be considered as prime suspects. By providing objective information, DNA analysis could serve as an antidote to the objectionable form of “racial profiling” in police work.

Of course, this is not to say that the government should institute a research program to develop more precise DNA markers for racial identity or that police should use existing markers that are demonstrated to be informative as to race. The conclusion is simply that these are policy choices to be made about a developing technology—these options are not foreclosed by the Constitution.

II. DNA ANALYSIS IN PROSECUTIONS

A. Standards and Procedures for Deciding Admissibility

1. Novel Scientific Methods

Two major standards exist for deciding whether scientific findings will be admitted into evidence: the “general-acceptance” test and the “sound-methodology” standard. If a timely objection is raised, the judge must determine whether the applicable standard has been met. The general-acceptance standard was first articulated in an influential 1923 federal court of appeals case, Frye v. United States. In jurisdictions that follow Frye, the proponent of the scientific evidence typically must establish that the underlying theory and methodology are generally accepted within the relevant portions of the scientific community.

The sound-methodology standard is derived from phrases in the Federal Rules of Evidence. In Daubert v. Merrell Dow Pharmaceuticals,
Inc., the U.S. Supreme Court held that these rules implicitly jettison general acceptance as an absolute prerequisite to the admissibility of scientific evidence. Instead of the *Frye* test, the Court prescribed a broader framework for deciding whether proposed testimony has sufficient scientific validity and reliability to be admitted as relevant “scientific knowledge” that would “assist the trier of fact.” In that framework, the lack of general acceptance weighs against admissibility but is not invariably fatal. It is circumstantial evidence that the technology has not been studied widely or that the bulk of the specialists who have studied the technique have found it wanting. The Court discussed other factors that might be considered. Its nonexhaustive list includes the extent to which the theory and technology have been tested, the existence of a body of peer-reviewed studies, and the known error rates of the procedure. Although the trend in the states appears to be toward the *Daubert* view, there still are jurisdictions that adhere to *Frye*.

Labels like “general acceptance” and “sound methodology” are just that—labels. Cases decided in each jurisdiction help to define the scientific community in which the degree of scientific acceptance is to be ascertained, the extent of disagreement that can be tolerated, the information that may be used to gauge the extent of consensus, and the specific factors other than general acceptance that bear on relevance and helpfulness. The degree of scientific consensus is important to the admissibility of scientific evidence in all jurisdictions, and pretrial hearings in hotly contested cases have lasted months and generated thousands of pages of testimony probing the opinions of experts on various aspects of DNA profiling. The courts have examined affidavits or testimony from scientists selected by the parties, specific papers in

---

209. *Id.* at 589.
210. *Id.* at 592.
211. *Id.* at 593–94.
scientific periodicals, the writings of science journalists, the body of court opinions, and other scientific and legal literature.

In particular, the history of the judicial treatment of DNA evidence can be divided into at least five phases. The first phase was one of rapid and sometimes uncritical acceptance. The first generation of DNA typing tests examined certain Restriction Fragment Length Polymorphisms (RFLPs) known as Variable Number Tandem Repeats (VNTRs). Bacterial “restriction enzymes” can be used to cut the strands of the DNA molecule when they encounter certain, short sequences of bases (“restriction sites”). The resulting fragments vary in their lengths; some have more base pairs between two adjacent restriction sites. One class of such length polymorphisms is due to repetitions of “core sequences” some fifteen to thirty-five base pairs long. These core sequences are repeated end-to-end different numbers of times in different individuals. Because of the many repeats, these VNTR alleles typically extend for thousands of bases. Between each pair of restriction sites, an individual usually has two fragments of distinct lengths (one from each chromosome). The two lengths can be measured by electrophoresis, a process that sorts fragments by length.

Initial praise for RFLP testing in homicide, rape, paternity, and other cases was effusive. Indeed, one judge proclaimed “DNA fingerprinting” to be “the single greatest advance in the ‘search for truth’ . . . since the advent of cross-examination.” In this first wave of cases, expert testimony for the prosecution rarely was countered, and courts readily admitted RFLP findings.

214. This history of the judicial reception of DNA evidence is adapted from 1 CHARLES T. MCCORMICK, MCCORMICK ON EVIDENCE § 205 (John W. Strong ed., 5th ed. 1999).

215. The nucleotide bases that connect the two helical strands of the DNA molecule often are referred to by the first letters (A, T, C, and G) of their names. The A on one strand always pairs with the T on another, and the C with a G, so that the base pairs are like a spiral staircase with four types of rungs (AT, TA, CG, GC). See, e.g., D.H. Kaye, Bible Reading: DNA Evidence in Arizona, 28 ARIZ. ST. L.J. 1035, 1044–45 (1996).

216. See, e.g., Kaye & Sensabaugh, supra note 9, at 494.

217. See id. at 562.

218. See, e.g., id. at 501.


220. See, e.g., Andrews v. State, 533 So. 2d 841, 847–51 (Fla. Dist. Ct. App. 1988); Wesley, 533 N.Y.S.2d at 643–59; State v. Woodall, 385 S.E.2d 253, 259–62 (W. Va. 1989) (taking judicial notice of general scientific acceptance where there was no expert testimony, but holding that
In a second wave of cases, however, defendants pointed to problems at two levels—controlling the experimental conditions of the analysis and interpreting the results. Some scientists questioned certain features of the procedures for extracting and analyzing DNA employed in forensic laboratories. It became apparent that determining whether RFLPs in VNTR loci in two samples actually match can be complicated by measurement variability or missing or spurious bands. Despite these concerns, most cases continued to find forensic RFLP analyses to be generally accepted, and a number of states have provided for admissibility of DNA tests by legislation. Concerted attacks by inconclusive results were properly excluded as irrelevant); Thomas M. Fleming, Annotation, Admissibility of DNA Identification Evidence, 84 A.L.R.4th 313 (1991).

221. For a comprehensive survey of possible sources of error and ambiguity in VNTR profiling, see William Thompson & Simon Ford, The Meaning of a Match: Sources of Ambiguity in the Interpretation of DNA Prints, in FORENSIC DNA TECHNOLOGY 93 (Mark A. Farley & James J. Harrington eds., 1990).

222. Electrophoresis involves the application of an electric field that pulls shorter fragments through a gelatinous material more rapidly than longer fragments. Thus, in a fixed time, shorter alleles run farther on the “gel,” as it is called, than the longer fragments. But the separation process is not exact. Two fragments that are very close in size may not be distinguishable, and the distance that a fragment travels varies slightly from one run to the next. Because of this measurement error inherent in the system the lengths are not known precisely. See, e.g., NRC II, supra note 1, at 139–42.

223. When the many copies of single-stranded DNA fragments extracted from a sample of many cells have been separated on the electrophoretic gel, they are transferred to a sheet of nylon that is easier to manage. The restriction fragments of interest are identified with a “probe”—a short stretch of single-stranded DNA that binds to the core sequence of the desired fragments. The probe includes a radioactive or chemical “tag” that “lights up” to mark the fragments to which it is bound. For example, if a nylon membrane with radioactively tagged fragments is placed next to sheets of suitable photographic film, the radiation will expose the film in two separate “bands” adjacent to the two regions where fragments of the two lengths have migrated. See, e.g., Kaye & Sensabaugh, supra note 9, at 501–02. But laboratory artifacts or contamination can produce extra bands or missing bands. See, e.g., United States v. Yee, 134 F.R.D. 161 (N.D. Ohio 1991), aff’d sub nom. United States v. Bonds, 12 F.3d 540 (6th Cir. 1993); Christopher Anderson, DNA Fingerprinting on Trial, 342 NATURE 844 (1989); William Thompson & Simon Ford, Is DNA Fingerprinting Ready for the Courts?, NEW SCIENTIST, Mar. 31, 1990, at 38; Kolata, supra note 2.

224. See, e.g., Yee, 134 F.R.D. at 166–67; State v. Pennington, 393 S.E.2d 847, 852–54 (N.C. 1990) (accepting uncontradicted expert testimony that false positives are impossible); Glover v. State, 787 S.W.2d 544, 548 (Tex. Ct. App. 1990) (finding analysis admissible in light of other decisions where “[a]ppellant did not produce any expert testimony”)

225. MD. CODE ANN., CTS. & JUD. PROC. § 10–915 (1998) (“In any criminal proceeding, the evidence of a DNA profile is admissible to prove or disprove the identity of any person.”); MINN. STAT. ANN. § 634.25 (West Supp. 2001) The statute states:

In a . . . criminal trial or hearing, the results of DNA analysis . . . are admissible in evidence without antecedent expert testimony that DNA analysis provides a trustworthy and reliable method of identifying characteristics in an individual’s genetic material upon a showing that the offered testimony meets the standards for admissibility set forth in the Rules of Evidence.
defense experts of impeccable credentials, however, produced a few cases rejecting specific proffers on the ground that the testing procedure was not sufficiently rigorous. Moreover, a minority of courts, perhaps concerned that DNA evidence might well be conclusive in the minds of jurors, added a “third prong” to the general acceptance standard. This augmented Frye test requires not only proof of the general acceptance of the ability of science to produce the type of results offered in court, but also a showing of the proper application of an approved method on the particular occasion. Whether this inquiry is properly part of the special screening of scientific methodology, however, is debatable.

Id.; see also Kenneth E. Melson, Legal and Ethical Considerations, in DNA FINGERPRINTING: AN INTRODUCTION 189, 199–200 (Lorne T. Kirby ed., 1990).

226. See State v. Schwartz, 447 N.W.2d 422, 428 (Minn. 1989) (recognizing that “DNA typing has gained general acceptance in the scientific community,” but “the laboratory in this case did not comport” with “appropriate standards”); People v. Castro, 545 N.Y.S.2d 985, 996 (Sup. Ct. 1989) (stating that principles of DNA testing have been generally accepted, but “[i]n a piercing attack upon each molecule of evidence presented, the defense was successful in demonstrating to this court that the testing laboratory failed in its responsibility to perform the accepted scientific techniques and experiments”); Colin Norman, Maine Case Deals Blow to DNA Fingerprinting, 246 SCIENCE 1556 (1989); Rorie Sherman, DNA Tests Unravel?, NAT’L L.J., Dec. 18, 1989, at 1, 24–25.

Effective December 1, 2000, Federal Rule of Evidence 702 was amended to require that the proponent show both the expert used a reliable methodology and the expert “has applied the principles and methods reliably to the facts of the case.” The Advisory Committee Note states:

The amendment specifically provides that the trial court must scrutinize not only the principles and methods used by the expert, but also whether those principles and methods have been properly applied to the facts of the case. As the court noted in In re Paoli R.R. Yard PCB Litig., 35 F.3d 717, 745 (3d Cir. 1994), “any step that renders the analysis unreliable . . . renders the expert’s testimony inadmissible. This is true whether the step completely changes a reliable methodology or merely misapplies that methodology.”

Id. (emphasis in original)

227. This innovation was introduced in Castro, 545 N.Y.S.2d at 995–99. It soon spread. See United States v. Two Bulls, 918 F.2d 56, 61 (8th Cir. 1990) (“[I]t was error for the trial court to determine the admissibility of the DNA evidence without determining whether the testing procedures . . . were conducted properly.”), vacated for reh’g en banc, app. dismissed due to death of defendant, 925 F.2d 1127 (8th Cir. 1991); Ex parte Perry, 586 So. 2d 242, 248 (Ala. 1991). For cases declining to graft a “third prong” onto Frye, see, for example, State v. Bible, 858 P.2d 1152, 1179–84 (Ariz. 1993); Hopkins v. State, 579 N.E. 2d 1297, 1302–04 (Ind. 1991); State v. Vandeboegart, 616 A.2d 483, 490 (N.H. 1992); State v. Cauthron, 120 Wash. 2d 879, 888–90, 846 P.2d 502, 505–07 (1993).

228. Later, some courts insisted on such a showing as part of the demonstration of scientific soundness required under Daubert. See, e.g., United States v. Martinez, 3 F.3d 1191, 1197–99 (8th Cir. 1993).

229. For an analysis concluding that such matters are better handled, not as part of the special test for scientific evidence, but as aspects of the balancing of probative value and prejudice, see
A different attack on DNA profiling that began in cases during this period proved far more successful and led to a third wave of cases. Even if the laboratory has found the true VNTR profile in the sample and has correctly determined that it matches the defendant’s, there is some chance that the match is a coincidence because the perpetrator actually was someone else whose VNTR profile happens to be the same as the defendant’s. To dismiss this possibility as remote, prosecutors called on experts to testify that the probability of a coincidentally matching VNTR profile (often called a “random-match probability”) is infinitesimal. However, these estimates relied on a simplified population-genetics model for the frequencies of VNTR profiles that treats each race as a large, randomly mating population. Some prominent scientists claimed that the applicability of the model had not been adequately verified. They suggested that within a broad population group such as Caucasians, subgroups such as Italian-Americans and Swedish-Americans tend to mate among themselves and might have very different frequencies for the VNTR alleles. Such a population structure could cause the simplified estimates to understate (or overstate) the profile frequency for Caucasians derived from data that fail to account for the effects of the subpopulations. A heated debate spilled over from courthouses to scientific journals and convinced the supreme courts of several states that general acceptance was lacking. A 1992 report of the National Academy of Sciences proposed a more “conservative” computational method as a compromise, and this seemed to undermine the claim of scientific acceptance of the less conservative procedure that was in general use.

At this juncture, the debate was poised to enter a fourth phase. In response to the population-genetics criticism and the 1992 NAS report, Margaret A. Berger, Laboratory Error Seen Through the Lens of Science and Policy, 30 U.C. DAVIS L. REV. 1081 (1997).

230. See Kaye, supra note 217, at 128; Thompson, supra note 5, at 22. In the light of the totality of information on the distribution of various genes in populations, the criticism of the simple random-mating model may have been overblown. See Bernard Devlin & Kathryn Roeder, DNA Profiling: Statistics and Population Genetics, in 1 MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY 710 (David Faigman et al. eds., 1997).

231. See NRC II, supra note 1, at 205–11 (tabulating cases); Kaye, supra note 5, at 101.

232. NATIONAL RESEARCH COUNCIL COMMITTEE ON DNA TECHNOLOGY IN FORENSIC SCIENCE, DNA TECHNOLOGY IN FORENSIC SCIENCE (1992) [hereinafter NRC I].


234. NRC I, supra note 232.
came an outpouring of both critiques of the report and new studies of the
distribution of VNTR alleles in many population groups. Relying on the
burgeoning literature, a second National Academy panel concluded in
1996 that the usual method of estimating frequencies of VNTR profiles
in broad racial groups was sound.235 In the fourth phase of judicial
scrutiny of DNA evidence, the courts almost invariably returned to the
earlier view that the probabilities estimated with the random-mating
model (or minor variations of it) are generally accepted and
scientifically valid.236

The fifth phase of the judicial evaluation of DNA evidence is well
underway. Harnessing the Polymerase Chain Reaction (PCR) enables
laboratories to produce millions of identical copies of DNA fragments
even from samples too small for RFLP typing.237 With these in hand,
many DNA polymorphisms can be analyzed quickly and unambiguously.238 Consequently, the RFLP methods “are being rapidly replaced”
with PCR-based methods.239 As results obtained with new methods enter
the courtroom, it becomes necessary to ask whether each such method
rests on a solid scientific foundation or is generally accepted in the
scientific community.240 Sometimes, the answer will be obvious even

---

235. NRC II, supra note 1, at 156–59. The 1996 report provides more refined methods for
estimating allele frequencies in ethnic subpopulations.
236. See, e.g., People v. Soto, 981 P.2d 958, 974 (Cal. 1999) (“Several developments since the
filing of Barney indicate the controversy over population substructuring and use of the unmodified
product rule has dissipated.”); People v. Miller, 670 N.E.2d 721, 731–32 (Ill. 1996) (“[W]hile there
has been some controversy over the use of the product rule in calculating the frequency of a DNA
match, that controversy appears to be dissipating.”); Armstead v. State, 673 A.2d 221, 238 (Md.
1996) (“[T]he debate over the product rule essentially ended in 1993.”); Commonwealth v. Fowler,
685 N.E.2d 746, 748–50 (Mass. 1997) (stating that the product rule with and without ceilings for
VNTRs now meets test of scientific reliability in light of 1996 NRC Report), departing from
Commonwealth v. Lanigan, 596 N.E.2d 311, 314–17 (Mass. 1992) (reasoning that dispute over
population structure evinces lack of general acceptance); 1 MCCORMICK supra 214, § 205, at 761.
237. See, e.g., Kaye & Sensabaugh, supra note 9, at 563–64.
238. For a description of one such system, see infra note 274.
239. NCFDNA, supra note 184, at 16; see also supra note 183 (describing STRs).
240. See, e.g., Harrison v. State, 644 N.E.2d 1243, 1251–52 (Ind. 1995) (reversing due to lack of
a Frye hearing on PCR-based method). But see State v. Scott, No. 01C01-9708-CR-00334, 1999
state statute providing that “the results of DNA analysis . . . are admissible in evidence without
antecedent expert testimony that DNA analysis provides a trustworthy and reliable method of
identifying characteristics in an individual’s genetic material upon a showing that the offered
testimony meets the standards of admissibility set forth in the Tennessee Rules of Evidence
obviated the need for a hearing on the scientific soundness of mitochondrial DNA testing).

457
without an extensive pretrial hearing. 241 The opinions are practically unanimous in holding that the more commonly used PCR-based procedures satisfy these standards. 242

In sum, in little more than a decade, DNA typing has made the transition from a novel set of methods for identification to a relatively mature and well studied forensic technology. However, one should not lump all forms of DNA identification together. New techniques and applications continue to emerge. These range from the use of new genetic systems and new analytical procedures to the typing of DNA from plants and animals. 243 Before admitting such evidence, it will be necessary to inquire into the biological principles and knowledge that would justify inferences from these new technologies or applications. 244 For example, a court’s prior approval of RFLP testing by gel electrophoresis 245 or reverse dot blot testing 246 of PCR-amplified frag-

241. For example, a procedure may be so similar to accepted protocols that acceptance or validity can be inferred from previous cases. See United States v. Johnson, 56 F.3d 947, 952–53 (8th Cir. 1995) (holding, in response to a defense expert’s testimony that a police department’s variation on the FBI protocol for RFLP-VNTR testing had not been validated, that the variation did not preclude admission where an FBI analyst testified that the difference was of no significance); People v. Oliver, 713 N.E.2d 727, 734 (Ill. Ct. App. 1999) (“[T]he minor variations . . . in the . . . second RFLP test did not render it a new scientific technique for the purposes of Frye.”). In general, trial courts have considerable “latitude in deciding how to test an expert’s reliability, and to decide whether or when special briefing or other proceedings are needed to investigate reliability.” Kumho Tire Co., Ltd. v. Carmichael, 526 U.S. 137, 152 (1999) (emphasis in original).


244. For suggestions to assist in this endeavor, see id.; Kaye & Sensabaugh, supra note 9, at 549–59.

245. Gel electrophoresis is described supra notes 222, 223.

246. Reverse dot blot testing “tests for the presence of a specific sequence. The procedure involves the use of probes specific for that particular DNA sequence. The analyst adds known probes . . . . The dot blot test is binary in character; there are only two possible outcomes, and they indicate whether or not the sequence is present. If the [sequence is present], color will develop; if the [sequence is absent], no color will develop at the DNA spot . . . .” 2 PAUL C. GIANNELLI & EDWARD J. IMWINKELRIED, SCIENTIFIC EVIDENCE § 18-3(C), at 18 (3d ed. 1999).
ments containing the HLA DQ-α gene does not dictate the conclusion that the court also must accept testing at STR loci or mitochondrial DNA sequencing. The newer technologies are gaining judicial approval, but a court should not confer approval until it is satisfied that the specific technology satisfies the applicable standard.

2. Proficiency-Test Records

In a study, the researchers empirically verify the ability of the technology to identify features of DNA molecules. In contrast a proficiency study tests how competently the laboratory’s analysts apply a technology that has been validated. The purpose of proficiency testing is to uncover difficulties that a particular technician or a particular laboratory might be encountering in applying established methods.

247. Human Leukocyte Antigen (HLA) is an “[a]ntigen (foreign body that stimulates an immune system response) located on the surface of most cells (excluding red blood cells and sperm cells). HLAs differ among individuals and are associated closely with transplant rejection.” Kaye & Sensabaugh, supra note 9, at 570. DQA is “[t]he gene that codes for a particular class of Human Leukocyte Antigen (HLA). This gene has been sequenced completely and can be used for forensic typing.” Id. at 568. Indeed, “[t]he first use of PCR-based typing for forensic application . . . employed the HLA-DQA1 locus (formerly called the DQ- ).” NCFDNA, supra note 184, at 17.

248. For descriptions of STR loci, see supra note 183; GIANNELLI & IMWINKELRIED, supra note 246, § 18-3(C), at 21.

249. Most DNA-typing techniques analyze DNA found in the chromosomes within cell nuclei. Mitochondrial testing analyzes DNA “found in the energy-producing material surrounding the cell nucleus.” GIANNELLI & IMWINKELRIED, supra note 246, § 18-2, at 3. A mitochondrion is “[a] structure . . . within nucleated . . . cells that is the site of the energy producing reactions within the cell. Mitochondria contain their own DNA (often abbreviated as mtDNA), which is inherited only from mother to child.” Kaye & Sensabaugh, supra note 9, at 571. Because there are hundreds of mitochondria in a cell, but only one nucleus, mtDNA sequencing can be used with samples containing too little nuclear DNA for PCR amplification to work. Id. at 495.

250. See, e.g., State v. Underwood, 518 S.E.2d 231 (N.C. Ct. App. 1999) (approving use of mitochondrial DNA sequencing); State v. Ware, No. 03C01-9705CR00164, 1999 WL 233592 (Tenn. Crim. App. Apr. 20, 1999) (holding that notwithstanding testimony from a defense expert that mtDNA sequencing had not been adequately validated for forensic use, the FBI’s mtDNA testing was properly admitted under the scientific soundness standard); see also Leigh Jones, mtDNA Ruled Reliable in Rape Trial, N.Y.L.J., Sept. 7, 2000; supra note 242.

251. Proficiency testing in forensic genetic testing is designed to ascertain whether an analyst can correctly determine genetic types in a sample the origin of which is unknown to the analyst but is known to a tester. Proficiency is demonstrated by making correct genetic typing determinations in repeated trials, and not by opining on whether the sample originated from a particular individual. Proficiency tests also require laboratories to report random-match probabilities to determine if proper calculations are being made. See Kaye & Sensabaugh, supra note 9, at 511.
Proficiency testing raises a variety of legal issues. It has been suggested that participation in a program of proficiency testing ought to be a prerequisite to the admission of evidence from a forensic laboratory, that proficiency-test results should be admissible to show how likely it is that the laboratory erred in the test at bar, and that random-match probabilities ought to be inadmissible unless they are combined with proficiency-test results to estimate the probability of a false match. If the second suggestion is followed, and the defense is allowed to introduce evidence of proficiency tests to suggest that the laboratory is prone to err, a further question arises: Should the prosecution be permitted to present testimony that the defense has not retested or even requested the opportunity to retest the samples?

a. Proficiency Testing As a Prerequisite to Admission

The first suggestion, that courts condition admissibility on proficiency testing, is a departure from the usual practice. As indicated in the previous section, the scientific-validity and general-acceptance standards relate to the capacity of an analytical procedure to generate accurate results when properly applied, and not to whether the individual or institution using a valid or generally accepted method is skilled and careful or is instead careless and prone to error. Of course, the latter issue can be of paramount importance, but usually it is said to be a matter affecting the weight of the evidence rather than its admissibility.


256. See Edward J. Imwinkelried, The Case Against Evidentiary Admissibility Standards that Attempt to “Freeze” the State of a Scientific Technique, 67 U. COLO. L. REV. 887 (1996). In extreme cases, where the laboratory departs so grossly from accepted practices that the reliability of its findings are in serious doubt, the court may well exclude the evidence on the ground that its probative value is too slight to warrant its admission.
b. **The Admissibility of Errors on Proficiency Tests**

The second suggestion, that testimony about proficiency-test results be used to reveal the chance of error in the case at bar, presupposes that such evidence is admissible at trial.\(^{257}\) In its 1992 report, a committee of the National Academy of Sciences took the position that “laboratory error rates must be continually estimated in blind proficiency testing and must be disclosed to juries.”\(^{258}\) There is authority that when the prosecution introduces testimony about the probability of a coincidentally matching profile, the defendant is entitled to introduce testimony about the laboratory’s proficiency tests.\(^{259}\) Indeed, it has been held that the opponent must be allowed to cross-examine one laboratory representative about errors committed by other analysts at the laboratory.\(^{260}\)

In contrast, in a report published in 1996, a second committee of the National Academy of Sciences declined to take a position on whether evidence of laboratory error rates, as estimated from proficiency studies, should be admissible at trial.\(^{261}\) However, the report’s discussion of proficiency testing raises questions about the probative value of such evidence. For example, the report notes that “[t]he pooling of proficiency-test results across laboratories” could mislead a jury and “penalize the better laboratories.”\(^{262}\) It adds that even a test of the same laboratory might be outdated, because the laboratory may have taken

---

257. The discussion of admissibility in this section is limited to the guilt phase of a case. In many jurisdictions, the formal rules of evidence are not binding during the sentencing phase. *See*, *e.g.*, Fed. R. Evid. 1101(d)(3). Moreover, the U.S. Supreme Court has indicated that during the sentencing stage of a capital case, the Constitution requires the sentencer to “consider[], as a mitigating factor, any [logically relevant] aspect for a defendant’s character or record and any of the circumstances of the offense that the defense proffers as a basis of a sentence less than death.” Lockett v. Ohio, 438 U.S. 586, 604 (1978) (emphasis in original). Although the past performance of the laboratory is neither an aspect of a defendant’s character or record nor a circumstance of the offense, a capital defendant might argue that the sentencer also must be able to consider all conceivable weaknesses in the state’s proof of the offense.

258. NRC I, *supra* note 232, at 89.


261. The report stated that the committee had chosen to limit its remarks to the question of “what aspects of the procedures used in connection with forensic DNA testing are scientifically valid.” NRC II, *supra* note 1, at 185.

262. *Id.* at 86.
The testimony could be vulnerable to an objection under Federal Rule of Evidence 403, which requires the exclusion of evidence whose probative value is substantially outweighed by the danger of prejudice, confusion of the issues, or undue consumption of time.264

A further objection is that the testimony represents inadmissible character evidence.265 If the theory of logical relevance is merely that the laboratory’s past commission of errors increases the probability that the laboratory erred on the occasion in question, then the theory amounts to forbidden character reasoning.266 This is precisely the theory of logical relevance generally banned by Federal Rule of Evidence 404.267 Moreover, to the extent that proficiency-test results constitute evidence of specific acts introduced to show a general tendency to make mistakes, they seem to run afoul of Rule 405, which forbids this form of character evidence.

263. Id. ("A laboratory is not likely to make the same error again."); see also NRC I, supra note 232, at 89 (recognizing that "errors on proficiency tests do not necessarily reflect permanent probabilities of false-positive or false-negative results").


265. See Edward J. Imwinkelried, Coming to Grips with Scientific Research in Daubert’s “Brave New World”: The Courts’ Need to Appreciate the Evidentiary Differences between Validity and Proficiency Studies, 61 Brook. L. Rev. 1247, 1273–78 (1995). A number of jurisdictions have abolished the character-evidence prohibition as it applies to a defendant’s character in certain types of cases such as rape or child abuse. See Fed. R. Evid. 413–15; Cal. Evid. Code §§ 1108–1109.

In these jurisdictions that allow the prosecution to rely on an accused’s past misconduct as circumstantial proof of the charged offense, the defense conceivably could argue that the differential treatment of the accused’s inculpatory misconduct and the exculpatory proficiency test results violates the equal protection guarantee. See, e.g., Nettles v. State, 683 So. 2d 9, 12 (Ala. Crim. App. 1996). However, the constitutional attacks on character-evidence restrictions on defense evidence have failed. EDWARD J. IMWINKELRIED & NORMAN M. GARLAND, EXCULPATORY EVIDENCE: THE ACCUSED’S CONSTITUTIONAL RIGHT TO INTRODUCE FAVORABLE EVIDENCE, 435–72 (2d ed. 1996).

266. See, e.g., Moorhead v. Mitsubishi Aircraft Int’l, 828 F.2d 278, 287 (5th Cir. 1987) (holding that lower court erred by admitting pilot’s low marks at flight school refresher course); see generally 1 MCCORMICK, supra note 214, § 186.

267. Federal Rule of Evidence 404(a) provides that “[e]vidence of a person’s character or a trait of his character is not admissible for the purpose of proving action in conformity therewith on a particular occasion.” Section (b) of the rule recognizes certain exceptions to this blanket rule of exclusion, but none are apposite here. There also is an exception permitting a witness’s opponent to impeach the witness by questioning the witness about prior untruthful acts. Fed. R. Evid. 608(b).

However, those acts relate to the witness’s character trait for untruthfulness, rather than the trait of competence or proficiency. The Federal Rules expressly carve out the exception for untruthfulness, but there is no comparable exception for the character trait of competence or proficiency.
One might argue that the character rules do not apply to entities such as a laboratory. However, the language of the rules is broad enough to extend to businesses and other entities; and the cases have generally construed the statutes as reaching entities as well as natural persons.

This issue is rarely recognized as a character-evidence problem in the trial court, but a trial judge might find it difficult to justify overruling a properly phrased character-evidence objection when the theory of relevance is nothing more than a general tendency of the laboratory to make mistakes. If there is a consensus that the jury sometimes needs the proficiency-test results as an antidote to overwhelmingly small random-match probabilities, then the federal and state rules governing character evidence should be altered to give the trial court the discretion to admit the evidence.

268. Federal Rule of Evidence 405(a) provides that “[i]n all cases in which evidence of character or a trait of character of a person is admissible, proof may be made by testimony as to reputation or by testimony in the form of an opinion. On cross-examination, inquiry is allowable into relevant specific instances of conduct.” FED. R. EVID. 405(a). Section (b) permits specific act evidence only when character is “in issue”—a term of art that has no application to the tendency of laboratory personnel to make mistakes in performing DNA tests. See 1 MCCORMICK, supra note 214, § 187.

269. Federal Rule of Evidence 404 refers to a "person," and the accompanying Advisory Committee Note states that the drafters drew on the California character statutes. FED. R. EVID. 404. California Evidence Code § 175 defines "person" as including a "firm, association, organization, partnership, business trust, corporation, limited liability company, or public entity." CAL. EVID. CODE § 175. Moreover, the complex of character rules includes the habit doctrine, and Federal Rule of Evidence 406, codifying that doctrine, expressly applies to "an organization." FED. R. EVID. 406.


271. But see United States v. Shea, 957 F. Supp. 331, 344 n.42 (D.N.H. 1997) (“The parties assume that error rate information is admissible at trial. This assumption may well be incorrect. Even though a laboratory or industry error rate may be logically relevant, a strong argument can be made that such evidence is barred by Fed. R. Evid. 404 because it is inadmissible propensity evidence.”); Unmack v. Deaconess Med. Ctr., 967 P.2d 783, 784–86 (Mont. 1998) (going to the brink of explicitly holding that the character evidence prohibition bars testimony about prior incidents of unskilful conduct).

272. Courts also generally have not addressed the impact of the character-evidence ban on expert testimony about the conditions under which eyewitness identifications are likely to be in error. Such testimony is unusual, and exclusion almost invariably is upheld on appeal. In the rare cases where appellate courts have held that the failure to admit the evidence was an abuse of discretion, they have not mentioned the rule against character evidence. See State v. Chapple, 660 P.2d 1208, 1217–24 (Ariz. 1983); People v. McDonald, 690 P.2d 709, 715–27 (Cal. 1984). Of course, much of this type of testimony falls outside the character-evidence rule. Thus, the rule does not ban testimony that “weapons focus” interferes with the accuracy of eyewitness identifications any more than it bans testimony that handling DNA samples from the suspect and the crime scene
Moreover, both the bench and bar should appreciate that in some circumstances proficiency tests of the laboratory involved in the case should be held admissible without relaxing the ban on character evidence. The ban applies only when the sole theory of logical relevance is that the existence of errors in the past suggests a tendency to err that might affect the result in the case at bar. There might be situations in which the defense can use the test data at trial on an entirely different theory of logical relevance. Assume, for instance, that the experts in a case disagree over whether a peak or a band observed in a DNA test is due to an allele or is an artifact. Evidence that spurious peaks or bands have occurred under similar circumstances in proficiency tests of the laboratory on known samples would lend support to the defense theory that the peak or band in the pending case is an artifact. In this situation, proficiency-test data are relevant because they provide information about the operating characteristics of the DNA test at that particular laboratory.

273 Artifacts in RFLP testing with gel electrophoresis and visualization by autoradiography are mentioned supra note 222.

274 STR typing uses capillary electrophoresis, in which much smaller amounts of DNA are drawn through sieving material in an exquisitely thin tube. A smaller STR fragment travels through this capillary faster than a longer one. The fragments are labeled with a fluorescent dye and a laser continuously illuminates a detection window near the end of the capillary. When the laser light strikes an allele passing by the window, the dye fluoresces, and a charged coupled device (CCD) in a camera responds to the intensity of the fluorescent light. The signal is flat before the allele passes by, peaks when the allele is in the window, and then drops back to the zero level. The allele size is determined by software that compares the time at which the peak appears to the times measured for a “ladder” of fragments whose lengths are known. See NCFDNA, supra note 184, at 52–53.

275 Of course, even when the proficiency test data would be admissible and the defense has a legitimate need to discover this type of information, there might be means of satisfying the need other than by furnishing proficiency-test results. By way of example, a sampling of the laboratory’s case work could meet the need. However, in most cases permitting discovery of proficiency-test data may be preferable. It will likely be more convenient for the laboratory to reveal the proficiency test data, because that data have already been compiled and giving the defense access to actual case work could compromise the privacy of the persons involved in those cases. When the defense needs to discover information about the operating characteristics of a laboratory’s test for a purpose other than merely establishing the laboratory’s general error rate, the data could prove to be admissible at trial; hence, the courts would not be justified in denying discovery of proficiency-test results on the ground that such discovery cannot lead to the production of admissible evidence at trial.
c. The Use of Proficiency Tests To Modify Random-Match Probabilities

The third argument relating to proficiency testing is that estimates of the probability that a randomly selected person would have the DNA type found in a crime-scene sample should be inadmissible unless accompanied by or blended with the laboratory’s error rate.\textsuperscript{276} The 1996 committee observed that combining the probability of a random match with the probability of error “would deprive the trier of fact of the opportunity to evaluate separately the possibility that the profiles match by coincidence as opposed to the possibility that they are reported to match by reason of laboratory or handling error.”\textsuperscript{277} The committee took the position that “a calculation that combines error rates with match

\textsuperscript{276} Combining the random-match probability with the probability of a false-positive laboratory error according to the rules governing conditional probabilities would give the jury an estimate of the probability that the laboratory would report a match if the source of the crime-scene DNA was not the defendant. To see how this might be done requires a few symbols. Let $T$ stand for the DNA type of the crime-scene sample, and let $R_T$ be the event that the laboratory reports this type when it tests a sample. Let $S$ be the hypothesis that a defendant is the source of the crime-scene sample, and let $-S$ be the hypothesis that someone else is. Finally, let $D_T$ be the event that defendant’s DNA actually is type $T$, and $D_T$ be the event the defendant’s DNA is of some other type. Then the chance that a person who is not the source of this sample would have DNA of type $T$ is $P(D_T|-S)$, where $P(D_T|-S)$ stands for the conditional probability of $D_T$ given $-S$. Suppose further that the chance that the laboratory would report that this individual is type $T$ when he is not is $P(R_T|D_T)$.

\textsuperscript{277} NRC II, supra note 1, at 85.
probabilities is inappropriate." 278 The reasoning supporting the committee’s position essentially sounds under Federal Rule of Evidence 403. 279 If anything, the Rule 403 objection is more substantial here than when it is urged as a basis for excluding testimony offered to impeach the laboratory’s competence. In this situation, the questions about the validity of industry-wide error rates and the staleness of even the laboratory’s own tests are equally applicable and call into question the probative worth of the testimony. Moreover, there is a heightened risk that the jury will be confused. Error rates and random-match probabilities relate to distinct hypotheses, and a lay juror may find it difficult to understand the significance of a computation which merges the rates and the probability. That mode of computation could place even greater strain on the jurors’ ability to comprehend the body of evidence submitted to them. 280 The few courts that have addressed the argument that error rates should be used to the exclusion of random-match probabilities have not been persuaded. 281

d. The Opportunity To Retest As a Response to Defense Arguments About Proficiency Testing

While defense counsel originated the first three suggestions, the fourth suggestion related to proficiency testing has been made by prosecutors. The thrust of this suggestion is that when the defense is allowed to introduce evidence of proficiency tests of the laboratory employing the prosecution’s expert to suggest that the laboratory is prone to err, the prosecution should be permitted to present testimony that the defense has not retested or even requested the opportunity to retest the samples analyzed by the prosecution’s expert.

The testimony would be logically relevant under several theories. First, if a defense expert testifies that the laboratory result is untrust-

278. Id. at 87.
279. See generally Margaret A. Berget, Laboratory Error Seen Through the Lens of Science and Policy, 30 U.C. DAVIS L. REV. 1081 (1997).
280. Jason Schklar & Shari Seidman Diamond, Juror Reactions to DNA Evidence: Errors and Expectancies, 23 LAW & HUM. BEHAV. 159, 179 (1999) (concluding that separate figures are desirable in that “[j]urors . . . may need to know the disaggregated elements that influence the aggregated estimate as well as how they were combined in order to evaluate the DNA test results in the context of their background beliefs and the other evidence introduced at trial”).
281. See, e.g., Armstead v. State, 673 A.2d 221, 245–46 (Md. 1996) (rejecting the argument that the introduction of a random-match probability deprives the defendant of due process because the error rate on proficiency tests is many orders of magnitude greater than the match probability).
worthy, the failure to retest would be relevant to impeach the defense expert’s credibility on the ground that a scientist who truly doubted the accuracy of the analysis normally would have retested the samples to resolve the matter. Inasmuch as replication is a crucial and common feature of scientific inquiry, it could be argued that neglecting to retest is prior inconsistent conduct. On this theory, the defense would be entitled to a limiting instruction to the effect that the expert’s failure to retest is not offered to show that the test result is correct, but only to demonstrate that the defense expert is not sincere in asserting that it is flawed.

The probative value of a failure to retest in showing an expert’s insincerity, however, is open to question. It is not uncommon for scientists to question in print or otherwise the adequacy of another researcher’s experiment before undertaking to replicate it. And even if such opinions were unheard of in the course of ordinary science, the expert may have been retained for the limited purpose of giving an opinion on the adequacy of the testing that was done rather than redoing that testing. Nevertheless, the inference of insincerity need not be particularly strong for the “inconsistent” conduct to be a proper, logically relevant subject for cross-examination.

Second, if the defense expert offers an opinion that the laboratory’s results may be in error, the expert’s failure to request or conduct an independent test would be relevant to suggest that the jury should give less weight to that opinion. The prosecution could argue to the jury that an expert who fails to use a more definitive and readily available procedure for ascertaining whether the initial test results are correct has not been thorough in evaluating those results, and that such experts

282. This theory does not apply if the defense introduces the proficiency-test data by cross-examining the prosecution’s experts rather than producing its own expert.


284. FED. R. EVID. 105.


286. Thus, in People v. Oliver, 713 N.E.2d 727 (Ill. App. Ct. 1999), the State was allowed to show that a defense expert who questioned the results of DNA tests had done no testing of his own. See id. at 736 (“[I]t was proper for the prosecution to bring out on cross-examination that the defense criticisms of the prosecution’s expert witnesses were not based on any independent testing that it had done.”).
deserve little credence because the basis for the opinion is not as complete as it could be. Again, the inference may be debatable, but the standard of relevance, particularly on cross-examination, is lenient.287

Third, whether or not a defense expert discusses proficiency tests, the prosecution could argue that the defense’s failure to retest (or to request a retest) amounts to an admission of the accuracy of the initial test by conduct by the defendant.288 The courts have applied the admission-by-conduct theory to a litigant’s failure to present evidence when “it would be natural” for the litigant to introduce such testimony.289 The prosecution might urge that it would be natural for a defendant affected by a false match to seek retesting and that it would be natural for a DNA expert who entertained serious doubts about the accuracy of a prior test to retest the samples.290

In short, there are reasonable arguments for permitting the prosecution to raise the issue of retesting when a defendant questions the laboratory’s ability to type DNA samples correctly. But even if the inquiry is probative of the insincerity or lack of thoroughness of the expert, or an admission by the defendant, there are potential objections to this counterthrust by the prosecution. One objection is that the inquiry is inconsistent with the prosecution’s burden of proof.291 To reinforce the allocation of the burden to the government, some courts generally forbid prosecution comment on the defense failure to produce evidence.292 The argument runs that the defense is entitled to rely on the burden and has no obligation to present any evidence at trial. According to this line of argument, it is improper to convert the defense’s failure to present testimony into prosecution evidence.293 Under this line of authority, the

287. FED. R. EVID. 401; United States v. Casares-Cardenas, 14 F.3d 1283, 1287 (8th Cir.) (“Relevance is established by any showing, however slight, that makes it more or less likely . . . .”), cert. denied, 513 U.S. 849 (1996); United States v. Nason, 9 F.3d 155, 162 (1st Cir. 1993) (“The threshold for relevance is very low under Federal Rule of Evidence 401.”).

288. 2 MCCORMICK, supra note 214, § 264.

289. Id. § 264, at 174.

290. On this theory, the defense is not entitled to a limiting instruction; an admission by conduct qualifies as substantive evidence. Id.; FED. R. EVID. 105.


293. A related argument looks to the privilege against self-incrimination. Griffin v. California, 380 U.S. 609, 614 (1965), teaches that the prosecution may not comment on the accused’s invocation of the privilege. However, a prosecutor’s statement that the defense has not introduced
Defense could bar prosecution comment about the defense’s failure to retest the DNA sample. However, even in such a jurisdiction, if the defense overreached, prosecution comment might be permitted as an invited response. In addition, some jurisdictions reject that line of authority and allow comment on the defense’s failure to present exculpatory evidence, so long as the trial judge clearly instructs the jury that the prosecution has the ultimate burden of proof.

A further objection is that the admission of the testimony is inconsistent with the defendant’s attorney-client privilege. A number of jurisdictions apply the attorney-client privilege when, as part of trial preparation, defense counsel hires an expert to evaluate private information from the defendant, such as the defendant’s mental or physical condition. The Advisory Committee Note to draft Federal Rule of Evidence 503 endorsed the application of the attorney-client privilege to experts and some courts have gone to the length of invoking the theory even when the expert did not evaluate information realistically originating from the defendant. Based on these authorities, the defense might contend that the attorney-client privilege applies to a defense expert’s retest of a DNA sample. The gist of the objection would be that if the result of a retest would be privileged, it is wrong-minded to penalize the defense for failing to retest.

Rebuttal expert testimony would not amount to impermissible comment. Courts have held that similar statements from the prosecution were improper only when the defendant was the only potential witness who could contradict the prosecution. See Bergmann v. McCaughtry, 65 F.3d 1372, 1377 (7th Cir. 1992). In a case involving DNA, the prosecutor’s comments would relate to potential rebuttal testimony by an expert witness rather than any testimony from the accused.


295. See Van Woudenberg ex rel. Foor v. Gibson, 211 F.3d 560, 570 (10th Cir. 2000) (“The prosecutor may . . . comment on the defendant’s failure to present evidence or call witnesses.”); People v. Guzman, 96 Cal. Rptr. 2d 87, 91 (Ct. App. 2000).


298. Riddle, 964 P.2d at 1063–64 (regarding accident-reconstruction expert).
As with the other suggestions related to proficiency testing, the case law offers little guidance. In principle, it would seem that once the defense has sharpened the issue of the prosecution expert’s use of proper test procedures, the prosecution should be allowed to elicit testimony about the defense’s failure to retest at least to probe the basis for the expert’s opinion and as circumstantial evidence of defendant’s belief that retesting would not yield a different result. The fact that the prosecution has the burden of persuasion does not make such inferences impermissible. The constitutional requirement for proof beyond a reasonable doubt regulates the quantum of proof the prosecution must present, but no court has invoked the requirement to preclude the prosecution from introducing an otherwise admissible item of evidence. In appropriate circumstances, the majority of courts permit prosecutors to comment on a defendant’s failure to produce evidence such as an available witness who would presumably corroborate the defendant’s testimony.

Neither should the attorney-client privilege pose an insurmountable barrier. Certainly, the prosecution cannot comment on a defendant’s decision to exercise a constitutional privilege, and comment on a defendant’s failure to produce a witness is often forbidden when the defendant stands in a privileged relationship with the witness. Consequently, it might be justifiable to apply the attorney-client privilege to a defense expert’s actual analysis of material that has become available because of the defendant’s exercise of the right to prepare a defense with the assistance of counsel. Perhaps material that both emanates from the defendant and is still confidential would fall into this category. However, these conditions do not seem to be satisfied in

299. Thus, in *Fluellen v. Campbell*, 683 F. Supp. 186 (M.D. Tenn. 1987), defense counsel argued that the state’s case was weakened by the fact that it failed to have blood tests performed, and the prosecutor remarked in rebuttal, “if he . . . thinks that is such good evidence, why didn’t he request that it be done?” Id. at 189. The federal district court found that “this comment in no way imposed upon the jury a presumption which conflicted ‘with the overriding presumption of innocence with which the law endows the accused and which extends to every element of the crime.’” Id. (quoting *Morissette v. United States*, 342 U.S. 246, 275 (1952)).

300. See Alan Stephens, Annotation, *Adverse Presumption or Inference Based on Party’s Failure to Produce or Examine Family Members Other than Spouse—Modern Cases*, 80 A.L.R.4th 337, 344 (1990); Alan Stephens, Annotation, *Adverse Presumption or Inference Based on Party’s Failure to Produce or Examine Friend—Modern Cases*, 79 A.L.R.4th 779, 785–86 (1990).

301. Most, but not all, jurisdictions also forbid comment on the invocation of a statutory or common law privilege. See 1 *MCCORMICK*, supra note 214, § 74.1, at 307–08.

DNA Typing Issues

this setting. The DNA sample that the defendant suggests has been improperly analyzed might be crime-scene material that was not obtained from the defendant, or it could be a sample that the prosecution lawfully acquired from the defendant. In these situations, the attorney-client privilege should not preclude adverse comment on the defense’s failure to retest.

B. Extending the Period of Statutes of Limitations

The power of DNA evidence has prompted proposals to create an exception to the statute of limitations for sexual assault when DNA profiling links the suspect to the assault.\(^{303}\) Moreover, some prosecutors have attempted to avoid the tolling of the statute by filing “John Doe” arrest warrants based solely on a description of the unnamed assailant’s DNA.\(^ {304}\) However, devising a workable “DNA exception” that would respect the interests of defendants and society in defining a point after

---

303. In May 1998, for example, the Illinois Senate approved House Bill No. 2257, which provides:

[I]f the identity of the accused is unknown and at the time of the offense physical evidence is collected that is capable of being tested for its DNA characteristics which would identify the accused, a prosecution for predatory criminal sexual assault of a child, criminal sexual assault or aggravated criminal sexual assault may be commenced at any time.

1997 H.B. 2257, 90th Gen. Assembly (Ill. 1998); see also, e.g., 2000 S.B. 2347, 209th Leg., 2d Sess. (N.J. 2001) (eliminating the statute of limitations for prosecuting criminal sexual contact and endangering the welfare of child “if the identity of the defendant may be determined by physical evidence capable of forensic deoxyribonucleic (DNA) testing”); Staff and Wire Reports, 77th Legislature, Ft. Worth Star-Telegram, Apr. 24, 2001, at 6, available at 2001 WL 5148152 (reporting the signing of legislation “that ends the statute of limitations for sexual assault if a DNA sample from the attacker is available but no one has been apprehended”); cf. 2001 S.B. 152, 141st Gen. Assem. (Del. 2001) (providing that “[i]n any indictment for a crime in which the identity of the accused is unknown it is sufficient to describe the accused as a person whose name is unknown but has a particular DNA profile.”).

304. See, e.g., Richard Willing, Mystery Suspects Charged Through DNA, USA TODAY, Apr. 3, 2000, at 3A, available at 2000 WL 5773965 (reporting that “California, New York, Oklahoma, Utah and Wisconsin have filed such DNA-based charges so far” and that “[s]imilar charges are imminent in two more states, and prosecutors say dozens more could be filed by year’s end”). The first case to be commenced on the basis of such charges is pending in Sacramento. See Erin Hallissy & Charlie Goodyear, Databank Match Brings Arrest on DNA Warrant, S.F. CHRON., Oct. 25, 2000, at A3, available at 2000 WL 6494988; Richard Willing, Police Expand DNA Use, Charge Man with Rape Using Only Genetic Profile, USA TODAY, Oct. 25, 2000, at 1A, available at 2000 WL 5793574.
which litigation no longer can be commenced is a formidable challenge. 305

Statutes of limitations serve a variety of purposes. Most obviously, they protect individuals against the risk that they will unable to assemble adequate evidence for a defense because too much time has passed since the alleged crime was committed. With time, memories fade, evidence is misplaced, witnesses become harder to locate, and the accused’s ability to defend himself is reduced. 306 Thus, the Supreme Court has described statutes of limitations as “the primary guarantee against bringing overly stale criminal charges.” 307 In addition, they give innocent (as well as guilty individuals) a certain peace of mind and encourage the police to move on to more recent cases that are more likely to be solved and for which punishment would be more effective. 308

A DNA exception attends only to the first justification for statutes of limitations. If a comparison of the defendant’s DNA with the trace evidence DNA collected many years ago were to establish conclusively that the defendant is guilty, 309 then it could be argued that any degradation in the defendant’s ability to mount a defense would be harmless because it could not affect the outcome of the trial. For example, even when a defendant’s alibi witness had died after the

---

305. See Jonathan W. Diehl, Note, Drafting a Fair DNA Exception to the Statute of Limitations in Sexual Assault Cases, 39 JURIMETRICS J. 431 (1999). Portions of this Article’s discussion of statutes of limitations are adapted, without further attribution, from this note.


307. United States v. Ewell, 383 U.S. 116, 122 (1966); see also Marion, 404 U.S. at 322; MODEL PENAL CODE § 1.06 cmt. 1, at 86 (1985); Tyler T. Ochoa & Andrew J. Wisstrich, The Puzzling Purposes of Statutes of Limitation, 28 PAC. L.J. 453, 458 (1997). The Due Process Clause of the Fifth Amendment also provides protection, but only when the defendant establishes that the delay not only substantially prejudiced the defense, but also that it was a tactical ploy by the prosecutor. See United States v. Lovasco, 431 U.S. 783, 790 (1977); Marion, 404 U.S. at 324. The Speedy Trial Clause of the Sixth Amendment offers the defendant no protection against preindictment delay. See Marion, 404 U.S. at 313.

308. The need for specific deterrence, incapacitation, and retribution usually fades with time. See MODEL PENAL CODE § 1.06 cmt. 1, at 86.

309. DNA evidence does not grow stale with the passage of time. Even ancient DNA sometimes can be analyzed successfully. See, e.g., ANCESTOR MATERIAL FROM PALEONTOLOGICAL, ARCHAEOLOGICAL, MUSEUM, MEDICAL, AND FORENSIC SPECIMENS (B. Herrmann & S. Hummel eds., 1994); Richard Willing, Fear Keeps Up as DNA Science Speeds Forward, USA TODAY, Mar. 29, 2000, at 26A, available at 2000 WL 577365 (reporting a case in Britain in which a thirty-seven-year-old saliva stain on a postage stamp helped exonerate a man hung in 1962 for murder). However, the ability of fragments of the DNA molecule to remain intact virtually indefinitely under certain conditions is not dispositive. Indeed, if the prosecution relies on DNA profiling of a stain shortly after a fairly fresh crime-scene sample first was collected, it is not even pertinent.
statutory period, if no reasonable jury could have believed the alibi in the face of the DNA proof, the availability of the witness could not have resulted in an acquittal.

The premise that DNA evidence is dispositive, however, is not always true. First, there are cases in which a defendant might succeed in raising a reasonable doubt about the reported results of the DNA tests. This situation would arise, for instance, when there is reason to think that samples were switched or cross-contaminated in the laboratory or in the collecting and handling of the trace evidence before it reached the laboratory. Many years later, the police officers and laboratory personnel involved could be impossible to locate, and the written records remaining might be inadequate to resolve these claims.

Second, even if one were to conclude that such cases are too rare to be an obstacle to creating an exception to the statute, DNA evidence can be conclusive only as to one factual issue—whether the DNA in the trace evidence somehow originated from the defendant. Without more, proof of that factual issue ordinarily falls far short of demonstrating guilt for every type of sexual assault. Thus, a defendant’s semen might be present on an alleged victim’s clothing or a bedsheets even if there had been no penetration, and it would be expected to be found in a vaginal swab if the sex had been consensual.310

The legislature might try to respond to these concerns by confining the DNA exception to cases in which identity is the only issue that needs to be resolved.311 But which cases are these? Can a defendant avoid the extension of the period of limitations by conceding his identity as the source of the trace evidence but alleging that he reasonably believed that the woman invited his actions, that he was acting under duress, or the like? Should the Court be required to find that these defenses have no basis in fact for the prosecution to proceed after the statute has run? It might be possible to draft a suitably sensitive DNA exception to the

310. For reports of successful consent defenses in sexual assault cases involving DNA evidence, see Smith v. State, 734 N.E.2d 706, 708 (Ind. Ct. App. 2000) (observing that the defendant who had advanced a consent defense at a previous trial for rape had been acquitted despite the DNA evidence against him in that case); Dean Wise, Jury Acquits Man of Rape, YORK DAILY REC., Oct. 25, 2000, at C04, available at 2000 WL 27979596 (reporting that although the state of Maryland established that “the chances of the semen having come from a [randomly selected] man other than Vinson within the African-American population was one in 2.3 quadrillion,” and although the defendant denied the allegation of intercourse, he had had a previous intimate relationship with the alleged victim for eighteen months).

311. The requirement in some bills that “the identity of the accused [be] unknown” points in this direction. See supra note 303.
statute of limitations, but the task is not so simple as, initially, it might appear to be.

In contrast, the situation is much simpler when a defendant brings forward DNA evidence in a case involving a single rapist that shows that the DNA in the trace evidence is not his. If that evidence is believed, then he is not the guilty party. But while DNA evidence can be conclusive of innocence, DNA evidence is not logically sufficient to prove guilt. Because identity is not the only element of the offense and because there are affirmative defenses that can be pled, even when the state brings forward incontestable DNA evidence of identity, the defendant might not be guilty of sexual assault. Consequently, it would be consistent to advocate an extension of the period in which post-conviction relief can be sought312 while opposing an extension of the statute of limitations.

CONCLUSION

This Article has canvassed a wide variety of issues. Some are constitutional in nature, others are statutory, and still others arise at common law. The character of the issues ranges from substantive criminal law to procedure to evidence. While the issues are diverse, they share two related common denominators. One is their relative novelty, and the other is the consequent paucity of case law analyzing the issues. In the past decade, a dizzying array of DNA technologies—from gel electrophoreses of single-locus RFLPs, to PCR-based studies of STRs, to mitochondrial DNA sequencing—has materialized in American courtrooms. The predictable result has been a proliferation of evidentiary issues. Yet, several important questions have received little in the way of careful consideration. This Article has identified and ventured answers to these questions. As stated at the outset, despite their novelty, some of these issues admit of relatively clear answers. In other cases, though, their resolution will demand a sophisticated balancing of competing public policy considerations. If the criminal justice system is to realize the full potential of DNA technology while maintaining its essential fairness, the system must come to grips with these issues in short order.

312. For recommended procedures of handling requests for DNA testing after conviction, see NATIONAL COMMISSION ON THE FUTURE OF DNA EVIDENCE, REPORT AND RECOMMENDATIONS POSTCONVICTION RELIEF (1999).
ONE STEP FORWARD, TWO STEPS BACK: VASQUEZ v. HAWTHORNE WRONGLY DENIED WASHINGTON’S MERETRICIOUS RELATIONSHIP DOCTRINE TO SAME-SEX COUPLES

Amanda J. Beane

Abstract: Washington’s property-division scheme for unmarried couples is among the most progressive in the nation. The scheme has evolved from a time when courts treated unmarried couples unfavorably and generally refused to divide their property equitably. The Washington Supreme Court took a step forward from this approach when it created the meretricious relationship doctrine. Under this doctrine, courts may equitably divide unmarried couples’ property at the termination of their relationship if the relationship was stable, marital-like, and the parties cohabited knowing they were not lawfully married. Now, however, the Washington Court of Appeals has restricted the application of this doctrine to heterosexual couples only, holding in Vasquez v. Hawthorne that same-sex relationships cannot qualify as meretricious relationships. Vasquez reasoned that because Washington law prevents same-sex couples from marrying, same-sex relationships cannot be “marital-like.” The Washington Supreme Court has granted review and should reverse Vasquez. “Marital-like” has been and should be defined by the conduct of the parties and not by their legal status. The Vasquez decision unjustly denies to same-sex couples an opportunity to benefit from a property-division scheme for unmarried couples by violating public policy and reflecting heterosexist views.

Until recently, Washington courts had treated unmarried couples unfavorably. The presumption governing property division, known as the Creasman presumption, provided that property acquired by an unmarried couple belonged solely to the one who held title.1 Recognizing the unfairness of this outcome, courts created certain exceptions to the Creasman presumption,2 until the Washington Supreme Court overruled the presumption in 1984 and created the meretricious relationship doctrine.3 Under the meretricious relationship doctrine, unmarried couples may take advantage of the community property scheme for married couples.4 For example, a woman who sacrificed her career, worked in the home, and helped with the family business, but who never married her male

partner, would have a claim for her share of the house, business, and other assets should the relationship end by separation or death. The woman would have to prove that she and her male partner were involved in a meretricious relationship: a stable, marital-like relationship where the partners cohabited knowing they were not lawfully married.5

A man, however, who also worked in the home and helped with the business, could not currently make the same claim should his relationship with his male partner dissolve. In *Vasquez v. Hawthorne*,6 the Washington Court of Appeals held that same-sex relationships cannot be meretricious relationships as a matter of law.7 The court held that because the Washington statute on prohibited marriages bars same-sex couples from marrying,8 same-sex relationships cannot be marital-like.9 Thus, same-sex relationships fail to qualify under a key element of the meretricious relationship doctrine.10 The court’s ruling denied Frank Vasquez his share of the estate he acquired with his male partner of twenty-eight years.11

The Washington Supreme Court should reverse *Vasquez*. The Washington Court of Appeals should have followed Washington precedent and defined “marital-like” in relation to the conduct of the couple involved instead of deferring to an unrelated marriage statute.12 The court confused marital-like conduct with marital-like status, ignoring that same-sex couples can and do engage in marital-like conduct and denying to Frank Vasquez what a Washington court would have granted to a woman in his place. The court also incorrectly applied marriage legislation to a property-division doctrine for unmarried couples and ignored the intent and purpose of the meretricious relationship doctrine.

In addition to being legally incorrect, the *Vasquez* decision reflects bad public policy and heterosexist bias. The *Vasquez* court violated public policy by deviating from just application of the law, the fair application of remedies, and the purpose underlying the meretricious

---

5. *Id.* at 346, 898 P.2d at 834.
7. *Id.* at 364, 994 P.2d at 241.
10. See *id*.
relationship doctrine. Finally, the Vasquez decision is heterosexist because it assumes that because same-sex couples may not enjoy marital status, they do not engage in marital-like relationships.

Part I of this Note explains the property rights of married, unmarried, and same-sex couples in Washington. Part II addresses the application of the meretricious relationship test since its development. Part III discusses the facts, procedural background, and the court’s opinion in Vasquez v. Hawthorne. Part IV argues that “marital-like” should be defined by conduct and not by legal status. Part V argues that Vasquez both violates public policy and is heterosexist. This Note concludes that Vasquez is a regressive, unjust step in Washington’s property-division scheme for unmarried couples and should be reversed.

I. THE PROPERTY RIGHTS OF MARRIED, UNMARRIED, AND SAME-SEX COUPLES IN WASHINGTON

Washington’s property-distribution scheme differs for married couples, unmarried heterosexual couples, and, after Vasquez, unmarried same-sex couples. Because Washington is a community property state, property of married couples is managed, characterized, and distributed according to principles of equality between husband and wife. For unmarried couples, courts may apply community property principles to the distribution of property at the termination of the relationship under the meretricious relationship doctrine. Property division at the end of a relationship that does not qualify as a meretricious relationship is dependent on contract, partnership and restitution remedies. Because same-sex couples may not marry, the Vasquez decision bars the only access same-sex couples have to community property principles.

A. Married Couples’ Property Rights Are Protected by Washington’s Statutory Community Property Regime

When a couple marries in satisfaction of the requirements of the Washington marriage statutes, the couple’s legal status automatically


15. Id. § 26.04.010 (stating marriage where husband or wife has not attained seventeen years of age is void except on showing of necessity); id. § 26.04.020 (stating marriage prohibited when either party has husband or wife living at time of marriage, husband and wife are of nearer relations.
enables each individual to claim community property rights.\textsuperscript{16} Community property rests on the presumption that the marital relationship is analogous to a partnership, with each spouse equally contributing to and benefiting from the marriage.\textsuperscript{17} Washington is one of nine states that apply community property principles regarding the characterization, management, and disposition of property acquired during a marriage.\textsuperscript{18} Washington statutes define community property and separate property, and, with the common law, govern the distribution of all property at the end of the marital relationship.

Washington statutory law labels community property as that property acquired during marriage by labor, industry, or other valuable consideration.\textsuperscript{19} In Washington, courts presume that property acquired during the marriage is community property,\textsuperscript{20} although each spouse may own property separately.\textsuperscript{21} Each spouse has equal management power over the community property and has testamentary power over his or her half of the community property.\textsuperscript{22}

Washington statutes define separate property as property acquired by gift, succession, inheritance, or other nonvaluable means.\textsuperscript{24} The Washington Supreme Court has also held that proceeds from a personal-injury action are separate property.\textsuperscript{25} Each spouse may manage his or her separate property as he or she desires.\textsuperscript{26}

Statutory schemes and common law govern the disposition of community property and separate property in the event of dissolution or

\textsuperscript{16} Id. § 26.04.160 (describing license and oath requirements for marriage).
\textsuperscript{17} Id. § 26.16.030.
\textsuperscript{18} Cross, supra note 17, at 17 n.3. The other community property states are Arizona, California, Idaho, Louisiana, Nevada, New Mexico, Texas, and Wisconsin. Id.
\textsuperscript{19} Id. at 27–28.
\textsuperscript{20} Id. at 28.
\textsuperscript{21} WASH. REV. CODE §§ 26.16.010–.020; Cross, supra note 17, at 19.
\textsuperscript{22} “Testamentary” is defined as relating to a will or testament. BLACK’S LAW DICTIONARY 1484 (7th ed. 1999).
\textsuperscript{23} Cross, supra note 17, at 18–19, 92.
\textsuperscript{24} WASH. REV. CODE §§ 26.16.010–.020; Cross, supra note 17, at 27.
\textsuperscript{25} In re Brown, 100 Wash. 2d 729, 730, 675 P.2d 1207, 1208 (1984).
\textsuperscript{26} Cross, supra note 17, at 19.
death. In marital dissolution, courts convert community property to separate property because the marital community ceases to exist, and then make a just and equitable distribution.\(^{27}\) If one spouse dies intestate,\(^{28}\) the surviving spouse takes the decedent’s one-half share of community property and retains his or her own one-half share.\(^{29}\) Furthermore, the surviving spouse takes the entire separate estate of the decedent spouse if that spouse has no surviving children, siblings, or parents,\(^{30}\) one-half of the separate estate if that spouse has surviving children,\(^{31}\) and three-quarters of the separate estate if that spouse has no offspring but has surviving parents or siblings.\(^{32}\)

**B. Cohabiting, Unmarried Couples’ Property Rights Are Protected by Washington Courts Through the Meretricious Relationship Doctrine**

Property distribution for cohabiting, unmarried couples in Washington has evolved over the last two decades from a rigid rule to an equitable approach. Historically, courts had held that property acquired by an unmarried couple belonged only to the one who held title.\(^{33}\) Although both in Washington\(^{34}\) and elsewhere\(^{35}\) courts developed various exceptions to the law, the Washington Supreme Court did not overturn the rule until 1984, when it created the meretricious relationship doctrine.\(^{36}\) That doctrine, further developed in 1995,\(^{37}\) dictates that parties to a stable, marital-like relationship who cohabit knowing they are not lawfully married may have their property justly and equitably distributed at the end of the relationship.\(^{38}\)

\(^{27}\) *Id.* at 113.

\(^{28}\) Intestacy occurs when one dies without a will. *Black’s Law Dictionary* 827 (7th ed. 1999).


\(^{30}\) *Id.* § 11.04.015(1)(d).

\(^{31}\) *Id.* § 11.04.015(1)(b).

\(^{32}\) *Id.* § 11.04.015(1)(c).


\(^{34}\) *Latham v. Hennessey*, 87 Wash. 2d 550, 553–54, 554 P.2d 1057, 1059 (1976); *infra* notes 44–51 and accompanying text.

\(^{35}\) *See*, e.g., *Marvin v. Marvin*, 557 P.2d 106, 116 (Cal. 1976).


\(^{38}\) *Id.* at 349, 898 P.2d at 835–36.
1. The Creasman Presumption: Parties Must Have Intended To Dispose of Their Property As They Actually Did

Before 1984, the Creasman presumption governed the property rights of cohabiting, unmarried couples in Washington. In Creasman v. Boyle, the Washington Supreme Court stated that the parties must have intended to dispose of their property as they actually did, such that property acquired by an unmarried couple belonged to the one who held legal title. The Creasman presumption disregarded contributions a party may have made to the acquisition of property. This holding may have been motivated by something other than property equity: Harvey Creasman was an African-American man living with a white woman, Caroline Paul. Although Creasman earned the money for the home he and Paul shared, the couple held title in Paul’s name only and made mortgage payments in Paul’s name. When Paul died intestate, the court awarded the home to her estate—not Creasman.

Washington courts devised exceptions to the Creasman presumption in order to divide equitably property acquired during an unmarried couple’s relationship. Situations qualifying for an exception included those where one or both parties entered a marriage in good faith that later proved to be void or where the title to property acquired during cohabitation could be traced to the separate property of one or both parties. Alternate theories also provided for equitable division of property acquired during marital-like cohabitation, such as joint venture, implied partnership, resulting trusts, and tenancy in com-

39. 31 Wash. 2d 345, 196 P.2d 835 (1948).
40. Id. at 356, 196 P.2d at 840–41.
41. At the very least, the holding ignored the social realities that may have discouraged an interracial marriage and kept the property in the white woman’s name. For instance, in 1967, seventeen states still had anti-miscegenation laws. Denise C. Morgan, Jack Johnson: Reluctant Hero of the Black Community, 32 Akron L. Rev 529, 544 n.58 (1999). The U.S Supreme Court invalidated such statutes in 1967. Loving v. Virginia, 388 U.S. 1, 2 (1967).
42. Creasman, 31 Wash. 2d at 345–50, 196 P.2d at 836–38.
43. Id. at 358, 196 P.2d at 842.
46. Shull v. Shepherd, 63 Wash. 2d 503, 507, 387 P.2d 767, 769–70 (1963); West v. Knowles, 50 Wash. 2d 311, 313, 311 P.2d 689, 691 (1957) (holding property acquired by both parties and held as tenants in common to be divided proportionally where separate property could be traced, otherwise divided equally).
47. In re Thornton, 81 Wash. 2d 72, 75, 499 P.2d 864, 865 (1972) (finding joint venture or implied partnership where woman contributed to success of business); Poole v. Schrichte, 39 Wash.
Furthermore, the Washington Supreme Court limited *Creasman* to its facts, where one party had died and the other was prohibited from testifying regarding the intent of the relationship because of the Dead Man’s Statute.51

2. *Marvin v. Marvin: The California Supreme Court Recognizes the Property Rights of Unmarried Couples*

In 1976, California became the first state to recognize the property rights of unmarried couples. In California, as in Washington, courts historically looked unfavorably on cohabiting parties who sued for a share in property acquired during the relationship.52 The California Supreme Court, however, recognized in *Marvin v. Marvin*53 the increasing number of couples living together without marrying54 and determined that agreements between nonmarital partners should be

---

2d 558, 564–65, 236 P.2d 1044, 1048–49 (1951) (finding joint venture or implied partnership where woman contributed money and labor to business and made house payments from profits). A joint venture is a business undertaking by two or more people. *See, e.g.*, id. at 564–65, 236 P.2d at 1048–49.

48. *Thornton*, 81 Wash. 2d at 73, 499 P.2d at 864–65; *Poole*, 39 Wash. 2d at 565, 236 P.2d at 1049. An implied partnership is a partnership implied by law when the actions and conduct of the parties demonstrate that they entered into a business relationship involving some combination of property, labor, skill, and experience. *Thornton*, 81 Wash. 23 at 79, 499 P.2d at 687–68.

49. *Walberg v. Mattson*, 38 Wash. 2d 808, 812–13, 232 P.2d 827, 829–30 (1951) (holding that Walberg owned property where evidence showed that Mattson only held title to avoid complications with Walberg’s estranged spouse and that Walberg had contributed nearly all of money for purchase of property). A court imposes a resulting trust when one party transferred property under circumstances implying the party did not intend for the transferee to retain legal title to the property. *See, e.g.*, id. at 812, 816, 232 P.2d at 829, 831. Compare a resulting trust to a “constructive trust,” where a court imposes a trust on equitable grounds against one who obtained the property by wrongdoing. BLACK’S LAW DICTIONARY 1514 (7th ed. 1999).

50. *Iredell v. Iredell*, 49 Wash. 2d 627, 631, 305 P.2d 805, 807 (1957) (holding that property acquired by both parties and held as tenants in common should be divided proportionally where separate property can be traced, otherwise divided equally). A tenancy in common is a tenancy shared by two or more persons in equal or unequal undivided shares with each having a right to possess the entire property but no right of survivorship. BLACK’S LAW DICTIONARY 1478 (7th ed. 1999).

51. *In re Lindsey*, 101 Wash. 2d 299, 302, 678 P.2d 328, 330 (1984). Washington’s Dead Man’s Statute is codified at WASH. REV. CODE § 5.60.030 (2000), which excludes testimony by an interested party regarding transactions with deceased, incompetent, or disabled persons. *Id.*


54. *Id.* at 109.
enforced unless the contract is founded explicitly on consideration for sexual services.\textsuperscript{55}

3. \textbf{The Washington Supreme Court Overruled Creasman and Developed the Meretricious Relationship Doctrine}

Eight years after \textit{Marvin}, the Washington Supreme Court expressly overruled \textit{Creasman} and replaced the \textit{Creasman} presumption with a new theory. Under \textit{In re Lindsey},\textsuperscript{56} courts were required to “examine the [meretricious] relationship and the property accumulations and make a just and equitable disposition of the property.”\textsuperscript{57} The court gave few guidelines for determining the existence of a meretricious relationship, stating that courts should examine each relationship on a case-by-case basis.\textsuperscript{58}

The Washington Supreme Court refined the meretricious relationship doctrine in \textit{Connell v. Francisco}\textsuperscript{59} by further defining a meretricious relationship and explaining how courts must distribute the property acquired in such a relationship.\textsuperscript{60} The court in \textit{Connell} defined a meretricious relationship as having three elements: the relationship must be stable, marital-like, and both parties must cohabit with knowledge that a lawful marriage between them does not exist.\textsuperscript{61} \textit{Connell} also listed five relevant factors for determining if a relationship is sufficiently stable and marital-like: continuous cohabitation, the duration of the relationship, the purpose of the relationship, any pooling of resources, and the intent of the parties.\textsuperscript{62}

\textit{Connell} then explained how a court must distribute the property acquired during a meretricious relationship. A court first finds whether a meretricious relationship existed; then, if it did, a court determines the interest each party has in the property acquired in the relationship and,

\begin{itemize}
\item \textsuperscript{55} \textit{Id.} at 113. \textit{Marvin} further held that in the absence of an express contract, a court should determine if the conduct of the parties created an implied contract, implied partnership, or joint venture or warranted the application of equitable remedies such as constructive trusts. \textit{Id.} at 110.
\item \textsuperscript{56} 101 Wash. 2d 299, 678 P.2d 328 (1984).
\item \textsuperscript{57} \textit{Id.} at 304, 678 P.2d at 331 (citing \textit{Latham v. Hennessey}, 87 Wash. 2d 550, 554, 554 P.2d 1057, 1059 (1976)).
\item \textsuperscript{58} \textit{Id.} at 305, 678 P.2d at 331.
\item \textsuperscript{59} 127 Wash. 2d 339, 898 P.2d 831 (1995).
\item \textsuperscript{60} \textit{Id.} at 349, 898 P.2d at 835–36.
\item \textsuperscript{61} \textit{Id.} at 356, 898 P.2d at 834.
\item \textsuperscript{62} \textit{Id.} See \textit{infra} Part II for further detail.
\end{itemize}
finally, justly and equitably divides it.63 Courts should only consider property for distribution that would have been community property had the parties been married.64 Community property legislation, however, is the only marriage legislation that may apply to meretricious relationships.65 The Connell court warned that courts may not equate meretricious relationships with marriage in other ways, thereby creating a type of common law marriage.66

With the meretricious relationship doctrine, Washington courts have developed an equitable-property doctrine from which any unmarried couple who has cohabited in a stable, marital-like relationship may benefit.67 The parties to a meretricious relationship do not have to prove the existence of a business partnership or establish any wrongdoing for a constructive trust. Instead, the party suing for equitable division of the property must show that the relationship was stable and marital-like, and that the couple cohabited knowing they were not in a lawful relationship.68 If these factors are met, the relationship is labeled “meretricious,” and the property is characterized as if it were community property and divided on a just and equitable basis.69

C. Same-Sex Couples’ Property Rights in Washington

Same-sex couples may only benefit from those property rights available to unmarried couples. Same-sex couples may not legally marry in Washington, as a result of Singer v. Hara70 and a 1998 amendment to the marriage statute defining prohibited marriages.71 Accordingly, same-sex couples must seek property remedies under the meretricious relationship doctrine or under contract and restitution remedies. If same-sex relationships cannot qualify as meretricious relationships, same-sex couples have more limited remedies than unmarried heterosexual couples

63. Connell, 127 Wash. 2d at 349, 898 P.2d at 835.
64. Id. at 352, 898 P.2d at 837.
66. Id. at 349–50, 898 P.2d at 836.
67. See id. at 349, 898 P.2d at 835–36.
68. Id. at 346, 898 P.2d at 834.
69. Id. at 349, 898 P.2d at 835.
when seeking equitable distribution of property at the termination of their relationships.\footnote{72}{See infra note 83 and accompanying text.}

Same-sex couples may not legally marry in Washington, according to Singer, which upheld the denial of a marriage license to two men in 1974.\footnote{73}{11 Wash. App. at 264, 522 P.2d at 1197.} The court held that a plain reading of the marriage statute did not authorize same-sex marriages.\footnote{74}{Id. at 249, 522 P.2d at 1189 (noting that although marriage statute at time had gender neutral language, 1970 amendment merely eliminated different marrying age requirements for men and women).} The court further denied the appellants’ claims under the newly enacted state Equal Rights Amendment\footnote{75}{Id. at 258–59, 522 P.2d at 1194; Equal Rights Amendment, WASH. CONST. art. 31, § 1.} and a federal equal protection claim,\footnote{76}{Singer, 11 Wash. App. at 261, 522 P.2d at 1196.} finding that the fundamental nature of marriage requires a union between a man and a woman.\footnote{77}{Id. at 253, 522 P.2d at 1191.}

The Washington Legislature amended the marriage statute on prohibited marriages in 1998 explicitly to ban same-sex marriages.\footnote{78}{Act of Feb. 6, 1998, ch. 1, 1998 Wash. Laws 1, codified at WASH. REV. CODE § 26.04.020(1)(c) (2000).} In the Defense of Marriage Act of 1996,\footnote{79}{28 U.S.C. § 1738C (Supp. IV 1998).} Congress provided that a state shall not be required to recognize marriages between persons of the same-sex where such a relationship is treated as a marriage in another state.\footnote{80}{Id.} Subsequently, the Washington Legislature passed the amendment to the marriage statute to ensure that Washington would not recognize same-sex marriages in the event they were legalized in other states.\footnote{81}{Act of Feb. 6, 1998, ch. 1, 1998 Wash. Laws 1, codified at WASH. REV. CODE § 26.04.020(1)(c)); E.S.H.B. 1130, 55th Leg., 1st Reg. Sess. (Wash.); see Baehr v. Lewin, 852 P.2d 44, 68 (Haw. 1993) (holding that state had burden of proving that prohibiting same-sex marriages furthered compelling state interests and did not abridge constitutional rights).} The marriage statute explicitly states that “it is a compelling interest of the state of Washington to reaffirm its historical commitment to the institution of marriage as a union between a man and a woman and to protect that institution.”\footnote{82}{WASH. REV. CODE § 26.04.010.}
Without legal marriage, same-sex couples are denied numerous benefits available to heterosexual couples upon marriage. Some of these benefits can be provided for by contract, such as creating decision-making powers by enacting durable powers of attorney, or distributing property by testamentary intent. However, without an express contract, same-sex couples will have to rely on equitable doctrines, such as constructive trust, resulting trust, or implied partnership, to determine property rights at the end of a relationship. Moreover, same-sex couples might seek to have their property equitably divided under the meretricious relationship doctrine.

Washington’s property-distribution scheme clearly differs for married and unmarried couples. Marriage offers the most equitable protection for the distribution of property because Washington adheres to the community property regime. Courts allow unmarried couples to take advantage of community property principles through the meretricious relationship doctrine. Before Vasquez, no court had distinguished Washington’s property scheme on the basis of sexual orientation and no rule barred same-sex couples from seeking the same remedies as unmarried heterosexual couples. Instead, courts evaluated unmarried couples’ property rights under the conduct-based meretricious relationship test.

II. WASHINGTON COURTS’ APPLICATION OF THE MERETRICIOUS RELATIONSHIP TEST

Washington courts determine if an unmarried couple’s relationship qualifies as a meretricious relationship by analyzing the couple’s conduct. Lindsey and Connell defined a meretricious relationship as having three elements: stability, marital-like behavior, and cohabitation with the knowledge that a lawful marriage does not exist. Courts analyze the couple’s conduct in five areas to decide if a relationship is

85. See, e.g., id.
“stable” or “marital-like”: continuous cohabitation, the duration of the relationship, the purpose of the relationship, the intent of the parties, and any pooling of resources.\(^\text{87}\) In practice, courts tend to analyze the continuity and duration of the relationship to determine stability and evaluate purpose, intent, and pooling of resources to determine marital-like behavior.\(^\text{88}\) These factors are not mutually exclusive; rather, courts view them as a whole to determine if a meretricious relationship exists.\(^\text{89}\)

A. A Meretricious Relationship Is a Stable Relationship

Courts examine the “stability” of a relationship to determine whether or not a relationship qualifies as meretricious.\(^\text{90}\) The continuity and duration of cohabitation indicate stability.\(^\text{91}\) Washington courts have refused to develop a bright-line rule regarding the continuity or duration of cohabitation to determine if a meretricious relationship is sufficiently stable.\(^\text{92}\)

Washington courts have alternatively affirmed and denied the existence of a meretricious relationship where the cohabitation was not continuous. For example, the Washington Court of Appeals has found a meretricious relationship despite the couple experiencing periods of separation where the couple had children and acquired a home together.\(^\text{93}\) Conversely, the Washington Supreme Court found that no meretricious relationship existed where in addition to the woman moving out twice and once living with another man, the parties did not share expenses for periods of time, did not mutually invest in a significant asset, and one refused to marry while the other wanted to.\(^\text{94}\)

Similarly, Washington courts have established no minimum time limit during which a couple must cohabit for the relationship to be

---

87. Connell, 127 Wash. 2d at 346, 898 P.2d at 834; Lindsey, 101 Wash. 2d at 304, 678 P.2d at 331.
89. Pennington, 142 Wash. 2d at 601–02, 14 P.3d at 770.
90. Connell, 127 Wash. 2d at 346, 898 P.2d at 834; Lindsey, 101 Wash. 2d at 304, 678 P.2d at 331.
91. See Pennington, 142 Wash. 2d at 603–04, 14 P.3d at 771.
92. Lindsey, 101 Wash. 2d at 305, 678 P.2d at 331.
94. Pennington, 142 Wash. 2d at 603–08, 14 P.3d at 771–73.
sufficiently stable. The Washington Court of Appeals held that a four-month relationship was sufficient to establish a meretricious relationship. Longer relationships may be helpful to prove the relationship was “meretricious,” but a lengthy relationship is not required.

B. A Meretricious Relationship Is a Marital-Like Relationship

A relationship must be marital-like in order for the courts to recognize it as a meretricious relationship. “Marital” is defined as “[o]f or relating to the marriage relationship,” while the marriage relationship can be defined by legal status or by the conduct of the parties. Therefore, when determining if a relationship is marital-like, courts can either evaluate the legal status or the conduct of the parties. Until Vasquez, Washington courts had used conduct to make this determination.

1. Marital Is Defined by Legal Status or the Conduct of the Parties

The word “marriage” can refer to a legal construct or to a social construct. As a legal construct, marriage refers to entrance into the marriage contract and to the assumption of legal benefits and obligations. As a social construct, marriage is defined by conduct befitting a marriage.

Legally, marriage is defined by its statutory requirements and a host of legal benefits and obligations. The essential requirements of a valid

100. See infra notes 101–04.
101. See supra note 15.
marriage are parties being legally able to contract, having mutual consent, and actually contracting as prescribed by law. Once two people are married, hundreds of Washington statutes and even more federal statutes confer various benefits and obligations upon the married couple as a result of the couple’s legal status.

Socially, marriage is manifested by the conduct between two people. For example, the act of marrying is often accompanied by a symbolic religious ceremony. In the marriage, the conduct typically includes living together and supporting each other physically, financially, and emotionally. The U.S. Supreme Court has recognized this conduct aspect of marriage by stating that marriage is “an association that promotes a way of life.”

2. Washington Courts Have Defined “Marital-Like” by Conduct

Washington courts have defined “marital-like” by examining the way the parties conduct themselves. Courts determine if a relationship is marital-like by examining the purpose of the relationship, the intent of the relationship, and the pooling of resources. Therefore, “marital-like” in the meretricious relationship context refers to “promot[ing] a way of life.”

Courts evaluate the purpose of a relationship to determine if a meretricious relationship is sufficiently marital-like. The Washington Court of Appeals has described the “purpose of the relationship” as the intent to “create a ‘long-term, stable, nonmarital family relationship.’” Evidence that the couple planned to spend the rest of their lives together is sufficient to establish this purpose. Other courts have found a
relationship to have a marital-like purpose if the relationship included elements like companionship, friendship, love, and mutual support. Courts have also found a meretricious relationship where the parties mutually support each other in work and leisure activities.

Courts also examine the intent to function as a family to determine if a meretricious relationship is sufficiently marital-like. Courts look for evidence that the couple intended to form a long-term, familial relationship such as intending to remain together as a couple. For example, the Washington Court of Appeals found that a meretricious relationship did not exist where the couple waited years to divorce their spouses and haphazardly shared a living space. The court could not determine from the evidence what the intent of the parties was regarding the relationship.

A party may evidence intent to remain together by showing that the couple modeled themselves after married couples. Courts, however, are inconsistent regarding whether or not the couple must have publicly acted as if they were married. Older, pre-Lindsey and Connell cases examined whether or not the couple “held themselves out as husband and wife.” More recently, the Washington Court of Appeals rejected the proposition that the parties must have publicly appeared as a marital-like couple.

---

112. Pennington, 142 Wash. 2d at 605, 14 P.3d. at 772.
114. See Connell, 127 Wash. 2d at 346, 898 P.2d at 834.
115. See, e.g., Pennington, 142 Wash. 2d at 603–05, 14 P.3d at 771–72 (finding one party’s refusal to marry and other party’s absences and relationship with third party negated intent; also finding that one party’s marriage to another negated intent); Zion Constr., Inc. v. Gilmore, 78 Wash. App. 87, 90, 895 P.2d 864, 866 (1995) (finding meretricious relationship where couple cohabited, pooled resources, and intended to marry).
117. Id.
118. The meretricious relationship cases involve one member of the relationship seeking a community property-like distribution of assets.
119. See In re Thornton, 14 Wash. App. 397, 403, 541 P.2d 1243, 1247 (1975) (holding facts did not support that couple ever held themselves out as husband and wife); Lailey v. Lailey, 43 Wash. 2d 192, 194–96, 260 P.2d 905, 906–7 (1953) (denying claim to title of car where parties did not live together, have marital relations, or hold themselves out as married couple).
120. Foster v. Thilges, 61 Wash. App. 880, 884–85, 812 P.2d 523, 525 (1991) (“[I]t is not necessary for a couple to represent themselves as husband and wife to establish a pseudomarital relationship.”).
Finally, a party did not have to be free to marry in order for a court to find that he or she was part of a meretricious relationship. In fact, marriage to a third person is not a per se rule against finding a meretricious relationship. Courts have found that marriage to another person is a significant factor that a meretricious relationship may not exist because it is evidence that at least one party lacked the intent to form a long-term, familial relationship. Courts, however, recognize that certain circumstances may support a finding of a meretricious relationship when one of the parties is married. Courts in at least three cases have found meretricious relationships despite one party’s marriage to a third party during a period of the meretricious relationship.

Courts also evaluate a couple’s pooling of resources to determine if a meretricious relationship is sufficiently marital-like. The pooling of resources may be tangible evidence that the couple existed in a mutually supportive, familial relationship. Courts generally find a meretricious relationship existed where couples pooled their resources for household and personal expenses, worked to improve their home together, and worked for each other without salary. The courts may look favorably upon couples having joint bank accounts, but sharing a bank account is not necessary.

---

123. *In re Pennington*, 93 Wash. App. 913, 919, 971 P.2d 98, 101–02 (1998), aff’d, 142 Wash. 2d 592, 14 P.3d 764 (2000) (distinguishing and not disapproving of *Foster*, where Washington Court of Appeals found meretricious relationship existed when one party was married to third party). The Washington Supreme Court affirmed *Pennington* and did not disapprove of *Foster* nor the court of appeals’s reading of *Foster*, *Pennington*, 142 Wash. 2d at 595–99, 14 P.3d at 767–69.
126. See *id*.
129. See *Anderson*, 1997 WL 6984, at *3 (finding meretricious relationship despite fact couple did not have joint bank account).
The Washington Supreme Court recently reaffirmed in *In re Pennington*\(^{130}\) that a meretricious relationship is defined by conduct. The court stated that “the term ‘marital-like’ is a mere analogy because defining meretricious relationships as related to marriage would create a de facto common-law marriage, which this court has refused to do.”\(^{131}\) Rather, courts evaluate factors such as the continuity of cohabitation and duration of the relationship, the purpose and intent of the parties, and the pooling of resources and services.\(^{132}\) One factor is not weighted more heavily than another.\(^{133}\) Instead, the factors are looked at as a whole to determine if a relationship is meretricious.

III. **VASQUEZ v. HAWTHORNE: THE WASHINGTON COURT OF APPEALS HELD THAT SAME-SEX RELATIONSHIPS CANNOT BE MERETRICIOUS AS A MATTER OF LAW**

In *Vasquez v. Hawthorne*,\(^{134}\) the Washington Court of Appeals held that one fact, the parties’ inability to marry, precluded the possibility of the couple existing in a meretricious relationship.\(^{135}\) Accordingly, unlike the trial court, the court of appeals found that same-sex relationships cannot be meretricious as a matter of law and refused to analyze the facts of the case.\(^{136}\) The decision is currently under review by the Washington Supreme Court.\(^{137}\)

**A. Factual Background**

The case stems from the long term relationship of Frank Vasquez and Robert Schwerzler;\(^{138}\) however, the trial court and the court of appeals provided limited information about Vasquez and Schwerzler’s relationship. The trial court simply found that Vasquez and Schwerzler

---

130. 142 Wash. 2d 592, 14 P.3d 764 (2000).
131. *Id.* at 602, 14 P.3d at 770.
132. *Id.*
133. *See id.*
135. *Id.* at 364, 994 P.2d at 241.
136. *Id.*
existed in a meretricious relationship. The court of appeals found that same-sex relationships cannot be meretricious as a matter of law. The court of appeals only noted the fact that Vasquez and Schwerzler cohabited for twenty-eight years, living together from 1967 until Schwerzler died in 1995.

Vasquez’s briefs, affidavits, and the affidavits of acquaintances and friends of Vasquez and Schwerzler support the view that Vasquez and Schwerzler existed in a marital-like relationship. These materials show that Vasquez and Schwerzler supported each other personally and professionally. Vasquez and Schwerzler also held themselves out in public as an intimate couple in certain circumstances.

Vasquez and Schwerzler extensively supported each other in their personal lives. Vasquez performed many household chores, such as cooking and cleaning, and cared directly for Schwerzler by cutting Schwerzler’s hair and nails. He referred to Schwerzler as “my husband.” Vasquez did not have a bank account of his own and relied on Schwerzler to take care of the finances. Schwerzler was seventeen years older than Vasquez and frequently assured Vasquez that

141. Id.
142. Because the trial court denied defendant Hawthorne’s motion to strike the affidavits supporting plaintiff Vasquez’s motion for summary judgment and therefore presumably relied upon this material in deciding that Vasquez and Schwerzler had a meretricious relationship, plaintiff’s briefs and affidavits are relied upon here. Order, supra note 139, at 3.
144. Id. at 3; Declaration of Brian Sword in Support of Motion for Partial Summary Judgment at 4, Vasquez, No. 96-2-07794-7 (Pierce County Super. Ct. Dec. 1, 1996) [hereinafter Sword Declaration]; Affidavit of Frank Vasquez in Support of Motion for Partial Summary Judgment at 4, Vasquez, No. 96-2-07794-7 (Pierce County Super. Ct. Dec. 20, 1996) [hereinafter Vasquez Affidavit].
146. Plaintiff’s Memorandum, supra note 143, at 2–3.
147. Id.
149. Vasquez Affidavit, supra note 144, at 2.
Vasquez v. Hawthorne

he would be taken care of if he died first,\textsuperscript{151} even offering to sell their home and give Vasquez half of the money.\textsuperscript{152}

Vasquez and Schwerzler also shared responsibilities in the bag-recycling business they managed from their home. Schwerzler bought and sold bags and managed the finances.\textsuperscript{153} Vasquez performed the physical labor required for this business and took phone calls from clients.\textsuperscript{154} Schwerzler never paid Vasquez for his work.\textsuperscript{155}

Schwerzler and Vasquez circulated as a couple among some friends,\textsuperscript{156} but among Schwerzler’s heterosexual friends and family, Schwerzler was apparently ambiguous or secretive about his relationship with Vasquez. For example, one of his friends assumed Vasquez and Schwerzler were a couple because of Schwerzler’s conduct, such as always letting Vasquez know when he expected to be home and bringing Vasquez lunch during the working day.\textsuperscript{157} Two of Schwerzler’s siblings believe he was not gay,\textsuperscript{158} perhaps because Schwerzler had been married and divorced three times.\textsuperscript{159}

Upon his death, Schwerzler left no will that was located.\textsuperscript{160} The home, the checking account, and other assets including a life-insurance policy, an automobile, and proceeds from the sale of a second automobile were in Schwerzler’s name.\textsuperscript{161} Schwerzler’s life-insurance proceeds were paid to the personal representative of his estate.\textsuperscript{162}

\begin{itemize}
  \item \textsuperscript{151} Plaintiff’s Memorandum, supra note 143, at 4; Vasquez Affidavit, supra note 144, at 6; Sword Declaration, supra note 144, at 3.
  \item \textsuperscript{152} Vasquez Affidavit, supra note 144, at 6.
  \item \textsuperscript{153} Id. at 4.
  \item \textsuperscript{154} Plaintiff’s Memorandum, supra note 143, at 3; Sword Declaration, supra note 144, at 4; Vasquez Affidavit, supra note 144, at 4.
  \item \textsuperscript{155} Plaintiff’s Memorandum, supra note 143, at 4; Vasquez Affidavit, supra note 144, at 5.
  \item \textsuperscript{156} Kinzner Declaration, supra note 145, at 1–2.
  \item \textsuperscript{157} Declaration of Robert Whitworth in Support of Motion for Partial Summary Judgment at 2, Vasquez v. Hawthorne, No. 96-2-07794-7 (Pierce County Super. Ct. Jan. 17, 1997).
  \item \textsuperscript{158} Brief of Appellant Joseph Hawthorne at 29, Vasquez, 99 Wash. App. 363, 994 P.2d 240 (No. 96-2-07794-7).
  \item \textsuperscript{159} Id. at 12.
  \item \textsuperscript{160} Vasquez, 99 Wash. App at 364, 994 P.2d at 241.
  \item \textsuperscript{161} Id. at 364, 994 P.2d at 241; Order, supra note 139, at 4.
  \item \textsuperscript{162} Plaintiff’s Memorandum, supra note 143, at 19.
\end{itemize}
B. Procedural Background and Status

Because Schwerzler died intestate, his estate passed to his siblings under Washington’s intestacy statute. Vasquez sued for his share of Schwerzler’s estate under the theories of meretricious relationship, implied partnership, and constructive trust. The trial court awarded the entire estate to Vasquez under the meretricious relationship theory on Vasquez’s motion for partial summary judgment. Joseph Hawthorne, the estate’s personal representative, appealed, and a unanimous court of appeals reversed, holding that a same-sex relationship could not be a meretricious relationship as a matter of law.

The Washington Supreme Court granted review of Vasquez. The issue of whether or not same-sex couples can sue for property division under the meretricious relationship theory has never before been decided in Washington. Vasquez’s theories of implied partnership and constructive trust have not yet been tried.

C. The Court of Appeals Opinion

The Washington Court of Appeals unanimously reversed the trial court judgment, holding that same-sex relationships cannot qualify as meretricious relationships in Washington. The court’s brief opinion, after summarizing the development of the meretricious relationship doctrine, focused solely on the “marital-like” element of the meretricious relationship test. The court stated that “because persons of the same sex may not be legally married, a meretricious relationship cannot exist between members of the same sex.” Thus, Vasquez received none of the property he acquired with Schwerzler during their twenty-eight year intimate relationship.

165. Id. at 364, 994 P.2d at 241.
166. Id. at 364–65, 994 P.2d at 241.
169. Id. at 365, 994 P.2d at 241.
170. Id. at 364, 994 P.2d at 241.
171. Id. at 367–68, 994 P.2d at 242–43.
172. Id. at 365, 994 P.2d at 241.
Vasquez v. Hawthorne

The court first reviewed the history of the meretricious relationship doctrine. The court noted its creation in Lindsey, its further development in Connell, and that the doctrine overturned the Creasman presumption. The court restated Connell’s holding that the only property that can be distributed upon a finding of a meretricious relationship is that property which would have been community property had the parties been married.

The Vasquez court then reasoned that if the only property available for distribution at the end of a meretricious relationship is property that would have been community property if the parties had been married, then a meretricious relationship must be one in which the parties may legally marry. To the Vasquez court, only potentially marital relationships can benefit from quasi-marital protection. Thus, the parties’ legal status is a conclusive factor in determining whether or not the couple can exist in a meretricious relationship.

The Vasquez court then noted that Washington’s marriage statute prohibits legal marriage between same-sex partners. The court concluded that this statutory definition of legal marriage prohibits the equitable doctrine of meretricious relationships from ever including same-sex relationships. As support for this conclusion, the Vasquez court held that Connell and its progeny must have “implicitly assumed that a meretricious relationship can exist only between a man and a woman.” This implicit assumption led the Vasquez court to conclude that only the legislature could apply the judicial doctrine of meretricious relationship to same-sex couples. As a matter of law, the court held that same-sex relationships could not be marital-like.

---

173. Id. at 366, 994 P.2d at 242.
174. Id. at 367.
175. Id. at 367, 994 P.2d at 242.
176. Id. at 369, 994 P.2d at 243. The Vasquez court used “quasi-marital” synonymously with “marital-like,” calling a meretricious relationship a quasi-marital relationship. See id. at 368, 994 P.2d at 243.
177. Id. at 367–68, 994 P.2d at 242–43.
178. Id. at 367, 994 P.2d at 242.
179. Id. at 369, 994 P.2d at 243. The court noted that Vasquez had the recourses of constructive trust and implied partnership. Id.
180. Id. at 364, 994 P.2d at 241.
IV. THE VASQUEZ COURT ERRONEOUSLY CONCLUDED THAT THE MERETRICIOUS RELATIONSHIP DOCTRINE DOES NOT APPLY TO SAME-SEX COUPLES

The Vasquez court should have defined “marital-like” according to the conduct of the parties and not according to their legal status. In determining that, as a matter of law, same-sex relationships cannot qualify as meretricious relationships, the Vasquez court erred in three ways. First, the court strayed from precedent by crafting a bright-line rule that status determines who may benefit from the meretricious relationship doctrine. Second, the court erred in using the Washington marriage statute regarding prohibited marriages to modify the judicial doctrine of meretricious relationships. Third, in applying the statute, the court wrongly interpreted the intent and purpose of the meretricious relationship doctrine.

A. The Vasquez Court Deviated from Precedent by Crafting a Bright-Line Rule

The bright-line rule established in Vasquez deviates from all previous application of the meretricious relationship doctrine. In determining whether a meretricious relationship exists, courts must analyze all the relevant evidence of the parties’ conduct on a case-by-case basis.\textsuperscript{181} No one factor weighs more than another in importance.\textsuperscript{182} The Vasquez court strayed from precedent by deciding that the parties’ legal ability to marry outweighed the totality of the parties’ conduct.\textsuperscript{183} The court further wrongly crafted a bright-line rule when deciding the parties must be male and female. Moreover, the Vasquez court misinterpreted precedent when it used marriage legislation to draw this bright-line rule.

Precedent does not support the Vasquez court’s conclusion that the marriage statute\textsuperscript{184} determines who may enter into meretricious relationships. Courts have found meretricious relationships when one of

\begin{itemize}
  \item \textsuperscript{181} In re Pennington, 142 Wash. 2d 592, 601–02, 14 P.3d 764, 770 (2000); In re Lindsey, 101 Wash. 2d 299, 305, 678 P.2d 328, 331 (1984).
  \item \textsuperscript{182} Pennington, 142 Wash. 2d at 602, 14 P.3d at 770.
  \item \textsuperscript{183} Vasquez, 99 Wash. App. at 369, 994 P.2d at 243.
  \item \textsuperscript{184} WASH. REV CODE § 26.04.020 (2000).
\end{itemize}
the parties has been married to a third party, yet, the marriage statute does not allow marriage when one or both parties are married to another person. In *Warden*, *Kinzer*, and *Foster* the Washington Court of Appeals found meretricious relationships existed even when one party was married to another for a portion of the meretricious relationship. The Washington Court of Appeals has indicated that one party’s marriage to a third party was not a dispositive factor where a relationship was long-term and the parties pooled their resources for many purposes. The Washington Supreme Court recently noted that marriage to a third party may be evidence of one party’s lack of intent regarding the relationship, but reaffirmed its commitment to determining the existence of meretricious relationship on a case-by-case basis and did not craft a bright-line rule regarding the parties’ legal status. Before *Vasquez*, Washington courts did not look to the legal status of the couple to determine if the conduct was marriage-like.

Furthermore, no court before *Vasquez* had held that one of the requirements for a meretricious relationship is that the parties be one man and one woman. The *Vasquez* court was unable to point to any specific language in previous cases and could only find “that these courts implicitly assumed that a meretricious relationship can exist only between a man and a woman.” Yet, because no fact scenarios had been presented other than those that dealt with opposite-sex couples, no cases had considered whether a meretricious relationship hinged on the gender of the parties. By giving legal significance to assumptions, the *Vasquez* court equated mind reading with the interpretation of precedent.

Finally, *Vasquez* misinterpreted *Connell* when it used marriage legislation to craft its bright-line rule that same-sex relationships could not be meretricious relationships. The *Vasquez* court reasoned that if

---

courts distribute property at the termination of a meretricious relationship based on what would have been community property if the parties had been married, then the parties must be legally able to marry in order to exist in a meretricious relationship. Yet this reasoning inverts the logic of Connell. Connell requires that courts first find that a meretricious relationship existed; second, determine the interest each party has in the property; and third, distribute the property according to principles of community property. Thus Connell called for the use of marriage legislation as a guide for the third prong. Vasquez found the marriage statutes not only determinative of how to distribute property, but determinative of whether to distribute the property at all, thus applying marriage legislation to the first prong. The court in Connell, however, explicitly stated that “a meretricious relationship is not the same as marriage.” Using marriage legislation to craft a bright-line rule for what constitutes a meretricious relationship finds no support in the Connell court’s recourse to community property statutes to make a just and equitable distribution of property.

The Vasquez decision is an anomaly among meretricious relationship cases that have used a case-by-case, factor-based approach and that have only used marriage legislation in determining how, not whether, to distribute the property. Courts have not established any bright-line rules when deciding if a relationship qualifies as meretricious. Vasquez is the first case to draw a bright-line rule, and it hides behind an irrelevant statute to do so.

B. The Vasquez Court Deviated from Statutory Interpretation by Applying a Washington Marriage Statute to the Common Law Doctrine of Meretricious Relationships

The Vasquez court, in holding that the statutory limitations on who may marry in Washington are determinative of who can engage in a

195. Vasquez, 99 Wash. App. at 369, 994 P.2d at 243. Although the court claims that the “statutory limitations on who may marry . . . are relevant in determining whether a relationship is sufficiently ‘marital-like,’” the use of the word “relevant” is misleading. The court’s holding, that same-sex relationships cannot be marital-like because same-sex couples may not marry, renders the statutory limitations determinative. Id. at 367–68, 994 P.2d at 242–43.
meretricious relationship, deviated from traditional statutory interpretation. Neither a plain reading of the marriage statute prohibiting certain marriages nor the legislative history of the amendment prohibiting same-sex marriages suggests that the marriage statute modifies the judicially created doctrine of meretricious relationships. The Vasquez court’s reliance on the Washington marriage statute was misplaced.

A plain reading of the marriage statute prohibiting certain marriages does not support the Vasquez court’s conclusion that the marriage statute determines who may enter into meretricious relationships. The statute is titled “Prohibited Marriages” and contains a list of who may not marry in Washington. The statute does not address any other legal rights or remedies, define who may cohabit together or form intimate relationships, or criminalize certain sexual relationships. The statute nowhere defines property rights. Rather, the plain language of the statute defines only who may undertake the legal rights and obligations of marriage. The Vasquez court, however, applied the statute in deciding who may benefit from the meretricious relationship doctrine, despite the Washington Supreme Court’s statements that meretricious relationships are “wholly unrelated” to and “not the same as” marriage.

The legislative history of the subsection of the marriage statute that specifically prohibits same-sex marriages also does not support the Vasquez court’s conclusion that the statute determines who may enter meretricious relationships. The Legislature adopted the amendment specifically prohibiting same-sex marriage in response to Congress’s Defense of Marriage Act, which authorized states not to recognize same-sex marriages in other states. The Defense of Marriage Act affected marriage and not any common law doctrines. The legislative history of the Washington amendment shows no indication the Legislature

199. Id.
200. Id.
201. Id.
203. In re Pennington, 142 Wash. 2d 592, 600, 14 P.3d 764, 769 (2000).
206. 28 U.S.C. § 1738C.
intended to restrict other types of rights and remedies applicable to same-sex couples.207 The amendment created a single legal disability: a limit on the right to marry. Neither a plain reading of the statute nor the legislative history supports the Vasquez court’s conclusion that the marriage statute prohibiting certain marriages should determine that same-sex relationships cannot be marital-like.

C. The Vasquez Court Deviated from the Intent and Purpose of the Meretricious Relationship Doctrine

The Vasquez court deviated from the intent and purpose of the meretricious relationship doctrine when it held that in order to be marital-like parties must have the legal ability to marry.208 The meretricious relationship doctrine is a judicial remedy, created with the intent to protect the property rights of unmarried couples. The meretricious relationship doctrine’s sole purpose, however, is to address property rights. This limitation explicitly prevents meretricious relationships from being common law marriages.209 Implicitly, it preserves the incentives for couples to marry.

Washington courts created the meretricious relationship doctrine to protect the property rights of people who are not legally married. When creating the meretricious relationship doctrine and overruling the Creasman presumption, the Washington Supreme Court recognized that the Creasman presumption unjustly rewarded the “cunning and the shrewd” who could deliberately hold title to all property at the end of a meretricious relationship despite the fact that the property may have been acquired through joint efforts.210 The meretricious relationship doctrine balances the scales between the shrewd party and the party without title by dividing property justly and equitably without regard to who holds title.211 By excluding same-sex couples from the meretricious

relationship doctrine, *Vasquez* excludes those who the doctrine was meant to protect.

While the meretricious relationship doctrine protects unmarried couples, the doctrine addresses a much narrower spectrum of legal obligations and benefits than marriage. The meretricious relationship doctrine is solely an equitable property doctrine that distributes property acquired during a stable, marital-like relationship at the termination of that relationship.\(^ {212}\) A meretricious relationship does not create a common law marriage,\(^ {213}\) as Washington does not allow common law marriages\(^ {214}\) and the court in *Connell* specifically held that couples in meretricious relationships may not be treated as if they were married.\(^ {215}\) The meretricious relationship does not establish rights of survivorship, affect tax status, or grant any decision-making power if one party to the relationship is debilitated. In contrast, legal marriage confers these benefits and many others upon the involved parties.\(^ {216}\)

Because the intent and purpose of the meretricious relationship doctrine is to provide only limited rights for unmarried couples, the meretricious relationship doctrine does not detract from the state’s role in encouraging couples to marry legally. The state values marriage as a “historical . . . institution.”\(^ {217}\) Any fear the *Vasquez* court had that allowing same-sex couples to benefit from the meretricious relationship doctrine threatens the institution of marriage is unfounded. Same-sex marriage is already prohibited by statute,\(^ {218}\) it is only heterosexual couples who may choose whether or not to legally marry, and courts have not hesitated in applying the meretricious relationship doctrine to heterosexual relationships. Moreover, because the meretricious relationship doctrine provides only limited rights for unmarried couples as opposed to those conferred by marriage\(^ {219}\) and does not create a common law marriage,\(^ {220}\) couples have more of an incentive to marry than to

\(^{212}\) *Id.*

\(^{213}\) A common law marriage is one where the couple is recognized as legally married after having cohabited for a certain period of time. *Cross*, *supra* note 17, at 20.

\(^{214}\) *Id.*

\(^{215}\) *Connell*, 127 Wash. 2d at 350, 898 P.2d at 836.


\(^{219}\) *See Connell*, 127 Wash. 2d at 350, 898 P.2d at 836.

\(^{220}\) *Id.*
cohabit without legal marriage. In fact, same-sex couples have continually sought the legal right to marry and take on the many benefits and burdens of legally sanctioned marriage.221

The Vasquez court ignored the intent and purpose of the meretricious relationship doctrine when it held that to create a meretricious relationship, the parties must be free to marry under Washington statutory law. The meretricious relationship doctrine is simply an equitable doctrine providing limited protection in the area of property distribution for unmarried couples. It is a judicially crafted doctrine that provides equity precisely where statutory definitions do not apply. The doctrine’s purpose, statutory interpretation of the statute on prohibited marriages, and precedent support applying the meretricious relationship doctrine to same-sex couples. By looking solely to Frank Vasquez’s statutory status, the Vasquez court neglected its equitable duty to define meretricious relationships on the basis of conduct.

V. THE VASQUEZ DECISION VIOLATES PUBLIC POLICY AND IS HETEROSEXIST

Because same-sex couples engage in marital-like conduct, the court should allow same-sex couples to benefit from the meretricious relationship doctrine. The Vasquez court violated public policy by denying the meretricious relationship doctrine to same-sex couples. The court further demonstrated its own heterosexism by holding that same-sex relationships cannot be marital-like.

A. Public Policy Dictates that Same-Sex Couples Should Benefit from the Meretricious Relationship Doctrine

The Washington Supreme Court should apply the meretricious relationship doctrine to same-sex couples because of public policy. The Vasquez court denied an equitable remedy to a group of Washington citizens. Public policy supports just applications of law, just remedies, and the protection of Washington citizens.

Just application of the law requires that same-sex relationships be eligible to qualify as meretricious relationships. Vasquez retains a right for unmarried couples who have the choice to marry but denies it to

couples who do not have the choice to marry but may wish to do so. Same-sex couples across the country desperately want to take advantage of both the rights and obligations offered by marriage. Legislatures have exercised their power to deny such legal benefits to same-sex couples. Vasquez punishes a group of Washington state citizens who have already been excluded from the legal institution of marriage by denying them a simple property remedy because of that legal status. If the courts create a property doctrine for unmarried couples, that property doctrine should justly apply to all unmarried couples, and not just heterosexual couples who already have the benefit of legal marriage.

The court attempted to justify its decision by noting that Frank Vasquez might seek relief under alternate theories; however, not all parties to a same-sex relationship will be able to prove implied partnership or constructive trust. Constructive trust has traditionally required evidence of fraud, breach of fiduciary responsibility or contract, or overreaching. Implied partnership requires evidence of the intention of the parties either by direct evidence or from the circumstances. It may be difficult for someone in Vasquez’s position to prove such intent where one partner is deceased. Under the meretricious relationship doctrine, intent is just one of the factors a court may rely on to find a meretricious relationship existed.

Nor are the remedies under the implied-partnership or constructive-trust theories just when compared to the property-distribution principles of the meretricious relationship doctrine. Parties that could prove such remedies as constructive trust or implied partnership do not have the advantage of the presumptive one-half interest inherent in community property principles or the just and equitable division used under the meretricious relationship doctrine. Just and equitable division allows

---

224. See id.
227. Id. at 401, 541 P.2d at 1246 (citing Nicholson v. Kilbury, 83 Wash. 196, 202, 145 P. 189, 191–92 (1915)).
229. See supra Part I.A.
courts to take into account the needs of the parties. On the other hand, remedies such as constructive trust are restitution remedies, and simply restore the parties to the positions they formerly occupied. For example, under a restitution remedy Vasquez might receive the equivalent of minimum wage for his labor in the bag business, but not the presumptive one-half interest in the assets acquired during the relationship that results when a couple has engaged in a marital-like partnership.

Finally, Vasquez rewards the “cunning and the shrewd.” The Washington Supreme Court specifically attacked the Creasman presumption for rewarding the “cunning and the shrewd” who may hold property in their name while acquiring it with an unmarried partner and later overturned the Creasman presumption. Vasquez reinstates the Creasman presumption for same-sex couples, requiring them to prove alternate contract or restitution theories. The meretricious relationship doctrine should protect same-sex partners who may not legally marry and may not possess the resources or legal knowledge to create contractual property rights.

B. The Vasquez Decision Is Heterosexist

Vasquez is a heterosexist decision because it holds that same-sex relationships cannot be marital-like. Heterosexism differs from homophobia. Homophobia implies an irrational fear and hatred of gay or lesbian people. Heterosexism “is the systematic implementation, through law, custom, and other vehicles, of the view that heterosexuality is ‘natural’ and that lesbian and gay sexuality are morally and socially inferior.” The Vasquez opinion reflects a belief that gay and lesbian relationships are inferior to heterosexual ones.

231. See Poole v. Schrichte, 39 Wash. 2d 558, 566–67, 236 P.2d 1044, 1049–50 (1951) (citing Buckley v. Buckley, 50 Wash. 213, 223, 96 P. 1081, 1083 (1908) (awarding one-quarter of property to ex-wife and innocent wife each and man one-half because man aged, feeble, and in need of medical support)).

232. RESTATEMENT OF RESTITUTION § 1 cmt. a (1937); see also id. § 160.

233. West v. Knowles, 50 Wash. 2d 311, 316, 311 P.2d 689, 693 (1957) (Finley, J., concurring).

234. Id.


237. Id.
Vasquez represents a heterosexist assumption that gay and lesbian couples are not capable of engaging in marital-like conduct. Vasquez referred to the marriage statute prohibiting certain marriages in order to hold that same-sex couples cannot exist in marital-like relationships, drawing a bright-line for the first time in the history of the meretricious relationship doctrine.238 The court thereby avoided a conduct-based analysis of the term “marital-like.”

By using the status-based construction of “marital-like,” the court assumed that status determines conduct. This is a false assumption. One can easily imagine an opposite-sex couple legally married for immigration reasons and acting both privately and publicly as if they are not married; or, alternatively, a couple acting privately and publicly as if they were married but not being legally married. This is also a false assumption as it applies to same-sex couples. Gay and lesbian couples can and do engage in marital-like relationships, holding commitment ceremonies, raising children, and staying together until death.239 Their status in society does not ultimately determine their ability to engage in such relationships.

Vasquez and Schwerzler’s relationship was an example of a meretricious relationship. A relationship must be stable and marital-like in order to be a meretricious relationship.240 By analyzing the duration and continuity of their cohabitation,241 it is evident that Vasquez and Schwerzler had a stable relationship, as Vasquez and Schwerzler cohabited without interruption for twenty-eight years.242 Furthermore, their conduct exemplified a marital-like relationship. Vasquez and Schwerzler mutually supported each other in a familial and business context, had an intent to form such a supportive relationship, and they pooled their resources.243

Neither state nor federal laws prevent heterosexist decisions by the courts. The court in Singer declined to extend Washington’s Equal Rights Amendment244 to provide marriage rights to same-sex couples.245

241. See supra Part II.A.
243. See supra Part III.A.
244. WASH. CONST. art. 31 §1.
The supreme court has not labeled gays, lesbians, and bisexuals a protected class for the purposes of federal equal protection. Until federal equal protection includes gays, lesbians, and bisexuals as a protected class, or Washington’s Equal Rights Amendment is interpreted to prohibit the denial of same-sex marriage, Vasquez represents a kind of decision that courts are technically free to make.

Reality, however, requires that the Washington Supreme Court recognize that gay and lesbian couples can and do engage in marital-like conduct. Justice, therefore, demands that same-sex couples be offered the limited protections of the meretricious relationship doctrine. Precedent, an interpretation of the statute prohibiting certain marriages, the intent and purpose of the meretricious relationship doctrine, and public policy support a conduct-based definition of marital-like that can and should apply to same-sex couples.

Creasman ignored the social realities of an interracial couple in 1948 and promulgated a presumption steeped in racism; Vasquez reinstitutes the Creasman presumption for same-sex couples because of heterosexist bias. The court should not impose the regressive Creasman presumption on same-sex couples. Rather, the court should remember what it said while admonishing lower courts’ application of the Creasman presumption in the pre-Lindsey era: “Property rights are not determined on the basis of social relationships, moral or immoral.”

VI. CONCLUSION

In 1948, the Washington Supreme Court ignored the social realities of an interracial couple and promulgated the Creasman presumption, holding that property acquired by an unmarried couple must belong to the one who holds title. For years, courts noted the injustice of this rule and created numerous exceptions. Finally, the Washington Supreme

246. See Bowers v. Hardwick, 478 U.S. 186, 195–96 (1986). It is arguable that denying the meretricious relationship doctrine to same-sex couples violates the Fourteenth Amendment under Romer v. Evans, 517 U.S. 620 (1996). This Note does not reach this constitutional issue.
247. See supra notes 181–96 and accompanying text.
248. See supra notes 197–207 and accompanying text.
249. See supra notes 208–21 and accompanying text.
250. See supra notes 222–35 and accompanying text.
Court recognized the need for a property doctrine that would equitably divide the property of unmarried couples upon the termination of their relationships, overruled the Creasman presumption, and created the meretricious relationship doctrine. In Vasquez, Washington took a step backward by withholding that doctrine from same-sex couples. By ignoring the social realities of same-sex couples, the Vasquez court denied to Frank Vasquez what a court would have given to a woman in his place, and it assumed that because same-sex couples may not be legally married, same-sex couples cannot engage in marital-like conduct.

The Washington Supreme Court should reverse Vasquez and hold that “marital-like” is determined by conduct and not by legal status. The previous application of “marital-like,” the plain language and history of the marriage statute prohibiting certain marriages, the intent and purpose of the meretricious relationship doctrine, public policy, and the need to combat heterosexist bias support such a holding. Moreover, same-sex couples can and do engage in marital-like conduct, just as Frank Vasquez and Robert Schwerzler did.
A SHIELD, NOT A SWORD: INVOLUNTARY LEAVE UNDER THE FAMILY AND MEDICAL LEAVE ACT

Megan E. Blomquist

Abstract: Under the Family and Medical Leave Act of 1993 (FMLA), covered employers must grant an eligible employee’s request for up to twelve weeks of unpaid leave to care for a new baby or an ill family member, or to accommodate the employee’s own serious health condition. The statute prohibits employers from interfering with, restraining, or denying an employee’s right to FMLA leave. Neither the statute itself nor its regulations directly address the practice of involuntary leave, a term that has been used to describe instances where an employer designates the leave of a qualifying employee as FMLA leave without an employee request or against the employee’s wishes and instances where an employer places a non-qualifying employee on FMLA leave. The few federal courts that have examined this issue have failed to make this distinction and have interpreted the statute’s silence to permit employers to put employees on involuntary FMLA leave in both instances. This Comment examines the text of the FMLA, its implementing regulations, and its legislative history to ascertain both Congress’s intent and the Department of Labor’s interpretation regarding involuntary leave. This Comment concludes that employers who put non-qualifying employees on involuntary FMLA leave violate the FMLA’s prohibition on employer interference with employees’ FMLA rights.

After recovering from complicated brain surgery to repair two aneurysms, Ed Keating returned to his job as a freight handler in the receiving department of a volume discount store. Although Mr. Keating suffered some permanent short-term memory impairment as a result of the surgery, he successfully performed his essential job functions for the next nine years. However, when a new supervisor began assigning him new and more complex tasks, Mr. Keating found he needed minor accommodations such as a daily task list. Despite Mr. Keating’s willingness and demonstrated ability to work with only minor accommodations, his employer placed him on twelve weeks of unpaid leave under the Family and Medical Leave Act (FMLA) while it purportedly determined whether accommodations were necessary and feasible. During this leave, Mr. Keating had no income and did not qualify for unemployment benefits. As a result, he had no choice but to declare bankruptcy when he could not pay his creditors.

1. This illustration is loosely based on a case pending in the District Court for the Western District of Washington. The plaintiff’s name has been changed to protect his privacy.

Given the prospect of twelve weeks without income, most employees in this situation would be forced to resign and seek work elsewhere. Even if Mr. Keating had found another job, he still would have been unable to take time off later that year in the event of another illness in his family, because he would no longer have been eligible for FMLA leave. Had Mr. Keating’s employer ultimately agreed to accommodate his disability and return him to his position, he eventually may have been able to recoup some of his financial losses and restore his economic security, but he would have been similarly unable to take FMLA leave later that year. In either case, Mr. Keating would have lost something else of great value: his annual entitlement to FMLA leave.

The FMLA entitles eligible employees to take twelve weeks of unpaid leave per year for specific reasons, including the employee’s own serious health condition, and to return to the same or an equivalent position upon completion of leave. Examined in light of the legislative history, the statutory provisions and accompanying regulations make clear the FMLA’s intent to grant employees a right to request leave when they find it necessary. The exercise of this right cannot be waived, interfered with, restrained, or denied. Nevertheless, in situations such as Mr. Keating’s, employers have used the FMLA as a sword to deal with difficult or undesirable employees by placing them on involuntary FMLA leave. The few courts that have examined the issue of involuntary leave, where an employer places an employee who does not meet the statutory requirements on FMLA leave, have held that such a practice does not violate the employee’s FMLA rights.

Part I of this Comment examines the development and implementation of the Family and Medical Leave Act. Part II analyzes court decisions

3. Employees may use their annual FMLA leave entitlement for a variety of reasons, including caring for a newly born or adopted child, 29 U.S.C. § 2601(b)(2) (1994), caring for a family member with a serious health condition, and caring for the employee’s own serious health condition, id. § 2601(a)(1). Employees most often use FMLA leave because of their own serious health condition. COMMISSION ON FAMILY AND MEDICAL LEAVE, A WORKABLE BALANCE: REPORT TO CONGRESS ON FAMILY AND MEDICAL LEAVE POLICIES 94 (1996) [hereinafter A WORKABLE BALANCE]. Because the issue of involuntary FMLA leave typically arises in situations involving an employee’s own serious health condition, the other reasons for FMLA leave are beyond the scope of this Comment.

5. 29 C.F.R. § 825.220(d) (2000).
addressing involuntary leave, including cases both in which the employee had an FMLA-qualifying serious health condition and in which the employee was able to perform essential job functions and thus did not have a qualifying serious health condition. Applying principles of statutory interpretation and analyzing the Department of Labor’s (DOL) regulations and advisory opinions, Part III critically evaluates the reasoning used in existing decisions on involuntary FMLA leave. It also considers the policy ramifications of permitting employers to put employees on involuntary FMLA leave. This Comment concludes that the statute, its accompanying regulations, and DOL advisory opinions prohibit employers from deducting leave from an employee’s annual FMLA leave allotment when the employee does not have an FMLA-qualifying serious health condition. In light of these provisions and the statute’s overall purpose of providing greater job security to employees with serious health conditions, employers who place non-qualifying employees who are willing and able to work on FMLA leave interfere with employees’ FMLA rights in violation of the statute.

I. THE FAMILY AND MEDICAL LEAVE ACT

President Clinton signed the Family and Medical Leave Act (FMLA) into law on February 5, 1993, making it the first major piece of legislation of his administration. The text of the statute does not directly address the issue of involuntary FMLA leave, whereby an employer places a non-qualifying employee on FMLA leave. However, it does describe the process by which employees may avail themselves of their FMLA rights, as well as the penalties for employer interference with these rights. The regulations issued by the DOL also fail to address involuntary FMLA leave explicitly, but do provide relevant guidance regarding designation of leave and intermittent leave. The DOL has issued a number of advisory opinions interpreting these regulations in light of specific employment situations, thereby illustrating their intended application. The legislative history of the FMLA describes the evolution of the statute and offers insight into its purpose, revealing

8. See President’s Statement on Signing the Family and Medical Leave Act of 1993, 29 WEEKLY COMP. PRES. DOC. 144 (Feb. 5, 1993).
Congress’s intent to instill rights in employees and safeguard those rights.

A. Provisions of the FMLA

The FMLA seeks to balance the demands of the workplace with the needs of families and to entitle employees to take reasonable leave for medical reasons or to care for a new baby or ill family member. The FMLA contains detailed provisions describing employee leave entitlements and permissive employer rights. The statute also prohibits certain acts by employers and provides extensive remedies to employees seeking enforcement of their FMLA rights. Courts have interpreted this provision to prohibit employers from forcing employees to forfeit other employment benefits in exchange for FMLA leave and from interfering with employee scheduling of FMLA leave. Finally, the statute’s construction provisions indicate that courts should liberally construe the FMLA in favor of employees.

1. Mandatory Employee Rights and Permissive Employer Rights Under the FMLA

In passing the FMLA, Congress determined there was inadequate job security for employees whose serious health conditions temporarily prevented them from working. Accordingly, the FMLA creates mandatory leave rights for employees and grants limited permissive rights to employers to prevent fraud and protect the employer’s economic viability.

11. Id. § 2601(b)(2).
12. Id. § 2601(a)(4). Congress had earlier extended job opportunities to persons with disabling health conditions under the Americans with Disabilities Act of 1990 (ADA), Pub. L. No. 101-336, 104 Stat. 330 (codified at 42 U.S.C. §§ 12101–12213 (1994 & Supp. IV 1998)). The ADA defines “disability” as “a physical or mental impairment that substantially limits one or more . . . major life activities.” 42 U.S.C. § 12102(2) (1994 & Supp. IV 1998); see also 29 C.F.R. § 1630.2 (2000). Under the ADA, employers may not discriminate against qualified individuals with disabilities, and employers have an affirmative duty to provide reasonable accommodation to such individuals, unless the employer can demonstrate that to do so would cause the employer undue hardship. 42 U.S.C. §§ 12112(a), (b)(5)(A).
13. The FMLA uses the term “shall” with regard to employees’ rights. 29 U.S.C. § 2612(a)(1) (stating that “eligible employee shall be entitled” to twelve weeks of leave); id. § 2614(a)(1) (stating that “eligible employee . . . shall be entitled” to be restored the same or an equivalent position on return from leave). In contrast, the statute uses the term “may” with regard to employer rights. Id.
Involuntary Leave Under the FMLA

The FMLA entitles eligible employees who work for covered employers to take up to twelve weeks of unpaid leave annually for specific qualifying reasons. While the FMLA guarantees twelve weeks of unpaid leave, an employee may elect or an employer may require the employee to substitute paid leave for unpaid leave. The reasons for taking FMLA leave include the birth or adoption of a child; caring for a seriously ill child, parent, or spouse; and attending to the employee’s own serious health condition. The statute defines “serious health condition” as an illness, injury, impairment, or physical or mental condition that involves inpatient care or continuing treatment by a health care provider. The statute also requires that the serious health condition make the employee unable to perform the functions of his or her position. Where the leave is for the employee’s own serious health condition, the employee is alternatively permitted to take leave on an intermittent or reduced schedule basis. An employee is eligible for leave when he or she has been employed by a covered employer for at least twelve months and has worked at least 1250 hours for that employer during the previous twelve months. Covered employers include private employers who employ at least fifty employees for each working day during twenty or more calendar weeks of the current or preceding year, as well as federal, state, and local governments. Employees returning

§ 2613(a)(1) (stating that “employer may require” certification from employee’s health care provider); id. § 2614(b)(1) (stating that “employer may deny restoration” of highly compensated employee to prevent substantial and grievous economic injury to employer’s operations). Use of the term “may” is usually precatory, whereas use of the term “shall” is mandatory. See, e.g., Mallard v. U.S. Dist. Court, 490 U.S. 296, 301–02 (1989).

14. 29 U.S.C. § 2612(d). Various forms of employer-provided paid leave may be substituted, including vacation, personal, family, and sick leave. 29 C.F.R. § 825.207. The employer may designate the substituted paid leave as FMLA leave and thus count it against the employee’s annual FMLA allotment pursuant to the designation procedures in 29 C.F.R. § 825.208.


16. Id. § 2611(11); see also 29 C.F.R. § 825.114(a).

17. 29 U.S.C. § 2612(a)(1)(D). This is the requirement that is usually at issue in involuntary-leave cases. Thus, the use of the term “non-qualifying employee” in this Comment refers to an employee who is, in fact, able to perform the essential functions of his or her position.

18. Id. § 2612(b)(1).

19. Id. § 2611(2)(A).

20. Id. § 2611(4)(A). Limiting coverage to employers with fifty or more employees made the statute, at the time of its enactment, applicable to approximately ten percent of private sector employers, who employed nearly sixty percent of the private sector workforce. A WORKABLE BALANCE, supra note 3, at 58–60. Some members of Congress have sought to expand coverage to include employers with twenty-five or more employees, which would cover thirteen million more employees, an additional fourteen percent of the workforce. See, e.g., S. 183, 105th Cong. (1997).
from leave are entitled to return to the same position they held before taking leave or an equivalent position. In addition, during any FMLA leave period, eligible employees are entitled to maintain their benefits under the employer’s group health plan.

The FMLA grants employers a limited set of permissive rights that allow an employer to verify the employee’s need for leave and preserve its economic viability. To prevent fraud or abuse, an employer may require employees seeking FMLA leave for their own serious health condition to provide medical certification from a health care provider, periodic status reports, and certification of the employee’s ability to return to work following the leave period. Employers may also preserve the economic viability of their businesses by denying restoration of certain highly compensated employees following a period of FMLA leave if such denial is necessary to prevent substantial and grievous economic injury to the employer’s operations. Additionally, the employer may recover premiums paid for group health plan coverage for an employee on FMLA leave if the employee fails to return to work after the leave period has expired. However, in contrast to the mandatory rights granted to employees, the language of the statute expressly indicates that the rights granted to employers are permissive.

21. 29 U.S.C. § 2614(a)(1). An “equivalent position” is one that has equivalent employment benefits, pay, and other terms and conditions of employment. Id. § 2614(a)(1)(B).

22. Id. § 2614(c)(1). Employers and employees both maintain their existing contributions to the payment of group health care premiums during the employee’s leave. See 29 C.F.R. § 825.209 (2000).


24. Id. § 2614(a)(5).

25. Id. § 2614(a)(4). The employer may only impose this requirement if it is uniformly applied. Id.

26. Id. § 2614(b). A “highly compensated employee” is a salaried employee who is among the highest-paid ten percent of the employees employed within seventy-five miles of the facility where the employee is employed. Id.

27. Id. § 2614(c)(2). However, if the employee fails to return to work because of the continuation, recurrence, or onset of a serious health condition, or because of circumstances beyond the employee’s control, the employer may not recover premiums paid to maintain the employee’s group health coverage during the period of FMLA leave. Id.

28. See, e.g., id. § 2612(b)(2) (“Employer may require such employee to transfer temporarily . . . .”); id. § 2613(a) (“Employer may require that a request for leave . . . be supported by certification . . . .”); see also supra note 13.
2. **Prohibited Employer Acts and Liability for Violations of Employees’ FMLA Rights**

In addition to describing what employers must do under the FMLA, the statute also describes what they cannot do. The FMLA prohibits interference with, restraint of, and denial of employees’ FMLA rights. Courts have held that employers who delay granting an employee’s FMLA leave request or force an employee to forfeit other employment benefits in exchange for FMLA leave violate the employee’s FMLA rights. The statute provides for both agency and private enforcement of this provision. Employers are strictly liable for violations of employees’ FMLA rights, and violations entitle employees to an array of remedies, including both monetary damages and equitable relief.

a. **The FMLA’s Prohibition of Interference with, Restraint, or Denial of Employee FMLA Rights**

Section 2615(a)(1) of the FMLA expressly prohibits employers from interfering with, restraining, or denying the exercise of or attempt to exercise any rights granted to employees under the FMLA. 29 The statute also contains a retaliation provision, which prohibits employers from discharging or otherwise discriminating against any individual for opposing any practice made unlawful by the FMLA. 30 In addition to providing for enforcement by the Secretary of Labor, 31 the FMLA grants employees a private right of action to enforce their FMLA rights 32 and specifically contemplates class action suits for groups of similarly situated employees. 33 Courts have concluded that employers guilty of interfering with, restraining, or denying employees’ FMLA rights are strictly liable for any such violations. 34 Under this standard, an employee

---


32. Id. § 2617(a)(2)(A).

33. Id. § 2617(a)(2)(B).

need not prove anything regarding the employer’s intent in violating the statute; the employee need only prove by a preponderance of the evidence that the employer deprived him or her of a benefit to which he or she was entitled under the statute.

In construing § 2615(a)(1), courts have found that certain acts not expressly prohibited by the statute nonetheless constitute violations of employee rights. For example, in *Mardis v. Central National Bank & Trust of Enid*, the Tenth Circuit held that forcing an employee to forfeit other employment benefits, such as accrued paid sick leave and vacation time, in order to take FMLA leave to care for her seriously ill husband interfered with the employee’s FMLA rights. Also, in *Sherry v. Protection, Inc.*, the district court found that delaying a decision on an employee’s requests for FMLA leave constituted denial of the employee’s FMLA rights, even where the employer ultimately granted the leave request. Similarly, the court in *Williams v. Shenango* held that where the employee requested a particular week off to care for his seriously ill spouse, the employer’s denial of the request interfered with the employee’s FMLA rights, despite the fact that the employer granted the request for leave during a different week. These cases all suggest that an employee has the unilateral right to decide when he or she needs FMLA leave and an employer cannot compel the surrender of other employment benefits.

### b. Employer Liability for Violations of Employees’ FMLA Rights

Employers who violate employees’ FMLA rights face serious penalties, including liability for the employee’s wages and out-of-pocket expenses, double damages, equitable relief, and attorney’s fees. The

---

36. *Id.*
38. *Id.* at *2; see also Goodwin-Haulmark v. Menninger Clinic, Inc.*, 76 F. Supp. 2d 1235, 1242 (D. Kan. 1999).
40. *Id.* at 1136.
42. *Id.* at 320–21.
Involuntary Leave Under the FMLA

statute explicitly provides for multiple legal remedies: lost wages, salary, employment benefits, or other compensation. The employer also may be liable for other economic losses sustained by the employee, such as out-of-pocket medical expenses. Furthermore, employers must pay interest on money damages for both lost compensation and out-of-pocket expenses. Congress further emphasized the gravity of employer violations of employee FMLA rights by creating a presumption that employers who violate such rights are liable to the employee for an additional amount of liquidated damages, equal to the sum of the two types of damages described above, plus interest. To overcome this presumptive liability for double damages, the employer must prove that it acted in good faith and with reasonable grounds for believing it was not in violation of the FMLA. Employees may also seek equitable relief for violations of their FMLA rights, which may include employment, reinstatement, or promotion. Finally, employees who prevail in FMLA enforcement actions receive reasonable attorney’s fees, reasonable expert-witness fees, and other costs. This award of fees and costs to a prevailing plaintiff is mandatory, even where an employee receives only nominal damages.

3. The FMLA’s Liberal Construction Provisions

The FMLA’s construction provisions demonstrate Congress’s intent to provide for and protect employees. The FMLA sets a minimum standard, a floor beneath which employee leave benefits may not drop. The

44. Id. § 2617(a)(1)(A)(i)(II).
45. Id. § 2617(a)(1)(A)(ii).
46. Id. § 2617(a)(1)(A)(iii).
47. Id.
48. Id. § 2617(a)(1)(B).
49. Id.
50. Id. § 2617(a)(3).
52. See McDonnell v. Miller Oil Co., 968 F. Supp. 288, 293 (E.D. Va. 1997), aff’d, 110 F.3d 60 (4th Cir. 1997), remanded, 134 F.3d 638 (4th Cir. 1998). The Fourth Circuit upheld an award of $19,698.81 in attorneys’ fees to an employee to whom the jury awarded only $2.00 in damages. Id. at 295.
FMLA states that nothing in the Act shall be construed to modify or affect any federal or state law prohibiting discrimination. The FMLA also specifies that it shall not be construed to supersede any state or local law or collective bargaining agreement or other benefit program that provides greater family or medical leave rights to employees. Finally, the FMLA is not meant to discourage employers from adopting or retaining leave policies that are more generous than those contained in the statute.

B. Department of Labor Regulations and Advisory Opinions Relevant to Involuntary FMLA Leave

Congress vested the Secretary of Labor with authority to administer the FMLA. While the DOL’s regulations do not specifically address involuntary leave, numerous regulatory provisions guarantee employee FMLA rights while allowing certain protections for employers. Regulations regarding the designation of leave and the amount of leave to be charged in cases of intermittent leave shed light on the issue of involuntary leave. In addition, the DOL has issued advisory opinions to employers seeking guidance on how to interpret the FMLA in particular circumstances. One such advisory opinion affirmed that employers may designate FMLA leave as such when the employer and employee have satisfied all of the statutory requirements for leave. Two additional opinions held that when an employee is on intermittent FMLA leave, employers may not deduct more leave from the employee’s annual FMLA allotment than is necessary to meet the employee’s specific qualifying need.

I. FMLA Regulations and Department of Labor Advisory Opinions Regarding Designation of Leave as FMLA Leave

While the DOL’s FMLA regulations do not expressly address involuntary leave, the regulations do describe the process by which an employer may charge leave against an employee’s annual FMLA leave.

---

54. 29 U.S.C. § 2651(a).
55. Id. § 2651(b).
56. Id. § 2652(b).
57. Id. § 2653.
58. Id. § 2654.
Involuntary Leave Under the FMLA

The FMLA regulations extensively address the employer’s designation of leave as FMLA leave, which occurs in response to an employee’s request for leave or upon receipt of information from the employee or his or her spokesperson indicating a need and desire to take FMLA leave.\(^{59}\) According to the regulations, “(i)n all circumstances, it is the employer’s responsibility to designate leave, paid or unpaid, as FMLA-qualifying, and to give notice of the designation to the employee.”\(^{60}\) The regulations explain that the employer must base this designation solely on information received from the employee or the employee’s spokesperson.\(^{61}\) Thus, while the regulations give the employer the responsibility for designating leave as FMLA leave, this responsibility only arises upon receipt of information from the employee or his or her spokesperson indicating a need and desire to take leave.\(^{62}\) A sample form entitled “Employer Response to Employee Request for Family and Medical Leave” further evidences this procedure.\(^{63}\)

In a 1995 advisory opinion,\(^{64}\) the DOL addressed the specific issue of whether an employer may designate leave as FMLA leave if the employee has not requested that it be designated as such. The opinion stated that if the employer meets the definition of a covered employer, the employee meets the definition of an eligible employee, and the reason for the leave meets the definition of a serious health condition, then the employer may designate the leave as FMLA leave and count it against the employee’s twelve-week entitlement, even if the employee has not requested that it be counted as such.\(^{65}\) This opinion affirmed that

---

59. 29 C.F.R. § 825.208 (2000). Designation is the process by which the employer determines whether the leave is for an FMLA-qualifying reason and thereby counts the leave against the employee’s annual twelve-week leave entitlement. Id.
60. Id. § 825.208(a).
61. Id.
62. See, e.g., id. § 825.208(a)(1) (Noting that employee must give notice of need for unpaid FMLA leave to employer); id. (“[I]n explaining the reasons for a request to use paid leave . . . an employee will provide sufficient information for the employer to designate the paid leave as FMLA leave.”); see also id. § 825.208(a)(2) (referring to employee’s request or notification to employer of intent to use accrued paid leave); id. (“[A]n employee . . . who seeks an extension of unpaid leave for an FMLA-qualifying purpose will need to state the reason.”).
63. Id. § 825 app. D, Form WH-381 (2000).
65. Id. A prior DOL advisory opinion indicated that an employer’s decision to make FMLA leave mandatory for employees taking leave for a qualifying reason was permissible under the statute but not required. 99 Wage & Hour Manual (BNA) 3044, Op. FMLA-49 (Oct. 27, 1994). A subsequent opinion confirmed that it is the employer’s prerogative to designate leave as FMLA leave, and that an employee may not bar the employer from designating any qualifying absence as FMLA leave. 99
the employer may designate leave as FMLA leave and count it against an employee’s annual allotment only where the employer and employee satisfy all of the statutory prerequisites.

2. Regulations and DOL Advisory Opinions Regarding the Calculation of Intermittent FMLA Leave

When an employee takes leave on an intermittent or reduced schedule basis, the regulations specify that only the amount of leave actually taken counts toward the annual twelve-week leave entitlement. Thus, if an employee needs only two hours off for an FMLA-qualifying medical appointment, an employer may only deduct those two hours from the employee’s FMLA leave balance. This is true even where the nature of the employee’s job prevents him or her from returning to work for the remainder of the workday. For example, if a flight attendant has a qualifying two-hour medical appointment in the morning that results in missing the flight on which he or she was to work that day, the employee can nonetheless only be charged with two hours of FMLA leave. The fundamental principle underlying this regulation is that employers may not require employees to take more FMLA leave than necessary to address the circumstance that precipitated the need for leave.

---

Wage & Hour Manual (BNA) 3086, Op. FMLA-83 (Aug. 7, 1996). The term “involuntary leave” has been used to describe this type of situation, where the employee qualifies for FMLA leave but has not requested such leave or has expressed a desire to have qualifying leave not count against the annual FMLA leave entitlement. However, the regulations and advisory opinions authorize the employer to designate FMLA-qualifying leave as such, regardless of the employee’s request or consent. This is therefore a designation issue and is not what is meant by the term “involuntary leave” as it is used in this Comment.

66. The statute allows for this type of leave in cases where it is medically necessary because of the employee’s own serious health condition or that of a family member. 29 U.S.C. § 2612(b)(1) (1994 & Supp. IV 1998). Both the employee and employer must agree to such leave. Id. Intermittent leave is leave taken in separate blocks of time for a single qualifying reason; a reduced leave schedule reduces the employee’s usual number of working hours per day or per week. 29 C.F.R. § 825.203.

67. 29 C.F.R. § 825.205(a).


69. Id.

70. 29 C.F.R. §§ 825.203(d), 204(e).
Involuntary Leave Under the FMLA

DOL advisory opinions clarify that it is the employee’s need, not the employer’s dictates, that determines the terms of FMLA leave. In August 1994, the Wage and Hour Division of the DOL issued an advisory opinion regarding intermittent leave in response to a series of inquiries from an employer. One of the questions the advisory opinion addressed involved a flight attendant who requested intermittent FMLA leave to care for a seriously ill parent. The employee needed three hours off every other Friday for two months. The three-hour absence caused the employee to miss her flight assignment for that day. The employer wished to know how much leave to charge against the employee’s twelve-week entitlement for these absences. In response, the DOL stated that the employee should be charged only three hours of FMLA leave per absence. The opinion cited regulations providing that an employer may only charge an employee taking intermittent leave with the amount of leave actually taken for the FMLA-qualifying reason.

The DOL issued a second advisory opinion on this topic in July 1997. In that case, the DOL examined intermittent FMLA leave for salaried executive, administrative, and professional employees who are exempt under the Fair Labor Standards Act (FLSA). The agency again concluded that requiring exempt employees to take a full day of leave in circumstances where the employee does not need a full day to attend to an FMLA-qualifying need violates FMLA regulations. These two opinions affirm that employers cannot require employees to take FMLA leave beyond what is necessary to meet the employee’s FMLA-qualifying need.

72. Id.
73. Id.
74. Id.
75. Id.
76. Id.
77. Id.
78. Id. (citing 29 C.F.R. § 825.205(a) (1994)).
81. 99 Wage & Hour Manual (BNA) 3097, Op. FMLA-89 (July 3, 1997) (citing 29 C.F.R. § 825.203(d) (1997)). In this situation, requiring exempt employees to use more FMLA leave than necessary would also violate FLSA regulations. Id. (citing 29 C.F.R. § 541.118(a)(1) (1997)).
C. The Legislative History of the FMLA

In response to a “demographic revolution”82 in the composition of the American workforce and the increasing difficulty employees faced in trying to balance work and family responsibilities,83 Congress introduced the first family leave legislation in 1985.84 Increasing numbers of women in the workforce, greater numbers of single-parent households, and the overall aging of the American population contributed to the tension employees felt between their jobs and personal obligations.85 Also, because other developed countries provide extensive family and medical leave benefits to employees,86 concerns about maintaining worker productivity and global competitiveness were instrumental in the FMLA’s development.87 After many years of revisions and compromises, Congress enacted the FMLA in 1993, guaranteeing employees a minimum amount of annual leave to balance the demands of their personal and professional lives.88

The legislative history of the FMLA affirms a consistent purpose: to provide both beneficial rights and greater job security for employees with serious health conditions.89 The FMLA’s legislative history contains numerous examples of ways in which Congress sought to protect employee rights. For example, Congress established a “stringent standard”90 requiring that employees returning from leave be returned to the same position or an equivalent position, not merely a comparable or similar position.91 Although the FMLA provides an exception to this restoration provision for certain highly compensated employees,92 Congress nonetheless mandated that such employees be provided with

84. Zuckman, supra note 9, at 267.
86. Id. at 19–20, reprinted in 1993 U.S.C.C.A.N. 3, 21–22. Congress noted that among the major industrialized countries, the average minimum amount of paid leave for medical reasons was twelve to fourteen weeks. Id. Many countries also provide at least one year of unpaid medical leave. Id.
88. 29 U.S.C. § 2601(b).
91. Id.
92. 29 U.S.C. § 2614(b).
continued group health benefits for the duration of the leave period, even after the employer notifies them that it will deny restoration. Furthermore, Congress modeled the FMLA’s remedies and enforcement mechanisms after those of the FLSA. Because of this explicit reference and the similarities between the FLSA and FMLA, courts have looked to the FLSA and its interpretive case law for guidance in interpreting the FMLA. Courts have overwhelmingly held that the FLSA warrants liberal construction in favor of employees.

Some of the legislators who opposed the FMLA did so on grounds that it was too favorable to employees. The legislative history contains references to comments by legislators noting that the Act “allows employees almost unrestrained discretion as to when to take leave” and “grants the employee a unilateral right to schedule and take the leave.” Yet, during the eight years in which Congress crafted the FMLA, it never explicitly considered the issue of involuntary leave, whereby an employer places an employee on FMLA leave without the employee requesting leave. Neither the committee reports submitted in support of the FMLA nor the floor debates contain any reference to involuntary leave.

95. See, e.g., Frizzell v. S.W. Motor Freight, 154 F.3d 641, 643–44 (6th Cir. 1998) (holding that while FMLA does not explicitly grant employees right to jury trial, legislative history of FMLA indicates its enforcement provisions mirror FLSA; courts have interpreted FLSA to provide employees with right to jury trial); Freemon v. Foley, 911 F. Supp. 326, 330–32 (N.D. Ill. 1995) (holding that identical definitions of “employer” in FMLA and FLSA and similarities between statutes require finding individual liability for supervisors under FMLA as under FLSA).
96. See, e.g., Tenn. Coal, Iron & R.R. Co. v. Muscoda, 321 U.S. 590, 597 (1944) (holding that FLSA is remedial and must be given liberal construction in accordance with its obvious intent and purpose); Wirtz v. Ti Ti Peat Humus Co., 373 F.2d 209, 212 (4th Cir. 1967), cert. denied, 389 U.S. 834 (1967) (citation omitted) (holding courts should liberally construe remedial social legislation such as FLSA to fulfill its purpose of protecting workers).
98. H.R. REP. NO. 103-8, pt. 1 at 70.
II. JUDICIAL OPINIONS ADDRESSING INVOLUNTARY FMLA LEAVE

In the eight years since President Clinton signed the FMLA, courts have decided many cases involving employees’ FMLA rights, but few of these decisions have examined the issue of involuntary leave. The few involuntary-leave cases decided to date have involved both employees with and without FMLA-qualifying serious health conditions, a definition that hinges upon whether the employee is able or unable to perform essential job functions. Most courts have failed to make this distinction and have therefore held that employers may place even non-qualifying employees such as Mr. Keating on FMLA leave reasoning that no FMLA provision expressly prohibits such a practice.

A. Cases in Which Employees Were Unable To Perform Essential Job Functions

Two cases purporting to deal with involuntary FMLA leave involved employees whose serious health conditions prevented them from performing an essential function of their respective positions and thus qualified them for FMLA leave. Because these were qualifying employees, the real issue in these cases was one of designation.102

In *Moss v. Formosa Plastics Corp.*,103 an employer placed an epileptic employee on disability leave after the employee suffered epileptic seizures at work.104 After receiving conflicting reports regarding the effect of the employee’s medical condition on his ability to perform essential job functions, the employer decided to terminate the employee because of his medical problems.105 Instead of firing the employee, however, the employer placed the employee on FMLA leave to allow him to “get his situation under control.”106 The employer ultimately terminated the employee, and the employee sued, claiming the employer violated his FMLA rights by unilaterally forcing him to take twelve weeks of FMLA leave.107 The court found the employee was unable to

---

102. *See supra* note 65 and accompanying text.
104. *Id.* at 738–39.
105. *Id.* at 738.
106. *Id.*
107. *Id.* at 740.
Involuntary Leave Under the FMLA

perform the essential functions of his position as a control-panel operator because he could not respond to emergency situations if rendered unconscious by a seizure.\textsuperscript{108} The court held that nothing in the FMLA prohibits employers from requiring employees to take unpaid leave.\textsuperscript{109}

In \textit{Harvender v. Norton},\textsuperscript{110} a pregnant employee’s physician advised her not to work with or be exposed to chemicals during her pregnancy.\textsuperscript{111} When budget constraints prevented the employer from placing the employee on light duty, the employer instead placed her on FMLA leave.\textsuperscript{112} The employee had not requested leave; in fact, she objected to it.\textsuperscript{113} She brought suit claiming the employer knowingly and intentionally violated the FMLA by committing a prohibited act under § 2615(a)(1),\textsuperscript{114} which prohibits employers from interfering with, restraining, or denying employee FMLA rights. The court reasoned that whether an employee requested FMLA leave is irrelevant because the statute does not specify that FMLA leave be granted only when the employee wishes it.\textsuperscript{115} Instead, the court held if the required conditions for leave are met, namely that the employee is eligible and has an FMLA-qualifying reason for leave, the employer is required to provide the employee with leave and may designate it as FMLA leave.\textsuperscript{116}

\textsuperscript{108} \textit{Id.} at 742–43.
\textsuperscript{109} \textit{Id.} at 741 (citing \textit{Harvender v. Norton}, No. 96-CV-653, 1997 WL 793085, at *7 (N.D.N.Y. Dec. 15, 1997)). Under the employment at-will doctrine, employers may place employees on unpaid leave, just as they may terminate them, for any reason or for no reason. \textit{See}, e.g., \textit{Smoot v. Boise Cascade}, 942 F.2d 1408, 1410 (9th Cir. 1991) (stating that, under Washington law, employer generally may terminate employment contract of indefinite duration at any time, with or without cause). \textit{Id.} However, there is a distinction between simply placing an employee on unpaid leave and placing an employee on unpaid FMLA leave in that the latter deprives the employee of a statutory right to a guaranteed amount of annual leave. Congress demonstrated its intent to grant employees this unique right and protect it by its inclusion of § 2615, prohibiting employer interference with, restraint of, and denial of the exercise of these rights. 29 U.S.C. § 2615(a)(1) (1994).
\textsuperscript{110} No. 96-CV-653, 1997 WL 793085 (N.D.N.Y. Dec. 15, 1997).
\textsuperscript{111} \textit{Id.} at *1.
\textsuperscript{112} \textit{Id.}
\textsuperscript{113} \textit{Id.}
\textsuperscript{114} \textit{Id.}
\textsuperscript{115} \textit{Id.} at *7.
\textsuperscript{116} \textit{Id.} at *7–8. Under the FMLA regulations and DOL advisory opinions regarding designation of leave, such designation would not be mandatory. 99 \textit{Wage & Hour Manual} (BNA) 3044, Op. FMLA-49 (Oct. 27, 1994). Thus, an employer would be permitted, but not required, to count the leave against the employee’s annual twelve-week entitlement.
B. Involuntary-Leave Cases in Which Employees Were Willing and Able To Perform Essential Job Functions

To date, only one circuit court has examined a case involving involuntary FMLA leave where an employer placed a non-qualifying employee on FMLA leave. Yet, the Sixth Circuit declined to decide the involuntary-leave issue and rendered an unpublished decision. In Hicks v. Leroy’s Jewelers, a pregnant employee requested FMLA leave to begin immediately after the birth of her child. Prior to the birth, the employee spent one night in a hospital because of a kidney infection. The employer designated the employee as being on FMLA leave as of the date of her hospitalization for the kidney infection, despite the fact the employee indicated she was willing and able to work for the final month remaining prior to her delivery date. Her twelve-week leave thus expired one month earlier than it would have as originally scheduled. After the expiration of the leave period, as calculated from the earlier date of hospitalization, the employer terminated the employee when she did not return to work.

The employee claimed that by placing her on involuntary leave at the time of hospitalization the employer interfered with her FMLA rights and thus violated § 2615(a)(1). The employee maintained that her hospitalization for a kidney infection did not meet the regulatory definition of a serious health condition; thus, because she was not eligible for FMLA leave, she argued that her employer could not place her on FMLA leave. The trial court found that the employee’s initial hospitalization was an FMLA-qualifying absence that made her eligible for FMLA leave. The trial court further concluded that because the employee was

---

118. Id. at *4.
120. Id. at *1.
121. Id.
122. Id.
123. Id.
124. Id. at *2. The employee also claimed the employer violated the FMLA by failing to restore her to her position. Id.
125. Id. at *3.
126. Id.
Involuntary Leave Under the FMLA

eligible for FMLA leave, the employer had authority to place her on involuntary FMLA leave. On appeal, the Sixth Circuit held that the employee’s stated willingness and ability to work created an issue of material fact regarding her eligibility for FMLA leave; therefore the trial court had erred in granting summary judgment to the employer on that ground. However, the court declined to remand on this issue because it was able to resolve the case on unrelated grounds.

A federal district court in Texas was the first to hear a case on the issue of involuntary FMLA leave. In Love v. City of Dallas, the employer placed two employees who suffered from carpal tunnel syndrome on light duty to conform to restrictions necessitated by their medical condition. These employees were willing and arguably able to perform the essential functions of their positions. The employer subsequently issued a new policy stating that it would not allow employees to remain on limited duty indefinitely. When the employer determined that the two employees were unable to return to regular duty, it placed them on involuntary FMLA leave for a total of twenty-four weeks over two years.

The employees argued that the employer violated their FMLA rights by placing them on involuntary FMLA leave because, without a qualifying serious health condition, they were not entitled to FMLA leave. The employer argued that the employees failed to state a claim under the FMLA because nothing in the statute prohibits the practice of placing employees on involuntary FMLA leave. The court dismissed the employees’ assertion that nothing in the FMLA or its regulations

127. Id.
128. Id. at *4. The court held that the employee’s failure to return to work after the expiration of her twelve-week leave period, when calculated either from the original date her leave was to begin or from the date of her hospitalization, precluded any recovery under the FMLA. Id.
130. Id. at *1.
131. The employees’ ability to perform raises issues of reasonable accommodation under the ADA. See supra note 12. However, these issues are beyond the scope of this Comment.
133. Id. at *2.
134. Id. at *6. The court had “difficulty” with the employees arguing in the same motion that they were disabled under the ADA but did not have a serious health condition under the FMLA. Id. The ADA’s definition of disability is markedly different from the FMLA’s definition of serious health condition; thus there is no inconsistency in such an argument. See supra note 12.
allows employers to put employees on involuntary FMLA leave. As in Moss, the court instead found it more relevant that “nothing in the statute prohibits an employer from doing so.”

Burton v. Neumann also involved an employer placing an employee on involuntary FMLA leave when the employee was willing and able to perform job functions. The employee in Burton suffered from iritis. Iritis is a permanent condition affecting the eyes and characterized by recurring periods of inflammation interspersed with periods of months or years without symptoms. The employee’s doctor determined that work-related stress contributed to the onset of symptomatic episodes, and the employee sought to discuss with her employer reasonable accommodations that might reduce her stress level at work. In response, her employer “summarily” placed her on involuntary FMLA leave. The employee claimed involuntary leave was unlawful under the FMLA. She also claimed the employer violated the FMLA by failing to restore her to her position following placement on involuntary FMLA leave. First, citing Love, the court held that the FMLA does not prohibit involuntary leave, and thus, the employee had no basis for an involuntary-leave claim. Second, although the employee’s physician provided a letter stating there was no medical reason why she could not continue working, with or without reasonable accommodation, the employee failed to submit this required letter in a timely fashion. Accordingly, she lost her FMLA claim for failure to restore her to her position.
III. EMPLOYERS WHO PLACE EMPLOYEES WHO CAN PERFORM THEIR ESSENTIAL JOB FUNCTIONS ON INVOLUNTARY FMLA LEAVE INTERFERE WITH EMPLOYEE FMLA RIGHTS

Courts deciding involuntary FMLA leave cases have wrongly held that the FMLA permits employers to place employees who are able to perform essential job functions on involuntary FMLA leave. The central issue courts must examine in involuntary-leave cases is whether the employee actually meets the statutory prerequisites for leave, particularly the requirement that the employee be unable to perform an essential job function of his or her position. Instead, most courts have focused on the fact that the text of the statute and its regulations are silent on the issue of involuntary leave. They have interpreted this silence to allow employers to force non-qualifying employees to take involuntary FMLA leave.

However, courts must resolve the ambiguity created by legislative silence by examining the language of the statute, the statutory scheme as a whole, the administrative regulations and interpretive guidance, and its legislative history. Such an analysis reveals not only Congress’s intent to provide employees greater job security by guaranteeing them the right to a minimum amount of family and medical leave, but also that allowing employers to place employees who are willing and able to perform their essential job functions on involuntary FMLA leave unlawfully interferes with the employees’ FMLA rights. If the requirements for FMLA leave are met, the issue becomes one of designation, and the employer may choose to designate FMLA-qualifying leave as such and count it against an employee’s annual entitlement. However, if the prerequisites for leave are not met, an employer may neither grant nor impose FMLA leave.

148. The *Hicks* court alluded to this distinction and noted that it is an issue of material fact in determining whether involuntary FMLA leave is permissible. *See* Hicks v. Leroy’s Jewelers, Inc., 225 F.3d 659, No. 98-6596, 2000 WL 1033029, at *3 (6th Cir. July 17, 2000), *cert. denied*, 121 S. Ct. 1084 (Feb. 20, 2001) (No. 00-891).

149. In borderline cases where it is unclear whether the statutory requirements for leave have been met, there may be a reasonable accommodation issue for employees who qualify under the ADA. *See supra* note 12. Employees who do not meet the ADA’s stringent definition of disability may be left unprotected by either the ADA or the FMLA. In light of the statute’s stated purpose of giving employees with serious health conditions greater job security, courts must not permit employers to use the FMLA as a weapon against employees in these types of borderline cases to precipitate termination of their employment.
A. Employers May Not Place Non-Qualifying Employees on Involuntary FMLA Leave

The FMLA prohibits employers from placing employees who are capable of performing their essential job functions on involuntary FMLA leave. Case law interpreting § 2615(a)(1) suggests that placing non-qualifying employees on involuntary FMLA leave interferes with employees’ FMLA rights. The statutory scheme as a whole reveals Congress’s strong intent to protect employee rights; because involuntary FMLA leave interferes with employee FMLA rights, it compromises the structural integrity of the Act. Finally, allowing involuntary leave contradicts the command of the statute’s construction provisions that the Act be broadly construed in favor of protecting employee leave rights.

1. Involuntary Leave Violates § 2615(a)(1) As Construed by Case Law

An employer who places an employee on involuntary FMLA leave interferes with the employee’s FMLA rights and thus violates § 2615(a)(1). In Mardis v. Central National Bank & Trust of Enid, the Tenth Circuit construed interference with FMLA rights to prohibit employers from requiring employees to forfeit employment benefits in exchange for being granted FMLA leave.150 In Williams v. Shenango151 and Sherry v. Protection, Inc.,152 the district courts also concluded that this provision of the statute prohibits employers from interfering with an employee’s scheduling of FMLA leave.153 These courts held that delaying or postponing the employee’s FMLA leave request constituted interference, even where the employee was ultimately granted FMLA leave.154

Placing an employee on involuntary FMLA leave in essence requires the employee to forfeit his or her right to use that FMLA leave at a later date. For example, when Mr. Keating’s employer placed him on twelve

153. Williams, 986 F. Supp. at 320–21; Sherry, 981 F. Supp. at 1136; see also supra notes 39–42 and accompanying text.
154. Williams, 986 F. Supp. at 320–21; Sherry, 981 F. Supp. at 1136; see also supra notes 39–42 and accompanying text.
Involuntary Leave Under the FMLA

weeks of involuntary FMLA leave, it precluded him from using that FMLA leave later in the year, were he or his wife to suffer from a serious medical condition.\textsuperscript{155} Alternatively, the practice of involuntary FMLA leave interferes with the employee’s ability to schedule his or her FMLA leave, which the statute also prohibits.\textsuperscript{156} For example, in Hicks, the employee was unable to schedule FMLA leave following the birth of her child because her employer had placed her on FMLA leave when she arguably did not have a qualifying serious health condition.

The Love and Burton courts failed to recognize this statutory violation. In both cases, the employees were arguably willing and able to perform the essential functions of their respective positions and thus did not qualify for FMLA leave.\textsuperscript{157} Because these courts failed to appreciate that these were non-qualifying employees, they also failed to recognize that by putting non-qualifying employees on FMLA leave, the employers in Love and Burton were interfering with the employees’ right to schedule and use their FMLA leave. Judicial interpretation of § 2615(a)(1) makes clear that the statute prohibits this type of interference; thus these employers violated their employees’ FMLA rights.

2. Involuntary Leave Compromises the Structural Integrity of the FMLA

Examination of the FMLA’s entire statutory and regulatory scheme reveals that allowing employers to place employees who are able to perform their essential job functions on involuntary leave violates § 2615(a)(1). This scheme is based on the employee’s initiation of the leave process, either by requesting leave in advance where the need is foreseeable, or by providing the employer with adequate notice of the need for leave where it is unforeseeable.\textsuperscript{158} The designation procedures, the recordkeeping procedures, and even the sample forms are all based on having the employer respond to the employee’s initial leave request or notification.\textsuperscript{159} No other manner of initiating the leave process appears in

\textsuperscript{155} See supra note 1 and accompanying text.
\textsuperscript{156} See supra notes 39–42 and accompanying text.
\textsuperscript{159} See 29 C.F.R. §§ 825.208, 500, app. D.
either the statute or the regulations. Allowing employers to place employees who are able to perform essential job functions on involuntary leave is an entirely different process that is outside the procedures described in the statute. Accordingly, Congress wanted employees, not employers, to initiate the leave process; thus, only employees may avail themselves of their FMLA rights when they deem it necessary.

In addition, the statutory scheme provides further evidence of Congress’s intent to instill leave rights in employees and protect those rights. For example, the FMLA’s strong enforcement provisions and generous remedies, including a presumption of double damages, shows Congress’s strong intent that employers not interfere with employees’ FMLA rights. Likewise, by modeling these enforcement provisions after the FLSA, a statute which protects employees’ most basic rights regarding wages and hours, Congress unmistakably signaled that it meant to guarantee employees FMLA rights of equal strength and provide strong incentives for employers to honor such rights. Finally, by making employee rights mandatory and employer rights merely permissive, Congress demonstrated that employee FMLA rights are stronger than employer FMLA rights and should thus control.

Courts that have examined involuntary FMLA leave have failed to review closely the entire statutory scheme. The Love and Burton courts looked only to whether the statute contained a provision regarding involuntary FMLA leave. Finding none, they concluded that involuntary leave was permissible. These courts failed to recognize that the statute’s silence on this issue creates an ambiguity, and thus failed to look at the statutory scheme as a whole to resolve this ambiguity. Had they done so, they would have found that the statutory scheme favors the creation and protection of employee rights that are stronger than those of employers. Allowing employers to place employees who are willing and

---

160. The canon expressio unius est exclusio alterius (“the expression of one excludes another”) dictates that where a law delineates a specific process for doing something it excludes or prohibits the use of other processes. See Raleigh & Gaston R.R. Co. v. Reid, 80 U.S. (13 Wall.) 269, 270 (1872).

161. See supra notes 31–36, 43–52, and accompanying text.


163. See supra note 13 and accompanying text.


165. Love, 1997 WL 278126, at *6; Burton, 5 Wage & Hour Cas. 2d (BNA) at 1140.
able to work on involuntary FMLA leave thus clearly violates the integrity of the statute as a whole.

3. **Permitting Involuntary FMLA Leave Contradicts the Command of the Statute’s Construction Provisions that the FMLA Be Construed Broadly in Favor of Employees**

Finally, the construction provisions of the FMLA and its remedial nature mandate liberal construction of the statute. While the statute does not explicitly call for liberal construction, §§ 2651–2653 indicate that courts should construe the FMLA both to maximize employee leave rights and to encourage employers to offer leave policies more generous than those contained in the statute. In addition, the legislative history of the FMLA reveals that the enforcement and remedy provisions of the statute were modeled after those of the FLSA. Courts interpreting the FLSA have repeatedly held that because the FLSA is a remedial statute, courts must interpret it liberally to benefit the workers it seeks to protect. Similarly, because the FMLA is a remedial statute designed to solve workplace problems involving job security and personal obligations, courts should liberally interpret it consistent with this stated purpose and prohibit employers from placing employees who are able to perform their job functions on involuntary leave.

B. **The FMLA Prohibits Employers from Placing Employees Who Can Perform the Essential Functions of Their Positions on Involuntary FMLA Leave**

Courts examining the issue of involuntary leave have ignored the DOL’s regulations and advisory opinions on designation of leave and intermittent FMLA leave. One set of regulations and advisory opinions addresses designation of leave and indicates that an employer may designate leave as FMLA leave and count it against an employee’s annual allotment when the employee and employer meet all of the statutory requirements. Another set of regulations and advisory opinions addresses intermittent leave and indicates that an employer may

168. *See supra* note 96 and accompanying text.
only charge an employee with the amount of FMLA leave necessary to meet the employee’s qualifying need. 170 Both sets of regulations and advisory opinions illustrate the broader principle that it is the employee’s need, and whether that need qualifies under the statute, that dictates whether leave may be taken, and if so, how much. Employers who place employees who are able to perform their essential job functions on involuntary FMLA leave violate this principle.

1. Employers May Only Designate Leave as FMLA Leave Where Such Leave Meets All Statutory Requirements

The regulations and advisory opinions regarding designation of leave affirm that the employer may not designate leave as FMLA leave and deduct it from an employee’s annual entitlement unless the employee meets all of the statutory prerequisites. The FMLA regulations give employers the responsibility of designating leave as FMLA leave.171 In 1995, the DOL issued an advisory opinion stating that where the employer is a covered employer, the employee is eligible for leave, and the reason for leave meets the definition of a serious health condition, the employer may designate leave as FMLA leave and count it against the employee’s allotment, regardless of whether the employee has requested it.172 Thus, where the employee satisfies the requirements for leave, including the statutory definition of a serious health condition, which requires that the employee be unable to perform an essential function of his or her position,173 employers may designate leave as FMLA leave and deduct it from the employee’s annual allotment.174 It follows that where the employee does not meet the statutory requirements for leave, such as not having a serious health condition that prevents him or her from performing essential job functions, the employer may not designate leave as FMLA leave.

Case law illustrates this important distinction. For example, in both Moss v. Formosa Plastics Corp.175 and Harvender v. Norton,176 the

170. See supra Part I.B.2.
171. 29 C.F.R. § 825.208 (2000).
172. 99 Wage & Hour Manual (BNA) 3067, Op. FMLA-68 (July 21, 1995); see also supra notes 64–65 and accompanying text.
employees were in fact unable to perform their respective job functions. In Moss, the employee’s epileptic seizures rendered him unconscious and thus unable to respond to emergency situations at the chemical plant where he was a control-panel operator. In Harvender, the employee’s pregnancy precluded her from handling chemicals, which was an essential function of her position as a lab technician. Thus, their employers acted in accordance with the FMLA regulations and the advisory opinion in placing them on involuntary FMLA leave. The Hicks court recognized the importance of whether the employee has a qualifying need and correctly held that summary judgment was inappropriate where there remained an issue of fact as to whether the employee had an FMLA-qualifying need for leave.

In contrast, in Love the employees did not have a serious health condition as defined by the statute. Although the employees suffered from carpal tunnel syndrome, they were arguably willing and able to perform their essential job functions. Likewise, in Burton the employee was willing and able to perform her essential job functions with reasonable accommodation. By placing these non-qualifying employees on involuntary FMLA leave, these employers deprived the employees of the right to schedule and use their FMLA leave should a genuine qualifying need arise and thereby unlawfully interfered with the employees’ FMLA rights in violation of § 2615(a)(1).

Furthermore, this distinction between those with a qualifying need and those without a qualifying need is very important for individuals with permanent disabilities. Individuals with permanent disabilities include

183. If the employee’s permanent condition meets the ADA’s definition of a disability, reasonable accommodation must be granted where it does not place an undue burden on the employer. See supra note 12. This accommodation could remedy the employee’s inability to perform an essential job function. In the Love decision, while the court denied the employees’ claim that their employer interfered with their FMLA rights when it forced them to take involuntary FMLA leave, the court suggested that putting them on involuntary leave may have given rise to a claim under another
Mr. Keating, whose brain injury resulted in memory impairment, and potentially the *Love* employees with carpal tunnel syndrome. 184 Many such employees have a permanent disability that limits their abilities but do not suffer from any recurring periods of illness. Without such flare-ups that meet the FMLA’s definition of a serious health condition, the FMLA simply does not apply. In these situations, if a non-qualifying employee were to request FMLA leave, the employer would be legally required to deny such a request. It follows that it is equally impermissible for an employer to place a non-qualifying employee on FMLA leave. If the employer needs time to assess whether and how to accommodate an employee who is disabled, as was the case with Mr. Keating’s employer, putting the non-qualifying employee on involuntary FMLA leave as an interim measure violates the FMLA. By allowing employers to do exactly that, an employee who could not request such leave faces the prospect of twelve weeks without pay and loses the right to use his or her FMLA leave later in the year should a genuine serious health condition or other qualifying need arise.

2. **Employers May Only Deduct Leave from an Employee’s Annual FMLA Allotment Where There Is a Qualifying Need**

Regulations regarding intermittent leave forcefully demonstrate the principle that the employee must have a qualifying need for FMLA leave. These regulations dictate that an employer may only charge an employee taking intermittent leave with the amount of leave necessary to meet his or her FMLA-qualifying need. 185 The DOL has issued two advisory opinions clarifying this regulation. 186 These opinions affirm that where an employee is taking FMLA leave on a reduced time or intermittent basis, the employer may not force the employee to take more leave.

---

184. The *Love* court held that the employees had presented evidence sufficient to create an issue of fact regarding whether their carpal tunnel syndrome met the ADA’s definition of disability. 1997 WL 278126, at *6.

185. 29 C.F.R. § 825.205(a) (2000).

leave than is necessary. ¹⁸⁷ For example, an employee who needs to take a few hours of leave for a qualifying medical appointment in the morning cannot be charged with an entire day’s worth of leave, even if the employee works on a plane, train, or boat and is unable to return to the worksite after the appointment.¹⁸⁸ Therefore, if an employer cannot require an employee to take more FMLA leave than actually necessary for a qualifying need, it follows that the employer cannot require an employee to take FMLA leave where the employee has no qualifying need. To hold otherwise would allow absurd results: an employee whose two-hour doctor’s appointment for a qualifying serious health condition caused him to miss a full day’s work assignment would be charged with two hours of FMLA leave, while an employee with no qualifying serious health condition could nonetheless be charged with up to twelve weeks of FMLA leave.

C. Employers Who Place Non-Qualifying Employees on Involuntary FMLA Leave Violate § 2615(a)(1)

Finally, the legislative history of the FMLA suggests that employers may not place employees who are able to perform their essential job functions on involuntary FMLA leave. The Love and Burton courts failed to review thoroughly the history of the FMLA to ascertain true legislative intent. Indeed, the Burton court made no reference to the statute’s legislative history or its purpose,¹⁸⁹ and the Love court only briefly mentioned of the demographic reasons that led to the statute’s enactment.¹⁹⁰ Neither of these courts recognized that permitting employers to place employees who are able to perform their essential job functions on involuntary FMLA leave undermines the FMLA’s stated purpose of providing greater employee job security.

Congress enacted the FMLA as a remedial statute in response to demographic trends that made it difficult for workers to balance the competing demands of their personal and professional lives.¹⁹¹

¹⁸⁹ Burton v. Neumann, 5 Wage & Hour Cas. 2d (BNA) 1138 (S.D. Fla. Apr. 8, 1999).
¹⁹¹ 29 U.S.C. § 2601 (1994); see also supra notes 82–87.
Congress’s primary purpose in enacting the FMLA was to provide employees greater job security by granting them this right to leave.192 As shown in the Love and Burton cases and in Mr. Keating’s case, permitting involuntary FMLA leave in fact results in less job security, especially for employees with disabilities that may place them at risk of falling victim to their employer’s use of the FMLA as a sword to precipitate their resignation. Even if an employer does not terminate an employee after a period of involuntary FMLA leave, that employee’s future job security will be at risk because he or she will be unable to take FMLA leave later that year should a true qualifying need arise. Had the Love and Burton courts delved into the legislative history when examining the issue of involuntary leave and given greater consideration to the social context in which Congress passed the FMLA, they would not have allowed employers to place employees who are able to perform their job functions on involuntary FMLA leave.

Furthermore, the Love and Burton courts failed to consider the undesirable social consequences of permitting employers to place employees on FMLA leave involuntarily. Allowing employers to place employees on involuntary FMLA leave results in undue hardship on employees who are forced to go without pay for the duration of any such leave.193 Because employees in this situation cannot collect unemployment benefits,194 they have no source of income and may be forced to quit their jobs to find new positions. Thus, permitting involuntary leave reduces an employee’s job security, which directly contradicts the statute’s express purpose of giving employees greater job security.195

Similarly, under the rationale of the Love and Burton courts, employers may use involuntary FMLA leave as a pretext for terminating “undesirable” employees who have permanent disabilities196 or who have a family member with such problems. The facts in Burton suggested this type of covert disability discrimination, where the employer summarily placed the employee on leave as soon as she expressed a desire to discuss

193. While the employee may choose or the employer may require the employee to substitute accrued paid leave for unpaid FMLA leave, leave guaranteed by the FMLA is unpaid leave. Id. §§ 2612(c)–(d); see also 29 C.F.R. § 825.207 (2000).
194. Because an employer who places an employee on involuntary FMLA leave has not terminated the employee, the employee is generally ineligible for unemployment benefits. See, e.g., WASH. REV. CODE § 50.04.310 (2000).
196. See supra notes 12 & 183.
reasonable accommodations for her health problems. When employers place employees, particularly disabled employees, on involuntary, unpaid leave, the employees are likely to face financial hardship and may have no choice but to seek other employment. By using involuntary FMLA leave, the employer can get rid of an employee without going through the disability-accommodation process and without terminating the employee, thereby avoiding the risk of discriminatory-discharge allegations.

IV. CONCLUSION

During the eight arduous years required for passage, the FMLA had one central purpose: to create and protect leave rights guaranteeing employees a minimum amount of unpaid leave in times of family or medical need. With such rights, employees would no longer be forced to choose between attending to their most important personal needs and keeping their jobs. Despite its detailed history, provisions, and procedures, the FMLA fails to address the practice of employers putting employees who are willing and able to perform essential job functions on involuntary FMLA leave. Congress probably never contemplated such a practice. However, courts examining the issue of involuntary leave have concluded, with little analysis, that the FMLA’s silence on the issue unequivocally permits the practice. Yet, the language of the statute, the overall statutory and regulatory scheme, the legislative history, and policy concerns are all evidence that Congress intended the FMLA to be a shield for employees, protecting their basic rights and affording them greater job security. The DOL’s regulations and advisory opinions explicitly effectuate this intent. Therefore, employers who use involuntary FMLA leave as a sword against employees who are able to perform their essential job functions, causing them financial hardship and precipitating their job separation, unlawfully interfere with employees’ FMLA rights.

UNOPENED PUBLIC STREET EASEMENTS IN WASHINGTON: WHOSE RIGHT TO USE THAT LAND IS IT, ANYWAY?

Alfred E. Donohue

Abstract: This Comment argues that landowners whose property abuts unopened public street easements have a right to reasonable, non-interfering use of such easements until the city or county opens the street for its intended purpose. Unopened public street easements are dedicated streets that a city or county has not developed or used. Often, landowners use this land to store firewood, park boats, or garden. In 1995, the City of Seattle enacted Municipal Code section 15.02.100, which prohibits all use of unopened public street easements. Several Washington court decisions purportedly support the Seattle ordinance. These decisions suggest that abutting property owners have no legal right to use unopened streets absent permission from the city. However, other Washington court decisions have held that abutting property owners have a right to reasonable, non-interfering use of unopened streets. Under these decisions, this right of use continues until the city or county uses the street for its intended purpose. Other major Washington cities follow this rule. In 1999, the Washington Court of Appeals attempted to reconcile the conflicting decisions but was unable to resolve the issue satisfactorily. This Comment argues that courts and municipalities have misconstrued cases purporting to prohibit all use by the abutting landowner and that an abutting landowner may make reasonable non-interfering use of an unopened street easement.

“The right of an owner of land abutting on public highways has been a fruitful source of litigation in the courts of all the states, and the decisions have been conflicting, and often in the same state irreconcilable in principle.”

Throughout Washington, public easements burden a significant amount of land. They give the city or county the right to construct a street at any time, but until the city or county actually does so, the land remains unused. For example, two Seattle neighborhoods, Magnolia and Queen Anne, have numerous unopened, unused, and unimproved public street easements. These easements total more than 21,000 linear feet in

3. A public street easement is a dedication to a county or city of specific land for use as a street. See State ex rel. York v. Bd. of County Comm’trs, 28 Wash. 2d 891, 904, 184 P.2d 577, 585 (1947). A street is opened when the county or municipality uses the street for travel or transportation. See Karb v. City of Bellingham, 61 Wash. 2d 214, 217, 377 P.2d 984, 985–86 (1963).
these two neighborhoods alone. This amounts to more than 650,000 square feet of unopened public street easements, or nearly fifteen acres of land in two of Seattle’s most expensive neighborhoods.

This Comment examines who has the right to use these unopened public street easements until a street is built. Washington court decisions are not consistent regarding the use of unopened public street easements, and two contradictory views have developed. The first line of decisions treats unopened public street easements akin to private easements. The abutting landowner may use the area covered by an easement in any reasonable way that does not interfere with the easement holder’s rights. However, the abutting landowner’s use may neither interfere with the opening of the street by the city or county nor with the right of neighboring landowners to use the unopened street for access to their own property when necessary. In contrast, the second line of decisions indicates that abutting landowners have no right to use an unopened public street easement, other than for access, light, air, and view.

Amidst this confusion, some cities have enacted ordinances that restrict use by abutting landowners. For example, Seattle and Bellevue both expressly prohibit the use of unopened public street easements

---

4. These figures are based on site surveys and examination of street maps by the author (Aug. 21–23, 2000) (maps available at Kroll Map Co.).

5. Based on site surveys and examination of street maps by the author, (Aug. 21–23, 2000), Magnolia and Queen Anne comprise less than ten percent of Seattle’s total residential area. According to a Building Industry Association of Washington officer, there are substantial dedicated but unopened and unused streets statewide. Interview with Kenneth Donohue, past President, Building Industry Association of Washington, in Olympia, Wash. (Sept. 12, 2000).


8. For the purposes of this Comment, an abutting landowner is the party that owns the land burdened by the easement.

9. Nystrand, 60 Wash. 2d at 795, 375 P.2d at 865 (permitting planting of shrubs and trees, and extension of garage); Thompson, 59 Wash. 2d at 407–09, 367 P.2d at 803–04 (permitting concrete slab, parking cars, “or other appropriate use”); Holm v. Montgomery, 62 Wash. 398, 400, 113 P. 1115, 1116 (1911) (permitting irrigation ditch). However, building a garage, an addition to a house, or a cement retaining wall on the easement might be unreasonable. See Thompson, 59 Wash. 2d at 409, 367 P.2d at 804.

10. Nystrand, 60 Wash. 2d at 795, 375 P.2d at 865.

11. Id.

12. Kemp v. City of Seattle, 149 Wash. 197, 201, 270 P. 431, 433 (1928) (explaining that property abuts street when there is no intervening land between it and street).

without the city’s permission. Others cities, such as Vancouver and Everett, allow reasonable use of unopened public street easements by an abutting landowner, provided the use does not interfere with the city’s use and is not commercial.

This Comment first explains unopened public street easements and their creation. Next, it examines the two conflicting lines of case law addressing the right of abutting owners to use these easements without first obtaining a permit. This Comment continues by examining various municipal ordinances, some of which, such as Seattle Municipal Code (SMC) section 15.02.100, prohibit any use of unopened easements. This Comment then argues that the legal premises for ordinances such as Seattle’s are invalid because the “no use” cases do not apply to the reasonable use of an unopened public street easement. Further, constraining all reasonable use by abutting landowners excessively restricts the rights of landowners to use their property. Such exclusions run counter to public policy favoring the efficient use of land. This Comment concludes that Washington case law, properly read, mandates that abutting owners may make non-interfering use of unopened public street easements.

I. CREATION AND OWNERSHIP OF PUBLIC STREET EASEMENTS

This Part first discusses the creation, opening, and extinguishing of public street easements. Next, it explains who owns the land over which the public street easement runs. Finally, it details the specific rights of use that the general public, neighboring owners, and abutting landowners have in public street easements.

16. For the purposes of this Comment, non-interfering use is any use of an unopened public street easement that will not prevent a city, county, or other party from later using the easement. See, e.g., Thompson v. Smith, 59 Wash. 2d 397, 408, 367 P.2d 798, 803–04 (1962).
A. Creating, Opening, and Extinguishing a Public Street Easement

An easement is one person’s right to make some limited use of land owned by another person. The dominant owner is the party who holds the easement, while the servient landowner is the person who owns the property burdened by the easement. The dominant owner’s right to use the land exists whether or not the easement is improved and opened. For example, when a landowner grants an easement for a driveway, that easement exists even though the driveway has not yet been built. Until the dominant owner constructs the driveway, the easement remains unopened and unimproved.

When a landowner dedicates an easement to the public for use as a street, the landowner creates a public street easement. These dedications, known as statutory dedications, occur when a landowner grants the governmental body a deed or when a city or county approves a plat that marks the location of streets dedicated to public use. A landowner may dedicate previously platted land for a street, provided the landowner follows the procedures set forth by the city or county. This later

17. For the purposes of this Comment, a person is any owner of property, including an individual or corporation.
18. City of Olympia v. Palzer, 107 Wash. 2d 225, 229, 728 P.2d 135, 137 (1986). An easement holder, the dominant owner, is entitled to only limited use of the fee owner’s, the servient owner’s, property. See State ex rel. Shorett v. Blue Ridge Club, 22 Wash. 2d 487, 494, 156 P.2d 667, 671 (1945).
20. An improved street is any street where the municipality or county has made improvements to provide for any primary use, such as grading or paving. See Albee v. Town of Yarrow Point, 74 Wash. 2d 453, 458–59, 445 P.2d 340, 344 (1968). An opened street is one where there is any primary or secondary use by a city or county; for example, laying a sidewalk opens a street, because transportation is a primary use. See State ex rel. York v. Bd. of County Comm’rs, 28 Wash. 2d 891, 904, 184 P.2d 577, 585 (1947). Likewise, a municipality laying gas, water, or power lines along an unimproved right of way opens it because the governmental agency has used the street easement for secondary uses. See id.
24. See WASH. REV. CODE § 58.17.215. See also generally id. §§ 58.08.050, 58.17.033.
Unopened Public Street Easements

dedication of a public street need not be accepted by any public authority to have statutory effect.25

A county or municipality opens a public street easement when it uses the street for any legal purpose.26 This use may be either primary or secondary.27 Primary uses relate to travel over the street, while secondary uses are incident to travel and commerce, such as utility lines and street railways.28 For example, if a city constructs a sidewalk or gravel road, the street easement is opened for its primary use.29 In contrast, if the city lays a sewer line or runs utility lines along the street easement, the street is open for a secondary use.30 Until a city or county uses the street easement in any or all of these manners, it remains unopened and exists only on paper.

Nevertheless, a public street easement exists even if it remains unopened, unused, or unimproved.31 Some street easements are not opened because they lie on steep slopes that make improvement impractical.32 Others are not opened because adequate access to abutting properties exists, making development of streets unnecessary and financially wasteful.33 Further, if a street easement continues through a ravine, creek, or other physical impediment, the improved street may end at the impediment while the unimproved easement or “paper street” continues.34

25. *Loose*, 25 Wash. 2d at 604, 171 P.2d at 852. In addition to statutory dedication, “common law dedications” are unrecorded public street easements created by implication or estoppel. See McConiga v. Riches, 40 Wash. App. 532, 537, 700 P.2d 331, 336 (1985). In this Comment, for uniformity, all references to easements assume a statutory grant is the basis for the dedication.


28. *Id.*

29. *See id.*

30. *See id.*


34. *See Finch*, 74 Wash. 2d at 164, 443 P.2d at 836 (explaining that unopened street ran through ravine and was unimprovable).
A municipality or county may lose a public street easement only by relinquishing the easement. Mere non-use will not extinguish a public or private easement. Equitable estoppel arises when a party’s prior conduct is inconsistent with a claim the party later asserts. For example, if a city is aware an individual is building a house on an unopened street easement and does nothing, a court may equitably estop the city from requiring the house’s removal. Washington courts do not favor extinguishment of a public easement by equitable estoppel, although cases exist in which courts have applied the doctrine.

B. Ownership of Land Underlying Public Easements

In Washington, a landowner generally holds fee simple title to the center of the street abutting the landowner’s property. This fee simple title is subject to the dedicated street easement, which gives a city or county the right to use the street easement. This easement gives the city or county superior rights of use, but title to the land remains in the hands of the abutting landowner. The chain of fee title runs back to the party who originally dedicated the street. If the easement is eliminated by

35. Adverse possession cannot run against a public easement. Erickson Bushling, 77 Wash. App. at 497, 891 P.2d at 752 (citing WASH. REV. CODE § 7.28.090 (1994), preventing adverse possession against municipalities, counties, or state).
37. Equitable estoppel occurs when a party makes admissions or acts inconsistent with a later claim, and an injury to another party relying on that admission or act would occur if that later inconsistent claim were allowed. Lybbert v. Grant County, 141 Wash. 2d 29, 35, 1 P.3d 1124, 1128 (2000).
40. In Washington, a conveyance of property without qualification conveys to the transferee fee simple title. See WASH. REV. CODE § 64.04.060 (2000); see also STOEBUCK, supra note 19, at 4–5.
41. Kemp v. City of Seattle, 149 Wash. 197, 201, 270 P. 431, 433 (1928) (explaining that property abuts street when there is no intervening land between it and street).
42. Burmeister v. Howard, 1 Wash. Terr. 207, 211 (1867) ("When an easement is taken as a public highway, the soil and freehold remain in the owner of the land encumbered only with the right of passage in the public."). Washington courts have followed this rule without exception. E.g., Rainier Ave. Corp. v. City of Seattle, 80 Wash. 2d 362, 366, 494 P.2d 996, 998 (1972); Puget Sound Alumni of Kappa Sigma, Inc. v. City of Seattle, 70 Wash. 2d 222, 226, 422 P.2d 799, 802 (1967).
43. 11 EUGENE M. MCQUILLIN, MUNICIPAL CORPORATIONS 3 (1991) (explaining that underlying owners must yield to municipality or county).
44. See supra note 18 and accompanying text.
45. See Rainier Ave., 80 Wash. 2d at 366, 494 P.2d at 998.
vacation, the exclusive right to use the property vests in the abutting landowners. Moreover, unless the original dedicator reserves the fee in the street, the fee runs with the abutting property when the owner subsequently sells the property.

However, several exceptions to this general rule exist. For example, a city may own the fee to a street. Another exception may arise when a street borders two separate plats dedicated by two parties. In this instance, one of the original dedicators may have originally conveyed the entire street easement, thus retaining the fee to the entire street.

C. Rights of Use in Open Public Street Easements

Washington courts have traditionally classified rights to use open public street easements in three categories. First, the general public has the right to use open public streets regardless of where these members of the public live. This right extends to all reasonable travel and communication of persons and movement of property over the public street easement. This right is granted through public dedication and manifests itself when the city or county opens the street. Thus, until the city or county opens the public street easement, the public has no right to use the easement. However, once the street is opened, the public has the full right to use the street.

---

46. “Vacation” is defined as the elimination of the easement from the underlying land by ordinance or judicial decree. See WASH. REV. CODE § 35.79.010 (2000).
47. Id. § 35.79.040; Puget Sound Alumni, 70 Wash. 2d at 226, 422 P.2d at 802.
49. This Comment assumes that the general principle of fee simple ownership applies.
50. For example, Seattle has owned fee simple title to Lake Washington Boulevard since 1913. See Martin v. City of Seattle, 111 Wash. 2d 727, 730, 765 P.2d 257, 258 (1988).
51. See Rowe v. James, 71 Wash. 267, 271, 128 P. 539, 541 (1912) (holding that deed carries title to farthest side of highway if easement is located wholly on grantor’s land).
52. Id.
54. 11 MCQUILLIN, supra note 43, at 3 (citing Yarrow First Assocs., 66 Wash. 2d 371, 403 P.2d 49).
55. See 10A EUGENE MCQUILLIN, MUNICIPAL CORPORATIONS 371 (3d ed. 1949); see also Moore v. City of Lawrence, 654 P.2d 445, 453 (Kan. 1982).
57. 11 MCQUILLIN, supra note 43, at 3.
Second, neighboring landowners have special rights arising out of property ownership. Whether the street easement is open or unopened, neighboring owner rights include access, air, light, and view. Even if a street abutting their property is unopened, closed, or vacated, neighboring owners may maintain a private easement to access their property, known as “easement implied from plat.” However, a neighboring owner may only use an unopened public street easement if that use is necessary for access to and from the owner’s property.

Third, the abutting landowner’s rights to use the street easement abutting their properties are greater still than neighboring landowners are. An abutting landowner generally holds title to the portion of the public street easement underlying the landowner’s property. Courts have recognized that abutting landowners have a special right to use the opened street abutting their properties. In Fry v. O’Leary, the court recognized that “the rights which an abutting owner of abutting property possesses in a street are different in kind from that possessed by [the general public].” However, the degree and nature of an abutting landowner’s use is less clear when the street is unopened.

58. For the purposes of this Comment, a neighboring owner is one whose property does not abut the particular street or portion of street an abutting owner holds title to, but rather is in the general vicinity of the property. The neighboring owner owns property in the same plat or subdivision and has special rights to all streets in that area. STOEBUCK, supra note 19, at 95.
59. 10A MCQUILLIN, supra note 55, at 355 (citing Shaw v. City of Yakima, 183 Wash. 200, 48 P.2d 630 (1935)).
60. Id.; see also Denman v. City of Tacoma, 148 Wash. 314, 320, 268 P. 1043, 1045 (1928).
61. Denman, 148 Wash. at 320, 268 P. at 1045; see also 10A MCQUILLIN, supra note 55, at 355.
63. Humphrey v. Krutz, 77 Wash. 152, 155, 137 P. 806, 807 (1913) (holding that abutting owner has both public rights in common with everyone and private rights arising from ownership of contiguous property).
64. See supra notes 42–52 and accompanying text.
65. James v. Burchett, 15 Wash. 2d 119, 123, 129 P.2d 790, 792 (1942) (stating that owners of lots bordering streets or ways have right to make reasonable use of street for convenience of their lots, not inconsistent with public’s right to use entire street); 10A MCQUILLIN, supra note 55, at 355.
67. Id. at 471, 252 P. at 113 (quoting Smith v. Centralia, 55 Wash. 573, 104 P. 797 (1909)).
68. The general rule for opened private easements is that a servient landowner may make any use of the area covered by the easement that does not interfere with the dominant owner’s use. Long v. Leonard, 191 Wash. 284, 295–96, 71 P.2d 1, 6 (1937); STOEBUCK, supra note 19, at 113.
II. USE OF UNOPENED PUBLIC STREET EASEMENTS

Although Washington courts have consistently construed the rights of abutting landowners with respect to open public street easements, two distinct lines of cases address the right of abutting owners with respect to unopened public street easements. In the first line of cases, exemplified by Nystrand v. O’Malley, courts focus on whether the use by the abutting landowner interferes with the city’s easement and whether the city has opened the easement. If neither condition is satisfied, courts allow abutting landowners “reasonable use” of the land over which the easement runs, provided the use does not interfere with access of other landowners. The second line of cases, based on Baxter-Wyckoff Co. v. City of Seattle, suggests that, although abutting landowners hold underlying title to the street, until the city vacates its street easement, abutting landowners have no exclusive right of use. Some municipalities have used these cases to justify legislation strictly regulating all uses of unopened public street easements. Others follow a reasonable-use standard as applied by the Nystrand and Thompson v. Smith courts.

A. Cases Permitting an Abutting Landowner’s Reasonable Use of an Unopened Public Street Easement

The Nystrand line of decisions treats unopened public street easements similarly to private easements, permitting abutting landowners reasonable use of the land. This right of reasonable use is subject to the city or county opening the easement and neighboring landowner’s right.

69. 60 Wash. 2d 792, 375 P.2d 863 (1962).
70. Id. at 795, 375 P.2d at 865.
72. 67 Wash. 2d 555, 408 P.2d 1012 (1965).
73. Id. at 560–61, 408 P.2d at 1015–16.
74. See, e.g., SEATTLE, WASH., MUNICIPAL CODE §§ 15.02.046, .048, .100 (1998) (prohibiting any use of unopened streets by any party without permission).
77. Nystrand, 60 Wash. 2d at 795, 375 P.2d at 865. The rule for unopened private easements is similar to opened private easements in that the servient landowner is entitled to any use that does not interfere with the dominant owner’s use. Compare Thompson, 59 Wash. 2d at 407–08, 367 P.2d at 803 (unopened private easement), with Long v. Leonard, 191 Wash. 284, 296, 71 P.2d 1, 6 (1937) (opened private easement).
78. Nystrand, 60 Wash. 2d at 795, 375 P.2d at 865.
to use the unopened street for access. However, if the municipality does not open the street and neighboring owners do not need it for access, these cases consistently hold that the abutting landowner may use the land in any manner consistent with what an owner burdened by a private easement might do with the land.

I. Nystrand v. O’Malley

In Nystrand v. O’Malley, the Washington Supreme Court held an abutting landowner may use the portion of an unopened street easement to which he or she holds fee title in any manner not inconsistent with the easement. The Nystrand court employed a two-part analysis. First, the court asked whether the use by the abutting owner was consistent with the public’s right to the easement. Second, the court examined whether the use challenged interfered with other abutting property owners’ access. The court held that if both questions were answered negatively, then the abutting landowner could make use of the portion of the easement overlying his or her land.

In Nystrand, a twelve-foot-wide unopened public street easement ran directly in front of the Nystrand property. The Nystrands held the underlying fee. Running parallel and abutting this right of way was a second easement owned by Northern Pacific Railroad. A street that provided access to the Nystrand and O’Malley properties lay on the railroad easement. Both the Nystrands’ and O’Malleys’ driveways ran across the twelve-foot unopened street easement to the street located on the railroad easement. The Nystrands extended a garage onto the unopened street, planted trees and hedges, and constructed a bulkhead.

79. Id.
80. Thompson, 59 Wash. 2d at 407–08, 367 P.2d at 803.
81. Nystrand, 60 Wash. 2d at 795, 375 P.2d at 865.
82. Id.
83. Id.
85. Nystrand, 60 Wash. 2d at 793, 375 P.2d at 863–64.
86. Id.
87. Id. at 793, 375 P.2d at 864.
88. Id.
89. Id.
90. Id. at 795, 375 P.2d at 865.
To improve access, the O’Malleys bulldozed an alternate route to their house.\textsuperscript{91} This alternate route crossed the public street easement over the Nystrand fee, and required removal of the bulkhead, trees, and hedge.\textsuperscript{92} When the Nystrands refused to remove the obstructions, the O’Malleys bulldozed them.\textsuperscript{93} The Nystrands brought suit.\textsuperscript{94}

The court rejected the O’Malleys’ argument that the Nystrands wrongfully encroached upon the public street easement.\textsuperscript{95} Because the Nystrands owned the fee to the land underlying the easement where they placed the bulkhead, trees, and hedges, the court determined that they could use the land to which they held fee until their use interfered with the city’s use of the street easement.\textsuperscript{96} First, the court reasoned that the Nystrands’ use of the abutting unopened street was not inconsistent with the public easement because the city had not asserted its right to open the easement.\textsuperscript{97} Second, it found the Nystrands’ use was not improper because the unopened easement was not necessary for reasonable ingress and egress to the O’Malleys’ property.\textsuperscript{98} Accordingly, the court held the Nystrands’ use was proper.\textsuperscript{99}

2. Additional Cases Supporting an Abutting Owner’s Reasonable Use of an Unopened Street Easement

Many other Washington courts have applied a reasonable-use standard. For example, the Washington Supreme Court applied the standard in \textit{Thompson v. Smith}, decided the same year as \textit{Nystrand}. In \textit{Thompson}, Smith, the abutting landowner, constructed a concrete slab over an unused portion of a private street easement.\textsuperscript{100} Thompson, a neighboring owner, sued to enjoin the use, claiming it interfered with the easement.\textsuperscript{101} The court concluded that, until Smith’s use interfered with Thompson’s use of the easement, Smith did not need to remove the...
concrete slab because such use by the abutting landowner was permitted. The court reasoned that such reasonable use must not interfere with the dominant party’s actual use of the easement.

Although Thompson involved a private easement, in dictum the court extended the abutting owner’s right of reasonable use, consistent with Nystrand, to public easements. Furthermore, the court noted that when possible, contemporaneous use by the underlying landowner not inconsistent with the city or county’s use was actually desirable. The Thompson court explained that although the city or county has the right to use its easement for any use consistent with a public right of way, until such use, the servient landowner has a right of reasonable use. As examples of consistent use, the court discussed two California cases. These cases found that neither the rights of a city possessing an easement nor the abutting landowner’s rights are absolute; instead, both the individual property owner and the municipality’s rights should be interpreted to permit reasonable enjoyment by both parties as long as mutual enjoyment is possible. While the city’s use of its easement is superior to the rights of an abutting landowner, the Thompson court reasoned that if the abutting landowner’s use can coincide with the city’s use, then the private use is permissible. Therefore, until a city asserts its rights by occupying and using the easement in some manner, non-interfering use by the abutting landowner would be consistent with the city’s rights, provided that use does not encroach on the easement by requiring substantial expense or effort for removal.

In addition, cases preceding Nystrand and Thompson support a right of reasonable use by the abutting landowner, so long as the use does not affect the rights of the city or county holding that easement. For example,

102. Id. at 407, 367 P.2d at 803.
103. Id. at 408–09, 367 P.2d at 804.
104. Id. at 407–08, 367 P.2d at 803.
105. Id. at 408–09, 367 P.2d at 803–04.
106. Id.
107. Id. at 407–08, 367 P.2d at 803.
110. Id.
111. Id. at 409, 367 P.2d at 804.
in *Holm v. Montgomery*, a neighboring landowner interfered with a ditch in the unopened portion of the public street easement. The abutting landowner sued to enjoin the use. The court, in ruling for the abutting landowner, held that the owner of abutting property was allowed to use land within the highway so long as the use did not interfere with its use as a highway. Because the use for ditches and drains was outside the traveled portion, his use was proper. Hence, the respondent’s interference with the abutting owner’s use constituted a trespass. The Washington Supreme Court subsequently applied this standard of consistent use in *Lanham v. Forney*, where the abutting landowner had installed a pipe under an opened street. The city sued to enjoin the use. The Washington Supreme Court cited to *Holm* and *Colegrove* to support the proposition that unless the use by an abutting landowner interfered with the public’s actual use of the street, it was permissible.

Recent Washington cases have reaffirmed the reasonable-use standard. For example, in *Meresse v. Stelma*, the Washington Court of Appeals found that until a private easement holder used property for the purpose reserved by the easement, the servient landowner could use the land in any manner not interfering with the easement’s proper enjoyment. Consistent with *Thompson*, the court held that this use is qualified; thus, any use considered permanent or requiring substantial cost to remove is not permissible.

Similarly, in *Sandpiper Condominium Ass’n v. Gaylard*, a recent unpublished decision, the Washington Court of Appeals analyzed an

---

112. 62 Wash. 398, 113 P. 1115 (1911).
113. Id. at 399, 113 P. at 1116.
114. Id.
115. Id. at 400, 113 P. at 1116.
116. Id.
117. 196 Wash. 62, 81 P.2d 777 (1938).
118. Id. at 65–66, 81 P.2d at 779.
119. Id. at 65, 81 P.2d at 778.
120. Id.
122. Id. at 867–68, 999 P.2d at 1274.
123. For the purpose of this Comment, “permanent uses” are those that might lead to equitable estoppel. See supra notes 35–39 and accompanying text.
abutting landowner’s use of an unopened public street easement.\textsuperscript{126} In Sandpiper, a landowner shared an unopened street end with a large condominium.\textsuperscript{127} The condominium began to use the entire unopened street easement for additional parking.\textsuperscript{128} The other landowner then constructed a fence extending to the centerline of the easement to eliminate the condominium’s use of the portion of the street easement overlying their land.\textsuperscript{129} Using a two-part “Nystrand test,” the court concluded that until a municipality opens a public street easement, the abutting landowner has a right to reasonable use.\textsuperscript{130} However, this right of use is qualified by the other abutting landowners’ need to access their own property.\textsuperscript{131} In applying this test, the court concluded that the fence did not interfere with the city’s use of the easement because the city had not asserted its right to open the easement and that the abutting landowner’s fence did not interfere with the condominium’s need for reasonable access.\textsuperscript{132}

\textbf{B. Cases Prohibiting an Abutting Landowner’s Use of an Unopened Public Street Easement Without Approval}

Despite the historical application of the reasonable-use standard, a distinct line of cases suggests an abutting landowner has no right to use any public street easement other than for access, light, air, and view.\textsuperscript{133} These cases focus on a municipality’s right to regulate any use of public street easements.\textsuperscript{134} Although a municipality or county may permit use by the abutting landowner, it may also prohibit all uses.\textsuperscript{135} Accordingly, even though a city or county has not opened a public street easement,

\textsuperscript{126} Id. at *2.
\textsuperscript{127} Id.
\textsuperscript{128} Id.
\textsuperscript{129} Id.
\textsuperscript{130} Id. at *5. The abutting owner’s need for access qualifies this right, along with the city’s assertion of the easement. \textit{Id}.
\textsuperscript{131} Id. at *5–6.
\textsuperscript{132} Id.
\textsuperscript{133} Baxter-Wyckoff Co. v. City of Seattle, 67 Wash. 2d 555, 561, 408 P.2d 1012, 1016 (1965) (“The lack of right of the abutting owner to use the street in front of his property does not depend on his interference with an actual or proposed public use of the street. The abutting owner simply has no legal right to make this kind of use of the dedicated public street.”).
\textsuperscript{134} Id.
these cases suggest it still may control and dictate any use of the land over which the easement runs. 136

1. Baxter-Wyckoff Co. v. City of Seattle

In Baxter-Wyckoff Co. v. City of Seattle, 137 Seattle lumber mills challenged the city’s authority to charge fees for its continued permanent use of an unopened public street easement. 138 The mills leased the property on both sides of the street and claimed the right to use the street. 139 The street was dedicated in 1897 but never opened. 140 Before 1955, Baxter-Wyckoff and Nettleton had built several structures. 141 In 1955, Seattle enacted an ordinance imposing fees for exclusive private commercial use of unopened public street easements. 142 In 1958, Baxter-Wyckoff and Nettleton refused to pay any further fees and sued the city to enjoin the collection of any fees. 143 They argued that because fee title to an unopened street easement lies with the abutting landowner, the city of Seattle could not charge fees that amounted to rent. 144

The Washington Supreme Court did not refer to Nystrand or the Nystrand analysis, but only considered whether Baxter-Wyckoff and Nettleton’s “Permanent Encroachments Are Inconsistent with the Public’s Easement for Travel in Southwest Florida Street.” 145 The City of Seattle argued that because streets were dedicated primarily for public travel, secondary and subordinate uses were allowed only when they did not interfere with public travel. 146 The court concluded: “The basic rule applicable to this case is that there is no inherent right in a private individual to conduct private business in the public streets.” 147 The court reasoned that neither individuals nor municipalities may permanently encroach on a street without a permit for private use. 148 The court held

137. 67 Wash. 2d 555, 408 P.2d 1012 (1965).
138. Id. at 557, 408 P.2d at 1014.
139. Id. at 556–57, 408 P.2d at 1013–14.
140. Id. at 556, 408 P.2d at 1013.
141. Id. at 557, 408 P.2d at 1013.
142. Id. at 557, 408 P.2d at 1014.
143. Id.
144. Id.
145. Id. at 559, 408 P.2d at 1015 (emphasis added).
146. Id.
147. Id. at 560, 408 P.2d at 1015.
148. Id. at 561, 408 P.2d at 1016.
the city could impose any terms and conditions it chose because the permanent use of a street was a mere privilege, not a right. To support this conclusion, the court cited several cases allowing regulation of private use of public streets. These cases all concerned private individuals conducting business on opened streets, not unopened ones.

In dictum, the court also discussed the rights of abutting property owners. It limited these rights to access, light, air, water, and certain temporary or nonexclusive uses of the street easement. It further indicated an abutting property owner has no right to use the street abutting its property exclusively, regardless of whether the use interferes with any actual or proposed public use of the street. Although the actual holding of the case solely addressed an abutting landowner’s lack of right to construct permanent commercial structures on street easements and a city’s right to regulate this extraordinary use in any manner it chooses, subsequent cases have expanded the Baxter-Wyckoff decision beyond this holding.

2. The Expansion of Baxter-Wyckoff To Prohibit Any Use of Unopened Public Street Easements by Abutting Owners

Subsequent cases have relied on Baxter-Wyckoff for the proposition that a city may regulate any use of a city street, and that any regulation is

149. Id. at 562, 408 P.2d at 1017.
150. Id. at 560, 408 P.2d at 1015–16 (citing McGlothern v. City of Seattle, 116 Wash. 331, 335, 199 P. 457, 458 (1921) (holding that regulating jitney buses on public streets permitted); State v. City of Spokane, 109 Wash. 360, 364–65, 186 P. 864, 866 (1920) (holding that regulating jitney buses by ordinance not precluded by state code); Hadfield v. Lundin, 98 Wash. 657, 662–63, 168 P. 516, 517–18 (1917) (holding that regulating taxicabs on public streets permitted); Allen v. City of Bellingham, 95 Wash. 12, 35, 163 P. 18, 26 (1917) (holding that regulating jitney buses permitted)).
152. Id. at 561, 408 P.2d at 1016.
153. Id. at 561–62, 408 P.2d at 1016–17.
154. Snohomish County Pub. Util. Dist. No. 1 v. Broadview Television Co., 91 Wash. 2d 3, 586 P.2d 851 (1978); City of Seattle v. Samis Land Co., 55 Wash. App. 554, 779 P.2d 277 (1989). Part of the confusion surrounding subsequent applications of Baxter-Wyckoff might arise from its first headnote, which states “abutting property owners [do not] have any inherent right to have the exclusive private use of, or to maintain permanent structures on, any area dedicated as a public street.” Baxter-Wyckoff, 67 Wash. 2d at 555. This headnote implies that the case overruled Nystrand by prohibiting all private use of unopened streets. However, Baxter-Wyckoff does not address this issue and discusses only permanent use. See infra notes 186–87 and accompanying text (discussing confusion surrounding “so use” language in Baxter-Wyckoff).
"entirely within the discretion of the city council."155 Under these cases, whether the street is opened or unopened does not affect the city’s regulatory authority.156 In addition, these cases do not distinguish whether the use regulated is permanent or non-permanent, commercial or non-commercial.157 Instead, such cases cite to Baxter-Wyckoff for the general proposition that a municipality or governmental agency may regulate any use of any opened or unopened street.

For example, in Snohomish County Public Utility District No. 1 v. Broadview Television Co.,158 the Washington Supreme Court reaffirmed its holding in Baxter-Wyckoff.159 It noted “there is no inherent right in a private individual to conduct private business in the public streets.”160 The court called the right to use a street “a mere privilege, which the city could grant or withhold,” and focused on the absolute right of the state and cities to control streets.161 Although Broadview concerned private use of public utility poles, rather than an unopened street easement, the Washington Supreme Court found the factual differences between Baxter-Wyckoff and Broadview immaterial and instead focused on the government’s authority to regulate use.162 By extending Baxter-Wyckoff to the use of public utility poles, the court focused on the general right to regulate public streets, rather than the type of use regulated.163

In City of Seattle v. Samis Land Co.,164 the Washington Court of Appeals held that the city could regulate any commercial use of an

159. Id. at 10, 586 P.2d at 856.
160. Id.
161. Id. at 10–11, 586 P.2d at 856 (“The state, and the city as an arm of the state, has absolute control of the streets in the interest of the general public. No private individual . . . has a right to the use of the streets . . . without the consent of the state.”) (quoting Hadfield v. Lundin, 98 Wash. 657, 660, 168 P. 516, 517 (1917)).
162. Id. (tracing authority to Hadfield v. Lundin, 98 Wash. 657, 660, 168 P. 516, 517 (1917)).
163. Id.
opened street easement by an abutting landowner.165 Samis, a property-
development company, owned several buildings that extended out
underneath opened public street easements.166 The city, relying on its
authority to regulate such uses, imposed a fee for this permanent use,
which Samis refused to pay.167 Samis argued that as the underlying fee
holder to the street, the city should not charge Samis for de minimis,
non-interfering use of the easement.168 Citing Baxter-Wyckoff, the court
held the city had the authority to regulate and charge fees for any type of
use.169 The court further held that an underlying landowner’s use need
not interfere with the public’s use because the city could regulate any
use.170 The City of Seattle and other municipalities have construed this
holding to justify land use ordinances restricting any use of unopened
public street easements by the underlying fee owner.171

C. Municipal Ordinances Regulating Abutting Landowner’s Use of
Unopened Public Street Easements Embody the Conflict Between
the Nystrand Rule and the Expansion of the Baxter-Wyckoff
Doctrine

Using Baxter-Wyckoff and Samis as support, some cities have enacted
ordinances that restrict any use of an unopened public street easement by
abutting landowners. The cities of Seattle and Bellevue, among others,
prohibit any use of unopened streets without permission.172 For example,
SMC section 15.02.100, enacted in 1995, prohibits activities on
unopened public street easements without a permit.173 The prohibited
activities include erecting fencing, storing material, planting trees or

165. Id. at 557, 779 P.2d at 278–79.
166. Id. at 556, 779 P.2d at 278.
167. Id. at 555–56, 779 P.2d at 278.
168. Id. at 557, 779 P.2d at 278.
169. Id. at 559–60, 779 P.2d at 280.
170. Id. at 560, 779 P.2d at 280. Baxter-Wyckoff’s headnotes might have caused this inter-
pretation. See supra note 154.
171. See SEATTLE, WASH., MUNICIPAL CODE, ch. 15.02 (1998) (citing Samis for proposition that
Seattle has authority to charge fees for permission to occupy portion of public street for private use).
This general provision provides the authority for chapter 15.02, which includes SEATTLE, WASH.,
MUNICIPAL CODE section 15.02.100 (barring “use” of unopened street without permit, as defined by
section 15.02.046).
172. See SEATTLE, WASH., MUNICIPAL CODE § 15.02.100; BELLEVUE, WASH., CITY CODE
173. See SEATTLE, WASH., MUNICIPAL CODE § 15.02.100 (regulating all “use” of “public
places”).
shrubs, or disturbing the surface without a permit. 174 This ordinance explicitly relies on Samis, which authorizes municipal regulation. 175 While this ordinance appears consistent with the holdings in the Baxter-Wyckoff line of cases, it conflicts with the Washington court decisions that allow abutting landowners to make reasonable non-interfering use of unopened street easements. 176

Other cities, such as Vancouver and Everett, do not restrict all use of unopened public street easements; instead, they are consistent with the Nystrand reasonable-use model. 177 Vancouver regulates only “improvements” in areas “legally open to the public use,” such as streets, sidewalks, roadways, and alleys. 178 Everett’s ordinance does not restrict any use of unopened public street easements for parking, landscaping, or single-family uses that “will not interfere with the public convenience or the health, safety or welfare of the general public.” 179

III. WASHINGTON CASE LAW DOES NOT PROHIBIT REASONABLE, NON-PERMANENT USES OF UNOPENED PUBLIC STREET EASEMENTS BY ABUTTING LANDOWNERS

On their facts, Baxter-Wyckoff and Samis both address permanent use of public streets. Neither case actually supports regulating non-interfering reasonable uses of an unopened public street easement by abutting landowners. This Comment argues that language in the Baxter-Wyckoff decision centers on permanent use and that the Samis court misapplied this language, creating confusion about the rights of abutting landowners. Seattle’s use of Samis to support its regulation of all uses of unopened public street easements exemplifies this confusion. Furthermore, because the Nystrand line of reasonable-use cases remains good law and unequivocally speaks to non-interfering uses of unopened public

174 See id. § 15.02.048 (defining “public place” in section 15.02.046 as street or right of way, whether opened and improved or not).
175 See id. ch. 15.02.
176 See supra notes 77–132 and accompanying text.
178 See VANCOUVER, WASH., MUNICIPAL CODE §§ 11.60.010 (defining public street), .050 (setting forth activities on rights of way that require permit).
179 EVERETT, WASH., MUNICIPAL CODE § 13.30.040(A)(2) (requiring permits but not fees for use of right of way by abutting owners), § 13.84.010 (defining right of way to not include unopened and unimproved street easements).
street easements, a reasonable-use model should apply. Properly limited to their facts, neither Baxter-Wyckoff nor Samis conflicts with this standard. Additionally, although case law clearly suggests that a city may comprehensively regulate opened street easements, case law does not support prohibiting abutting owners from non-interfering use of unopened public street easements. Thus, municipalities such as Seattle should not prohibit or limit non-interfering reasonable use of unopened street easements by abutting landowners, and instead should encourage the efficient use of urban land.

A. Ordinances Such as SMC Section 15.02.100 Should Not Rely on Samis To Restrict Use by Abutting Owners

Washington case law does not support SMC section 15.02.100 because the Seattle ordinance regulates all use of unopened public street easements by abutting landowners regardless of the nature of that use. 180 Samis, the case purportedly providing explicit support for the Seattle ordinance, only addresses the regulation of permanent use of opened streets. 181 Likewise, Baxter-Wyckoff does not support a city prohibiting all uses of an unopened public street easement because Baxter-Wyckoff dealt with permanent commercial use of an unopened public street easement.

1. Samis Provides No Authority for SMC Section 15.02.100 Because the Samis Court Erroneously Interpreted the Washington State Supreme Court’s Holding in Baxter-Wyckoff

Decisions relying on Baxter-Wyckoff mistakenly conflate the permanent and non-interfering use of public street easements. 182 Furthermore, Samis concerned permanent use of opened streets and did not address unopened street easements. The Samis court determined only that a Washington statute allowed cities to regulate permanent encroachments under sidewalks. 183 Because sidewalks are opened public street easements, Samis does not hold that a city may regulate any use of

180 See SEATTLE, WASH., MUNICIPAL CODE § 15.02.046 (defining “public place”), .048 (defining “use”).
181 See supra notes 164–69 and accompanying text.
182 See supra notes 137–69 and accompanying text.
an unopened public street easement. The court’s analysis centered on whether Samis’s use of the public street easement needed to interfere with the public’s use of the easement for the city to regulate the use.\textsuperscript{184} Citing to \textit{Baxter-Wyckoff}, the court concluded that there did not need to be interference with the public’s use to allow municipal regulation.\textsuperscript{185}

However, the \textit{Samis} court misinterpreted dictum from \textit{Baxter-Wyckoff}.\textsuperscript{186} The \textit{Samis} court quoted \textit{Baxter-Wyckoff} for the proposition that “the lack of right of the abutting owner to so use the street . . . does not depend on his interference with an actual or proposed public use of the street.”\textsuperscript{187} The court emphasized the language “lack of right” instead of the words “so use” in the quoted passage. This emphasis is important because “so use” in \textit{Baxter-Wyckoff} specifically referred to the permanent use of the street.\textsuperscript{188} By taking this dictum out of context, \textit{Samis} employed the proposition of no use incorrectly. This misinterpretation dramatically changes the \textit{Baxter-Wyckoff} holding from prohibiting permanent commercial use of easements to prohibiting any use of easements. Moreover, \textit{Baxter-Wyckoff} and other decisions simply do not support prohibiting abutting owners from making reasonable non-permanent use of an unopened public street easement.\textsuperscript{189} By incorrectly applying \textit{Baxter-Wyckoff} and expanding its holding to all uses of street easements, the \textit{Samis} court grossly misinterpreted \textit{Baxter-Wyckoff}. To the extent that cities such as Seattle use \textit{Samis} as a foundation for ordinances restricting use of unopened public street easements by abutting landowners, their validity is specious.

\begin{itemize}
\item \textsuperscript{184} Id. at 561, 779 P.2d at 281.
\item \textsuperscript{185} Id.
\item \textsuperscript{186} Id. at 559–60, 779 P.2d at 279–80 (citing Baxter-Wyckoff v. City of Seattle, 67 Wash. 2d 555, 561, 408 P.2d 1012, 1016 (1965)).
\item \textsuperscript{187} Samis, 55 Wash. App. at 559–60, 779 P.2d at 280 (quoting Baxter-Wyckoff v. City of Seattle, 67 Wash. 2d 555, 561, 408 P.2d 1012, 1016 (1965)).
\item \textsuperscript{188} Baxter-Wyckoff v. City of Seattle, 67 Wash. 2d 555, 561, 408 P.2d 1012, 1016 (1965). “The rule applies particularly to exclusive private use of the street and the construction and maintenance of permanent structures.” Id.
\item \textsuperscript{189} For the purposes of this Comment, a “non-permanent” use is a long term use by an abutting owner that cannot lead to equitable estoppel.
\end{itemize}
2. Baxter-Wyckoff Addresses Only the Regulation of Permanent Uses of Unopened Public Street Easements, Not Reasonable Non-Permanent Uses by Abutting Landowners

Baxter-Wyckoff did not address an abutting landowner’s non-interfering and non-permanent use of an unopened street easement; instead, it centered on the permanent use of an unopened street easement. The court expressly stated that the only issue it was considering was whether “Respondent’s [Baxter-Wyckoff and Nettleton’s] Permanent Encroachments Are Inconsistent with the Public’s Easement for Travel in Southwest Florida Street.” The court concluded “there is no inherent right in a private individual to conduct private business in the public streets.” It also emphasized that permitting the construction of permanent buildings in the street “is so unusual, and beyond the ordinary authority and power of a municipality” that enabling state legislation is needed for the city to permit such use.

Baxter-Wyckoff was unique because it involved an abutting owner’s permanent use of a public street easement, a use that could have led to equitable estoppel. In Baxter-Wyckoff, the mills constructed numerous permanent buildings and structures in the unopened street. It was likely the concern of the court that by not allowing the city to regulate this extremely unusual use, the city might be estopped from later opening the easement. Because equitable estoppel may extinguish an easement, cities may restrict permanent use on unopened streets. While the permanent use of an unopened easement can affect the city’s rights, an

190. Id. (holding that use of street by abutting owners for permanent structures is so unusual that it may be regulated regardless whether it interferes with public).
191. Id. at 559, 408 P.2d at 1015 (emphasis added).
192. Id. at 560, 408 P.2d at 1015 (emphasis added).
193. Id. at 561, 408 P.2d at 1016.
194. See supra notes 37–39 and accompanying text (discussing equitable estoppel).
abutting owner’s non-interfering “reasonable use” does not pose the same threat of equitable estoppel.\footnote{198}{Thompson v. Smith, 59 Wash. 2d 397, 407–08, 367 P.2d 798, 803 (1962).}

Additionally, the language used by the court restricting an abutting owner’s rights is confusing. The court stated that the “recognized text authorities” limit an abutting owner’s rights to temporary or nonexclusive use of unopened street easements.\footnote{199}{Baxter-Wyckoff, 67 Wash. 2d at 562, 408 P.2d at 1016–17 (citing EUGENE A. MCQUILLIN, MUNICIPAL CORPORATIONS 647–56 (3d ed. 1949), and CHESTER JAMES ANTEAU, MUNICIPAL CORPORATION LAW 606, 622 (1955)).} The Baxter-Wyckoff court’s use of the language “temporary and non-exclusive” is distinct from “reasonable” or “non-interfering” use.\footnote{200}{Id. at 562, 408 P.2d at 1016–17.} The meaning intended by the court is very specific and includes such temporally limited uses as parking a car on an unopened street easement and unloading it, or playing catch on the unimproved street easement.\footnote{201}{See, e.g., Gabrielsen v. City of Seattle, 150 Wash. 157, 168–69, 272 P. 723, 727 (1928) (recognizing deposit of building materials, machinery, and equipment on street, necessary for construction of abutting building as temporary); see also 10A MCQUILLIN, supra note 55, at 356 (listing temporary uses).} The court’s use of “temporary or nonexclusive” is distinct from “non-interfering” uses, which would include the construction of a fence or horseshoe pit, for example. Moreover, because the court did not distinguish “temporary and non-exclusive” uses from “non-interfering” uses in its discussion, a mistaken belief might arise that they are the same, when they are not.

More importantly, the recognized authorities cited by the Baxter-Wyckoff court did not clearly support the court’s dictum limiting an abutting owner’s right to use unopened street easements. McQuillin, the treatise cited extensively as support for this proposition, relies on rules allowing regulation of opened streets, not unopened ones.\footnote{202}{10A MCQUILLIN, supra note 55, at 356–57. The court’s use of McQuillin is inappropriate because the quoted portion focuses on the right to use opened streets and streets in general. Id. McQuillin’s discussion of abutting owner’s rights focuses on use of the street and sidewalk (discussing awnings, signs, steps, and other encroachments found on opened streets). In addition, the Baxter-Wyckoff court ignores non-permanent use in its brief discussion. Baxter-Wyckoff, 67 Wash. 2d at 561, 408 P.2d at 1016–17.} This omission resulted in a conflation by the Washington Supreme Court of opened and unopened streets that is troubling. In fact, Antieau, cited by the court as support along with McQuillin, does at times distinguish between opened and unopened streets. It is in its discussion of an abutting owner’s rights to an opened street that the Antieau treatise restricts use to temporary obstructions of streets, and access, light, air,
and view. However, Antieau also discusses reasonable use, stating that, “when an abutter owns the fee he can use the property in the street in any way not interfering with the public easement.” Antieau’s discussion contradicts the Baxter-Wyckoff court’s assertion that abutting owner’s rights are limited to access, light, air, water, view, and temporary uses. Further, the most recent edition of Antieau on Local Government Law reaffirms the right of an abutting owner to use a street area so long as it does not interfere with the public’s use. Prior and subsequent cases allowing abutting owners to make non-interfering reasonable use of unopened public street easements expose the vulnerability of the Baxter-Wyckoff court’s dictum. Yet, this dictum has led some municipalities to believe that abutting landowners have no right to use unopened streets for any purpose in any circumstances. However, on close analysis, neither the holding, underlying purpose, nor dictum of Baxter-Wyckoff suggests a city may prohibit an abutting owner’s reasonable non-interfering use of an unopened public street easement.

B. Reasonable Nonpermanent Use of an Unopened Public Street Easement by Abutting Owners Is Consistent with Prior Case Law

By requiring abutting landowners to seek permission to use unopened easements in any manner, SMC section 15.02.100 ignores Nystrand and other Washington case law allowing abutting landowners to use the unopened street as long as it does not interfere with the city’s use. The legislature may delegate the authority to regulate streets to cities through enabling legislation. See State ex rel. York v. Bd. of County Comm’rs, 28 Wash. 2d 891, 898, 184 P.2d 577, 582 (1947); see also 10A McQuillin, supra note 55, at 301–02. As an enabling statute, WASH. REV. CODE section 35.22.280(7) (2000) comes closest to providing a basis for SMC section 15.02.100. However, the language of that statute focuses on the rights of cities to “lay out [and] establish . . . streets . . . and to regulate the use thereof,” not on regulating unopened streets. WASH. REV. CODE § 35.22.280(7). The statute does not authorize regulation before laying out or establishing a street, and courts have never interpreted it to apply to non-permanent use in unopened streets. Therefore, if the sole authority for SMC section 15.02.100 is Samis, there may be no legal basis for section 15.02.100. The courts have not addressed whether Washington Revised Code
Unopened Public Street Easements

*Nystrand* line of cases remains good law and holds that abutting landowners may make reasonable use of an unopened public street easement in any manner not inconsistent with public street easements or neighboring owner’s access. In *Nystrand*, the court stated that the O’Malleys’ use of the unopened street, in extending their garage onto the unopened easement, planting trees and a hedge, and constructing a bulkhead, “was not inconsistent with the public’s easement since the right to open the street for the public’s use had not been asserted by the city.” Thus, the use by the abutting landowner did not interfere with the city’s use of the easement, satisfying the first part of the *Nystrand* test. Furthermore, the landowner’s use of the unopened public street easement did not interfere with the neighboring owners’ access, satisfying the second part of the test. By satisfying both prongs of the test, the O’Malleys’ use was permitted.

The foundation of this reasonable-use principle is *Colegrove Water Co. v. City of Hollywood*, consistently cited by the Washington Supreme Court to support the proposition that abutting landowners may reasonably use unopened public street easements, so long as the use does not interfere with the public’s full enjoyment of the easement. In *Colegrove*, the California Supreme Court held that as long as the use by an abutting landowner of the street did not interfere with the public’s rights, it was permissible. This holding was endorsed by the Washington Supreme Court in 1911, when it adopted this rule and decided that an abutting landowner could construct a ditch in the unopened portion of the abutting public street. Likewise, in *Lanham v. Forney* the court permitted reasonable use by abutting landowners, where the abutting

---

210. See supra notes 81–99 and accompanying text.
212. *Id.*
213. *Id.*
214. 90 P. 1053 (Cal. 1907).
218. 196 Wash. 62, 81 P.2d 777 (1938).
landowner had installed a pipe under what later became an opened street, and two city officials sued to enjoin its use.\textsuperscript{219}

Since the Washington Supreme Court adopted \textit{Colegrove}, it has consistently allowed abutting owners to make use of the land subject to the public easement where such use is not inconsistent with that easement.\textsuperscript{220} These cases are consistent with \textit{Nystrand}’s approach of allowing reasonable use by abutting landowners. For example, the court in \textit{Thompson v. Smith}\textsuperscript{221} permitted reasonable use by the abutting landowner until that use interfered with the opening of the street easement.\textsuperscript{222} The Washington Court of Appeals recently followed \textit{Thompson}’s rule of reasonable use in \textit{Meresse v. Stelma},\textsuperscript{223} where it determined that an abutting landowner could reasonably use a private street while unopened.\textsuperscript{224} Thus, Washington still adheres to the reasonable-use standard for unopened and unimproved public street easements.

\textbf{C. Cities May Not Prohibit the Non-Permanent Reasonable Use of Unopened Public Street Easements}

Washington cases have repeatedly held that public easements do not convey complete control over the property dedicated. In fact, Antieau, one of the treatises cited by the \textit{Baxter-Wyckoff} court, states that fee owners may use their fee so long as their use does not interfere with the municipality’s use.\textsuperscript{225} If the municipality is not using the easement, reasonable use by an abutting owner cannot interfere with the city’s rights to the easement. Additionally, a long line of Washington decisions supports limiting the scope of public street easements by restricting the rights of the public easement holder to uses connected with transportation and commerce.\textsuperscript{226} This limitation arises because an easement dedication grants only limited rights to use the land, not fee ownership.

\textsuperscript{219} \textit{Id.} at 66, 81 P. 2d. at 779.

\textsuperscript{220} \textit{See, e.g.}, \textit{Bd. of County Comm’rs}, 28 Wash. 2d at 902, 184 P.2d at 584 (listing cases permitting non-interfering use of easement); \textit{see also} \textit{Thompson v. Smith}, 59 Wash. 2d 397, 407–08, 367 P.2d 798, 803 (1962).

\textsuperscript{221} 59 Wash. 2d 397, 367 P.2d 798 (1962).

\textsuperscript{222} \textit{Id.} at 407–08, 367 P.2d at 803.

\textsuperscript{223} 100 Wash. App. 857, 999 P.2d 1267 (2000).

\textsuperscript{224} \textit{Id.} at 867–68, 999 P.2d at 1274.

\textsuperscript{225} 2 \textit{STEVENVSON}, \textit{supra} note 206, at 30–82 (quoting \textit{Nystrand v. O’Malley}, 60 Wash. 2d 792, 375 P.2d 863 (1962) in its text to support this proposition).

\textsuperscript{226} \textit{E.g.}, \textit{State ex rel. York v. Bd. of County Comm’rs}, 28 Wash. 2d 891, 902, 184 P.2d. 577, 584 (1947); \textit{Motoramp Garage Co. v. City of Tacoma}, 136 Wash. 589, 591, 241 P. 16 (1925).

566
Generally, Washington courts have not held that municipalities and counties have anything more than an easement of use in public streets. The only right a municipality or county acquires from the easement is the right to use the land for a public street. Nystrand and Thompson support this proposition that a public street easement does not convey unlimited control over the property dedicated. If the public has not asserted the easement by opening it, then reasonable use by an abutting landowner is consistent with the public’s present and future rights to use the easement.

Consider under Thompson a situation in which a private easement holder, who had yet to assert the easement by opening it, put up a sign stating: “The fee holder may not use this land except for access without my permission.” A court would conclude that this attempt to prevent an abutting landowner’s reasonable use was improper because the easement holder had not begun to use the easement. Likewise, following Nystrand, to exercise direct control over the easement, the city or county must assert its rights by opening the easement. Merely placing a sign at the beginning of the easement, or regulating the easement by enacting a statute, does not open it. The hypothetical situation above would also apply to public easements because the easement conveys to a city or county the right to open a street. To hold otherwise, prohibiting reasonable non-interfering uses by abutting landowners, would essentially equate dedication of an easement with conveyance of fee title ownership.

Therefore, it is contradictory to allow municipalities and counties to act as owners by effectively precluding abutting landowners from making any non-interfering use of the land over which the easement runs. Additionally, these regulations effectively preclude landowners

227. See, e.g., Bd. of County Comm’rs, 28 Wash. 2d at 898, 184 P.2d at 582 (“[N]ormally, the interest acquired by the public is but an easement.”); Erickson Bushing Inc. v. Manke Lumber Co., 77 Wash. App. 495, 498, 891 P.2d 750, 752 (1995) (“[W]hen land is dedicated to the public for a street or road, the public acquires only an easement.”).

228. Finch v. Matthews, 74 Wash. 2d 161, 168, 443 P.2d 833, 838 (1968) (holding that only right acquired by municipality was easement for passage and use for street purposes).


230. See Nystrand, 60 Wash. 2d at 795, 375 P.2d at 865.

231. See supra notes 100–07 and accompanying text.

232. See Nystrand, 60 Wash. 2d at 795, 375 P.2d at 865.

233. Rowe v. James, 71 Wash. 267, 270, 128 P. 539, 540–41 (1912) (“We have uniformly held that a city acquires only an easement in a street in consequence of a dedication.”).
from challenging the city because the cost to litigate these rights is prohibitive. In this manner, municipal regulations that prohibit use by abutting owners remain unchallenged.

D. Public Policy Favors Permitting Abutting Landowners’ Non-Permanent Reasonable Use of Unopened Public Street Easements

Restricting abutting landowners’ reasonable non-interfering use of unopened public street easements prevents the efficient use of land. Municipalities regulate opened streets primarily to ensure public safety.234 Municipalities may seek to regulate unopened public street easements for similar reasons. However, cities are not liable for injuries on unimproved (and unopened) public street easements.235 Therefore, the need to regulate these unopened streets is minimal. If the city or county has not asserted the easement by opening it, the public is not using the street and the need to protect the public through regulation diminishes.236

Municipal restrictions that prohibit non-interfering reasonable use by abutting landowners furthers waste. When an abutting landowner uses an unopened street easement to store firewood, establish a dog run, or simply to extend a backyard, the fee owner enjoys this previously unused land. Restricting this reasonable use by the abutting landowner serves no benefit, except to increase revenue for municipalities.237

Furthermore, many unopened street easements were dedicated in the 1920s and 1930s and lie over terrain where it is impossible to build a street. Abutting landowners should be able to make non-interfering reasonable use of this marginal land until the city or county needs it. Such use leads to an efficient utilization of resources. Consequently, abutting landowners should have rights of reasonable non-interfering use to the unopened public street easements abutting their property.

234. 10A MCQUILLIN, supra note 55, at 332.
235. Barton v. King County, 18 Wash. 2d 573, 575, 139 P.2d 1019, 1020 (1943) (holding that “municipality is not liable for injuries sustained outside the improved portion of the street or highway”).
236. Id.; see also 10A MCQUILLIN, supra note 55, at 332.
237. However, regulating non-interfering land use as a revenue source is suspect. See Baxter-Wyckoff v. City of Seattle, 67 Wash. 2d 555, 564–67, 408 P.2d 1012, 1018–19 (1965) (Hunter, J., dissenting) (focusing on whether fees relate to cost of administration, or instead are revenue raising).
CONCLUSION

Reasonable, non-interfering use of unopened public street easements by abutting landowners should be exempt from prohibitions on use and fees by municipalities. Neither Baxter-Wyckoff nor Samis addresses reasonable non-interfering use of an unopened street easement. If courts and lawmakers properly confine Baxter-Wyckoff and Samis to their facts and holdings, these cases do not actually conflict with the Nystrand line of decisions. Accordingly, ordinances such as Seattle Municipal Code section 15.02.100 should not rely on Samis to support restricting abutting landowners’ reasonable non-interfering use of unopened public street easements because Samis applies to neither unopened street easements nor non-interfering use of these easements. Further, the Nystrand line, which explicitly addresses abutting landowner’s use of unopened public street easements, does not support such restrictions; instead, these cases permit reasonable non-interfering use. If the use by the abutting landowner does not interfere with the city’s potential use of the easement, or neighboring owners’ rights of access, then the use is reasonable. Finally, as land in Seattle neighborhoods such as Queen Anne and Magnolia becomes both more valuable and exceptionally scarce, the desire and need to use this land will increase. Thus, abutting landowners should have the right to reasonable non-interfering use of unopened public street easements.
GRANDMA GOT RUN OVER BY THE SUPREME COURT: SUGGESTIONS FOR A CONSTITUTIONAL NONPARENTAL VISITATION STATUTE AFTER TROXEL v. GRANVILLE

Eric B. Martin

Abstract: Every state in the Union has a statute allowing for court-ordered child visitation by non-parents. Until the summer of 2000, the U.S. Supreme Court had never ruled on the constitutionality of such statutes. When the Court finally tackled Washington’s statute in Troxel v. Granville, the Court left the most significant questions unanswered, while casting doubt on the validity of Washington’s statute. Prior to Troxel, the Washington Supreme Court had held Washington’s nonparental visitation statute facially unconstitutional, finding that the statute violated the Fourteenth Amendment rights of parents. After granting certiorari, the U.S. Supreme Court held Washington’s statute unconstitutional as applied and refused to reach the question of facial unconstitutionality. This Comment proposes three changes to Washington’s nonparental visitation statute that would ameliorate the objections voiced by the U.S. Supreme Court regarding the application of the statute: the Washington Legislature should limit the classes of persons allowed to petition for visitation, codify the common law rebuttable presumption that a parent’s decision regarding visitation is in the best interest of the child, and add a purpose section to the nonparental visitation statute. This Comment concludes that with these changes, Washington’s nonparental visitation statute would be constitutional and Washington’s lower courts would have the guidance needed to constitutionally apply the statute in a manner consistent with precedent.

Consider fourteen-year-old Riley. Riley’s mother never got along with her parents-in-law, but Riley and his father visited them frequently, allowing Riley to develop strong emotional ties to those family members and their ethnic and religious traditions. Riley’s father recently died. Riley’s mother now refuses her former in-laws permission to visit Riley. Alternatively, imagine a nine-year-old boy named Royce. Royce has an absentee father, who has never had any significant contact with his child. From the time Royce was two, his mother lived with her boyfriend Kevin, who served as a de facto parent to Royce. Recently, his mother

1. These scenarios are hypotheticals created by the author.
2. For purposes of this Comment, a de facto parent is a person who, while having no legal or biological relationship to the child, has nonetheless acted as a parent to the child, usually while in a relationship with one of the child’s parents. De facto parents are increasingly common in homosexual relationships in which one of the partners has a biological child, but the other partner has not adopted, or legally cannot adopt, the child. See generally Recent Case, E.N.O. v. L.L.M., 711 N.E.2d 896 (Mass.), Cert. Denied, 120 S. Ct. 500 (1999), 113 HARV. L. REV. 1551 (1999) [hereinafter Recent Case]; see also In re B.G. v. San Bernardino County Welfare Dep’t, 523 P.2d 244, 253 n.18 (Cal. 1974).
and Kevin separated for reasons unrelated to Royce. Kevin would like to continue to see Royce and continue to provide him the parental guidance he has provided up to this point in time. However, Royce’s mother does not want to see her ex-boyfriend and therefore will not allow Kevin to spend time with her son. In the case of both Royce and Riley, continued visitation with the non-parents would be in the best interest of the child.

In 1996, twenty-eight percent of U.S. children lived with only one parent and four million children lived with their grandparents. In addition, although no state recognizes same-sex marriages, an increasing number of same-sex partners are raising children. Because one of the same-sex partners has no biological relationship with the child, statutory law may prevent that partner from adopting the child. Finally, second marriages or subsequent non-marital relationships are increasingly common. Often a stepparent does not legally adopt children from a partner’s or spouse’s previous relationship, perhaps because the child’s other biological parent wishes to maintain a legal relationship, or because the stepparent simply does not see the value in completing the legal steps.

Until 1998, Washington law allowed a court to order visitation rights for non-parents if such visitation would be in the best interest of the child. However, a 1998 Washington Supreme Court holding, affirmed on other grounds by a 2000 U.S. Supreme Court decision, raised serious doubts regarding the law’s validity. This Comment argues that the Washington State Legislature should amend Washington’s nonparental visitation statute by authorizing only certain relatives and de facto parents to bring nonparental visitation actions. Also, the Legislature should require courts to apply a rebuttable...
Nonparental Visitation Rights

presumption that parental decisions regarding visitation are in the best interest of their children. Part I of this Comment provides an overview of the constitutional rights of parents, the tension between parental rights and the state’s parens patriae power, and the history of nonparental visitation laws, in particular Washington’s law. Part II explores the conflict between the state’s parens patriae power and the Constitutional rights of parents. Further, it analyzes the holdings of the cases—In re Smith in the Washington Supreme Court and Troxel v. Granville in the U.S. Supreme Court—that questioned the continuing validity of Washington’s nonparental visitation law and explains the status of current state law in light of these two cases. Part III argues that three specific amendments to Washington’s nonparental visitation statute would address the U.S. Supreme Court’s concerns in Troxel and allow for the proper balance between the state’s parens patriae power and parents’ constitutional rights. This Comment concludes that, with these changes, Washington courts could constitutionally apply Washington’s nonparental visitation statute and not misinterpret precedent in light of the Smith decision.

I. CONSTITUTIONAL RIGHTS OF PARENTS, THE STATE’S PARENS PATRIAE POWER, AND NONPARENTAL VISITATION STATUTES

Parents’ constitutional rights to control the upbringing of their children are liberty interests derived from the Fourteenth Amendment. However, these rights are not unlimited. While the line of cases establishing parental rights is well-developed, the Court has not clearly explained the extent to which a state, through its parens patriae power to act in its citizens’ interests, may interfere with these rights. Reconciling these competing interests is necessary to evaluate properly nonparental visitation statutes.

10. See infra Part I.B.


12. Troxel, 120 S. Ct. at 2066 (Souter, J., concurring).
A. Parents Have a Limited Right To Raise Their Children Without State Interference

It is beyond question that parents have some right to direct the upbringing of their children. The U.S. Supreme Court has repeatedly affirmed this principle since *Meyer v. Nebraska*\(^{13}\) in 1923. Yet the Court has limited these rights by allowing state involvement in the welfare of children where there is a compelling state interest since *Prince v. Massachusetts*\(^{14}\) in 1944.

Parents’ rights regarding the care and custody of their children are well-established,\(^{15}\) but these rights are not unlimited. The U.S. Supreme Court articulated these rights in *Meyer*, *Pierce v. Society of Sisters*,\(^{16}\) and subsequent cases.\(^{17}\) In *Meyer*, the Court determined that the liberty interest guaranteed by the Fourteenth Amendment included at least a partial right to raise children as a parent sees fit.\(^{18}\) Invalidating a state law prohibiting the teaching in schools of any language other than English, the Court said that the liberty interest denoted freedom “to engage in any of the common occupations of life, . . . [including bringing] up children.”\(^{19}\) Similarly, in *Pierce*, the Court held unconstitutional a statute requiring children to attend only public schools because it unreasonably interfered with the liberty of parents and guardians to direct the development of their children.\(^{20}\) In striking down the law, the Court noted that constitutional rights may not be abridged by legislation that has no reasonable relation to some purpose within the state’s regulatory competence.\(^{21}\)

However, in *Prince*, the Court held that the family is not beyond state regulation.\(^{22}\) While responsibility for “custody, care and nurture” resides

\(^{13}\) 262 U.S. 390 (1923).

\(^{14}\) 321 U.S. 158 (1944).

\(^{15}\) See *Smith*, 137 Wash. 2d at 13, 969 P.2d at 27.

\(^{16}\) 268 U.S. 510 (1925).


\(^{18}\) See *Meyer*, 262 U.S. at 399–400.

\(^{19}\) Id. at 399.

\(^{20}\) See *Pierce*, 268 U.S. at 520; see also Kathleen S. Bean, Grandparent Visitation: Can the Parent Refuse?, 24 J. FAM. L. 393, 410 (1985).

\(^{21}\) See *Pierce*, 268 U.S. at 520.

Nonparental Visitation Rights

first in the parents of the child, the state has authority in matters affecting a child’s welfare and thus has a broad range of powers to limit parental freedom and authority in those areas. In *Prince*, the Court upheld an application of Massachusetts’s child-labor law prohibiting girls under eighteen years of age from selling publications (here religious literature) on the streets. Rejecting assertions of parental control and free exercise of religion, the Court held that the state’s power to protect the child from harm does not depend on the presence or direction of the child’s guardian. The Court did not discuss the exact parameters of the state’s power to intrude into parental rights and religious freedom. Regarding the limits of state power, the Court held only that “the rightful boundary of [the state’s] power has not been crossed in this case” and noted that “[o]ur ruling does not extend beyond the facts the case presents.

In later cases, the Court described the liberty interest as applying to the family unit instead of strictly to parents. In *Moore v. City of East Cleveland*, a case involving living arrangements, the Court noted that:

> [O]urs is by no means a tradition limited to respect for the bonds uniting the members of the nuclear family. The tradition of uncles, aunts, cousins, and especially grandparents sharing a household along with parents and children has roots equally venerable and equally deserving of constitutional recognition. . . . Even if conditions of modern society have brought about a decline in extended family households, they have not erased the accumulated wisdom of civilization . . . that supports a larger conception of the family.

Similarly, in *Quilloin v. Walcott*, the U.S. Supreme Court upheld the termination of a father’s parental rights based on the best interest of the

23. Id.
24. Id.
25. Id.
26. After noting the “custody, care and nurture of the child reside first in the parents,” the Court went on to hold that “neither rights of religion nor rights of parenthood are beyond limitation . . . the state has a wide range of power for limiting parental freedom and authority in things affecting the child’s welfare.” *Id.* at 166–67 (citations omitted).
27. *Id.* at 170.
28. *Id.* at 171.
30. *Id.* at 504–05.
child, even though there was no finding that the father was unfit or that
the child suffered harm from the parental relationship. Quilloom
involved the termination of a biological father’s parental rights in order
to allow for the adoption of the child by the child’s stepfather. The
stepfather had been involved in the child’s life for eleven years and was
married to the child’s mother, while the father had taken little or no
interest in the child and had never married the child’s mother. Likewise
in Lehr v. Robertson, the Court noted that a parent’s liberty interests do
not “spring full-blown from the biological connection between the parent
and child. They require relationships more enduring.”

B. The State’s Parens Patriae Power Allows for Regulation of Child
Welfare

The parens patriae doctrine allows a state to protect its quasi-
sovereign interests in the “health, comfort, and welfare of its citizens.” The
document of parens patriae, literally “parent of the country,” comes
from English common law. Originally, the monarch assumed this role
to act as a guardian for minors, incompetents, and those “who have no
other protector.” States have used this authority to pursue environ-
mental, antitrust, and mass tort litigation, such as tobacco suits on behalf
of citizens. This authority is also used to pursue child-welfare cases. The
state’s parens patriae power thus fits within the state’s police power
“to regulate public health and safety, maintain the peace, and provide for
the general welfare.”

32. Id. at 255–56.
33. Id. at 257.
34. Id.
36. Id. at 260.
38. Natalie Loder Clark, Parens Patriae and a Modest Proposal for the Twenty-First Century:
39. See Ratliff, supra note 37, at 1850.
40. See Clark, supra note 38, at 382.
41. See id. (quoting 2 FREDERICK POLLOCK & FREDERIC WILLIAM MAITLAND, THE HISTORY
   OF ENGLISH LAW 445 (2d ed. 1968)).
42. See Ratliff, supra note 37, at 1848–49.
43. Clark, supra note 38, at 382; see also Ratliff, supra note 37, at 1847.
44. Justice Philip A. Talmadge, The Myth of Property Absolutism and Modern Government: The
Nonparental Visitation Rights

*Parens patriae* is the most common basis for state involvement in child-custody disputes. In *In re Sumey*, the Washington Supreme Court upheld the temporary residential placement of a child outside the home as a valid exercise of the state’s *parens patriae* power, even though there was no assertion of parental unfitness or harm to the child. Similarly, in *In re Welfare of Key*, the court upheld the state’s *parens patriae* power to hold a dependency hearing over the objection of the child’s mother. The court stated that the interest of the mother did not turn on her fitness, rather that “presence or absence of unfitness would seem to affect only the weight of the State’s interest.” Even in child-custody decisions, courts exercise the *parens patriae* power to make decisions based on the best interests of the child. Thus, although the right to custody, care, and nurturing of children resides first in their parents, parents’ power must give way to the child’s best interest when the state exercises its *parens patriae* powers.

The Washington Supreme Court has utilized *parens patriae* power in holding that the family is not beyond regulation in the public interest. The state utilizes a wide range of *parens patriae* power for limiting parental freedom and authority in matters affecting a child’s welfare. The Washington Supreme Court held in *Sumey* that:

[T]he liberty and privacy protections of the due process clause of the Fourteenth Amendment establish a parental constitutional right to the care, custody, and companionship of the child. . . . The parents’ constitutional rights, however, do not afford an absolute protection against State interference with the family relationship . . . [g]rowing concern for the welfare of the child and the disappearance of the concept of the child as property has led to a gradual modification in judicial attitude. It is now well established

---

46. 94 Wash. 2d 757, 621 P.2d 108 (1980).
47. *Id.* at 762–65, 621 P.2d at 110–12.
49. *Id.* at 610–13, 836 P.2d at 206–07.
50. *Id.* at 611, 836 P.2d at 206.
that when parental actions or decisions seriously conflict with the physical or mental health of the child, the state has a parens patriae right and responsibility to intervene to protect the child.\textsuperscript{55}

Similarly, in \textit{State v. Koome},\textsuperscript{56} the court recognized that “although the family structure is a fundamental institution of our society, and parental prerogatives are entitled to considerable legal deference . . . they are not absolute and must yield to fundamental rights of the child or important interests of the State.”\textsuperscript{57} The concept is again found in \textit{In re K.R.},\textsuperscript{58} where the court accepted the application of the state’s parens patriae power to terminate parental rights without requiring an explicit finding of parental unfitness.\textsuperscript{59} The court\textsuperscript{60} stated that “when the rights of parents and the welfare of their children are in conflict, the welfare of the minor children must prevail.”\textsuperscript{61}

\section*{C. The History of Nonparental Visitation Laws and Washington’s Nonparental Visitation Statute}

At common law, no legal right for nonparental visitation existed. However, every state has enacted some sort of nonparental visitation statute.\textsuperscript{62} While the U.S. Supreme Court has not directly addressed the limits of parents’ constitutional rights in the context of nonparental visitation statutes,\textsuperscript{63} the majority of state courts addressing the issue has

\begin{footnotes}
\item[55.] \textit{In re Sumey}, 94 Wash. 2d 757, 762, 621 P.2d 108, 110 (1980) (internal citations and quotation marks omitted).
\item[56.] 84 Wash. 2d 901, 530 P.2d 260 (1975).
\item[57.] \textit{Id.} at 907, 530 P.2d at 264.
\item[58.] 128 Wash. 2d 129, 904 P.2d 1132 (1995).
\item[59.] \textit{Id.} at 146, 904 P.2d at 1141 (Madsen, J.).
\item[60.] The \textit{In re K.R.} opinion was authored by Justice Madsen, who also authored the court’s opinion in \textit{Smith}. \textit{Id.}; \textit{Smith v. Stillwell-Smith}, 137 Wash. 2d 1, 969 P.2d 21, \textit{aff’d on other grounds sub nom. Troxel v. Granville}, 120 S. Ct. 2054 (2000).
\item[61.] \textit{In re K.R.}, 28 Wash. 2d at 146, 904 P.2d at 1141.
\item[62.] While most statutes address grandparent visitation, some statutes allow for other nonparental visitation, including visitation by other relatives, non-relatives who have significant relationships with the child, including de facto parents, and, in the case of Washington’s statute, “any person.” \textit{See generally Phyllis C. Borzi, Note, Statutory Visitation Rights of Grandparents: One Step Closer to the Best Interests of the Child}, 26 CATH. U. L. REV. 387 (1977). While there may be different policy questions raised by the different statutes, the constitutional questions remain the same regardless of the parties permitted to petition for visitation. This Comment uses the term “nonparental visitation” to refer to all of these types of statutes.
\item[63.] \textit{See infra} Part II.D.1.
\end{footnotes}
found these statutes comport with the federal constitutional protections given to the liberty and privacy interests.64

1. Nonparental Visitation Under the Common Law

At common law, grandparents and other non-parents generally had no legal right to petition for visitation with children that were not their own.65 Society recognized a grandparent’s right to visitation with a child “as the custodial parent’s moral obligation, but not an enforceable legal right.”66 The few exceptions to the common law rule were based on a custodial parent’s voluntary relinquishment of exclusive control over the child.67 As a result, those seeking visitation had no legal recourse other than to challenge the fitness of the parent.68 If the custodial parent was found to be a fit parent, the party seeking visitation had to accept the decision of the parent.69 The common law rule resulted in an ironic situation that made it easier to sue for custody of children, because that decision was predicated on the “best-interest-of-the-child” standard, than to petition for visitation, where the decision was based on the stricter “harm-to-the-child” standard.70

64. See infra Part II.A.
67. See Collins, supra note 65, at 59; see also Henry H. Foster, Jr. & Doris Jonas Freed, Grandparent Visitation: Vagaries and Vicissitudes, 23 ST. LOUIS U. L.J. 643 (1979);
Exception to the usual common law rule is made in three instances: (1) when there was an agreement or stipulation as to visitation made, for example, incident to a divorce proceeding; (2) when the child has resided with the person seeking visitation, as for example, in a case in which, the child’s custody originally having been awarded to a parent who lived with grandparents, the custodial parent died and the surviving parent seeks a change of custody; and (3) when it is demonstrated that the parent seeking custody is “unfit” under the prevailing notions at the time.
Id. at 645–46 (citations omitted).
68. Borzi, supra note 62, at 387.
69. See id.
70. See Foster & Freed, supra note 67, at 647.
2. Nonparental Visitation Statutes

In the late 1960s, state legislatures, seeking to provide visitation rights to grandparents, relatives, stepparents, de facto parents, and others, began to enact nonparental visitation statutes. Currently all fifty states have some form of nonparental visitation statute. These statutes commonly give trial courts the discretion to award visitation rights to non-parents after a showing that it would be in the best interest of the child to do so. Many states limit the jurisdiction of the court to certain circumstances such as divorce, death of a parent, or other pending custody or visitation action. Many states also limit who may bring nonparental visitation actions to, for example, grandparents or other blood relatives, de facto parents, or stepparents.

71. For a definition of “de facto parent” see supra note 2.
72. See Borzi, supra note 62, at 392.
74. See Borzi, supra note 62, at 392.
75. See Ladd, supra note 66, at 639; see also Collins, supra note 65, at 61.
Nonparental Visitation Rights

Washington has two nonparental visitation statutes. The one at issue in *Smith* and *Troxel* allows for visitation petitions to be brought by “any person” at “any time” so long as it is in the “best interest of the child.” The other statute allows for a petition for visitation only upon a divorce proceeding, and requires that the judge make a determination that visitation would be in the “best interest of the child.” However, the visitation statute at issue in *Smith* and *Troxel* provides no guidance for trial courts to use in determining the best interest of the child. In contrast, a trial court using the divorce statute is guided by several statutory factors:

(a) The strength of the relationship between the child and the petitioner;
(b) The relationship between each of the child’s parents or the person with whom the child is residing and the petitioner;
(c) The nature and reason for either parent’s objection to granting the petitioner visitation;
(d) The effect that granting visitation will have on the relationship between the child and the child’s parents or the person with whom the child is residing;
(e) The residential time sharing arrangements between the parents;
(f) The good faith of the petitioner;
(g) Any criminal history or history of physical, emotional, or sexual abuse or neglect by the petitioner; and
(h) Any other factor relevant to the child’s best interest.

When the Washington Supreme Court was presented with the *Smith* case, it confronted not the divorce statute, but the other nonparental visitation statute. The court sought to reconcile parental constitutional rights and the state’s *parens patriae* power.

---

78. *Id.* § 26.10.160(3) (“Any person may petition the court for visitation rights at any time including, but not limited to, custody proceedings. The court may order visitation rights for any person when visitation may serve the best interest of the child whether or not there has been any change in circumstances.”); *Smith v. Stillwell-Smith*, 137 Wash. 2d 1, 5, 969 P.2d 21, 23 *aff’d on other grounds sub nom.* *Troxel v. Granville*, 120 S. Ct. 2054, 2057 (2000).
80. *Id.*
II. THE CONSTITUTIONALITY OF WASHINGTON’S NONPARENTAL VISITATION STATUTE UNDER THE SMITH AND TROXEL DECISIONS

At the heart of the constitutional question regarding nonparental visitation statutes is the inherent conflict between the constitutional rights of parents and the parens patriae power of the state. The Washington Supreme Court, in three consolidated cases, bucked the trend of other state courts and held in Smith v. Stillwell-Smith that Washington’s nonparental visitation statute facially violated the U.S. Constitution.81 The U.S. Supreme Court, without a majority opinion, affirmed the Washington Supreme Court’s holding in Troxel, but only as applied.82 In Troxel, the Court did not rule on the facial constitutionality of the interpretation given the statue by the Washington Supreme Court.83 The constitutionality of nonparental visitation statutes therefore remains unaddressed by the ultimate arbiter of constitutionality.

A. State Court Reactions to Nonparental Visitation Statutes

While the U.S. Supreme Court has yet to truly weigh in on the matter, a majority of the states faced with the question have framed the primary issue presented by nonparental visitation statutes as the tension between the state’s parens patriae power and a parent’s constitutional rights regarding child rearing.84 Nonparental visitation statutes have generally met with acceptance by state courts.85 The majority view accepts nonparental visitation statutes that are premised on the best interests of the child.86

81. See Smith, 137 Wash. 2d at 5, 969 P.2d at 23.
83. Id.
86. See, e.g., King, 828 S.W.2d at 631–32.
The few courts that have held nonparental visitation statutes unconstitutional have done so on state constitutional grounds.\textsuperscript{87} For example, in \textit{Beagle v. Beagle},\textsuperscript{88} the Supreme Court of Florida held that the state’s nonparental visitation statutes violated the enhanced privacy rights found in article 1, section 23 of the Florida Constitution, which provides privacy protections “broader in scope” than the U.S. Constitution.\textsuperscript{89} However, Washington’s constitution affords no greater protection than the protections provided by the U.S. Constitution on matters of privacy, other than in the area of search and seizure.\textsuperscript{90}

\textbf{B. State Court Proceedings in Smith v. Stillwell-Smith}

The Washington Supreme Court’s holding in \textit{Smith v. Stillwell-Smith} is the result of three consolidated, somewhat typical,\textsuperscript{91} nonparental visitation actions.\textsuperscript{92} \textit{Clay v. Wolcott}\textsuperscript{93} involved the former male companion of the mother of a six-year-old boy.\textsuperscript{94} The man had developed a close relationship with the child while living with the child’s mother for the child’s second through sixth years of life.\textsuperscript{95} In \textit{Smith v. Stillwell-Smith}, the parents of a child had filed for divorce.\textsuperscript{96} While the divorce was pending, the maternal grandmother of the child shot and killed her son-in-law.\textsuperscript{97} The father’s surviving family petitioned for visitation with his child.\textsuperscript{98} In \textit{In re Visitation of Troxel},\textsuperscript{99} Tommie Granville and Brad Troxel


\textsuperscript{88} 678 So. 2d 1271 (Fla. 1996).

\textsuperscript{89} \textit{Id. at 1275–76.}

\textsuperscript{90} \textit{Ramm v. City of Seattle, 66 Wash. App. 15, 27, 830 P.2d 395, 402 (1992); see also WASH. CONST. art. I, § 7.}

\textsuperscript{91} \textit{See Brief for Petitioners at 3, Troxel v. Granville, 120 S. Ct. 2054 (2000) (No. 99-138), available at 1999 WL 1079965 (citing more than 700 reported cases).}

\textsuperscript{92} \textit{Smith v. Stillwell-Smith, 137 Wash. 2d 1, 969 P.2d 21 (1998), aff’d on other grounds sub nom. Troxel v. Granville, 120 S. Ct. 2054 (2000).}

\textsuperscript{93} 85 Wash. App. 468, 933 P.2d 1066 (1997).

\textsuperscript{94} \textit{Smith, 137 Wash. 2d at 5–6, 969 P.2d at 23–24.}

\textsuperscript{95} \textit{Id.}

\textsuperscript{96} \textit{Id.}

\textsuperscript{97} \textit{Id.}

\textsuperscript{98} \textit{Id. at 7, 969 P.2d at 24.}

had a long-term relationship that resulted in two daughters.\(^{100}\) The couple separated in 1991.\(^{101}\) At that time, Brad Troxel moved in with his parents Jenifer and Gary Troxel.\(^{102}\) In 1993, two years after the separation, Brad Troxel committed suicide.\(^{103}\) His parents continued to see their granddaughters on a regular basis after the death of their son.\(^{104}\) However, in October 1993, Ms. Granville informed them that she wished to limit their visitation with her daughters.\(^{105}\) The Troxels responded by filing a petition for visitation under Washington’s nonparental visitation statute\(^{106}\) in county superior court.\(^{107}\) The trial court found that visitation was in the best interest of the children and issued an order granting the Troxels visitation.\(^{108}\)

In a five-to-four decision, the Washington State Supreme Court held that Washington’s relevant nonparental visitation statute\(^{109}\) was contrary to the liberty interest of the Fourteenth Amendment; therefore, it held the statute facially unconstitutional.\(^{110}\) The court relied primarily on the Meyer-Pierce line of cases describing the scope of parents’ constitutional rights.\(^{111}\) The court held that a parent’s fundamental right to autonomy in child-rearing decisions is unassailable absent a showing of harm to the child.\(^{112}\) Likewise, it held that the state cannot invoke its parents patriae power absent a showing of parental unfitness or harm to the child.\(^{113}\) The court held that the only compelling state interests authorizing state intrusion on family life were protecting citizens from threats to health or safety or from injuries inflicted by third persons.\(^{114}\) Thus, the decision, in addition to invalidating the statute, rejected as unconstitutional the best-

\(^{100}\) Smith, 137 Wash. 2d at 6, 969 P.2d at 25–26.
\(^{101}\) Id.
\(^{102}\) Id.
\(^{103}\) Id.
\(^{104}\) Id.
\(^{105}\) Id.
\(^{107}\) Smith, 137 Wash. 2d at 5–6, 969 P.2d at 23–24.
\(^{108}\) Id.
\(^{109}\) WASH. REV. CODE § 26.10.160(3).
\(^{110}\) Smith, 137 Wash. 2d at 5, 969 P.2d at 23.
\(^{111}\) See id. at 13–18, 969 P.2d at 27–29; supra Part I.A.
\(^{112}\) Smith, 137 Wash. 2d at 19, 969 P.2d at 30.
\(^{113}\) See id. at 20, 969 P.2d at 30.
\(^{114}\) See id.
interest-of-the-child standard and thus sharply limited the *parens patriae* power of the state.

In addressing the *parens patriae* power of the state, the court held that there must first be harm to the child before the state may exercise this power.\(^{115}\) It attempted to distinguish its previous holding in *In re Sumey*.\(^{116}\) That case held that the state need not demonstrate unfitness of a parent before exercising its *parens patriae* power but instead must weigh the interest of the state action against the intrusion on the parent’s rights.\(^{117}\) The *Smith* court declined to use the *Sumey* balancing test, believing that the proper standard was only whether, absent visitation, there was harm to the children.\(^ {118}\) While this standard contradicted the precedent of *Sumey*, the *Smith* court indicated that the statute in question in *Sumey* involved potential harm to children.\(^{119}\) While suggesting that the *Sumey* court improperly used the balancing test, the court went on to find that “[n]evertheless, the court’s result was correct because the interests of the state in that case . . . were compelling and the statute was narrowly tailored to serve the state’s interest.”\(^ {120}\) This odd sidestep around the *Sumey* balancing test gave rise to the dissent’s major disagreement with *Smith*.

Writing for a four-justice dissent, Justice Talmadge condemned the decision’s “cruel and far-reaching effects on loving relatives.”\(^ {121}\) The dissent contended that the majority opinion misinterpreted the U.S. Supreme Court’s opinions dealing with parental rights.\(^ {122}\) In addition, the dissent questioned the majority’s reliance on cases that involved other substantial infringements of parental or others’ rights.\(^ {123}\) The dissent argued that the rights of parents are not unlimited and that “the state has a wide range of power for limiting parental freedom and authority in

\(^{115}\) See id.


\(^{117}\) Id. at 763–65, 621 P.2d at 111–12.

\(^{118}\) *Smith*, 137 Wash. 2d at 19, 969 P.2d at 30.

\(^{119}\) Id. at 19 n.4, 969 P.2d at 30 n.4.

\(^{120}\) Id.

\(^{121}\) Id.

\(^{122}\) Id.

\(^{123}\) Id. at 23, 969 P.2d at 32 (Talmadge, J., concurring in part, dissenting in part).

\(^{124}\) Id. (Talmadge, J., concurring in part, dissenting in part).

\(^{125}\) Id. at 26–27, 969 P.2d at 33–34 (Talmadge, J., concurring in part, dissenting in part) (discussing Santosky v. Kramer, 455 U.S. 745 (1982) (termination of parental rights); Wisconsin v. Yoder, 406 U.S. 205 (1972) (free exercise of religious beliefs); Stanley v. Illinois, 405 U.S. 504 (1972) (termination of parental rights); Pierce v. Soc’y of Sisters, 268 U.S. 510 (1925) (finding that parental authority to send children to private religious or military schools is otherwise acceptable)).
things affecting the child’s welfare.” 124 The dissent further averred that Sumey was controlling and that its balancing test was appropriate. 125

C. Standards of Review Used in Reviewing State Nonparental Visitation Statutes

When the U.S. Supreme Court reviews a challenge to a state statute based on the U.S. Constitution, the state’s highest court and the U.S. Supreme Court have combined responsibility for resolving the issue. 126 The state court interprets the statute in question, and the U.S. Supreme Court rules on the constitutionality of that statute as interpreted. 127 However, even if the U.S. Supreme Court holds a statute as interpreted by the state supreme court unconstitutional, either facially or as applied, the state court may salvage the statute by either reconsidering the interpretation or severing the statute’s unconstitutional language. 128

The Court applies a strict scrutiny test when analyzing state legislation that may interfere with a family’s liberty interest. 129 In order to survive strict scrutiny, the state government must show a compelling state interest in the area encompassed by the legislation. 130 In addition, the state must narrowly tailor legislation to obtain the results desired by the legislation and cause minimal infringement on fundamental rights. 131

D. The U.S. Supreme Court’s Holdings in Troxel v. Granville

In Troxel, the Supreme Court accepted review of one of the three cases ruled on in Smith. The Court affirmed the Washington Supreme Court’s ruling.

125. See Smith, 137 Wash. 2d at 23, 969 P.2d at 32 (Talmadge, J., concurring in part, dissenting in part).
127. Id. at 284.
128. Id.
130. Roe, 410 U.S. at 155.
131. Id. Additionally, the constitutionality of any standard for awarding visitation rights turns on the specific manner in which that standard is applied. Troxel v. Granville, 120 S. Ct. 2054, 2064 (2000). The constitutional protections in this area are best elaborated with care. Id. Because much state court adjudication in this context occurs on a case-by-case basis, the U.S. Supreme Court should be hesitant to hold specific nonparental visitation statutes per se unconstitutional. See id.
Court, finding the statute unconstitutional, but only as applied. The Court specifically refused to rule on the facial constitutionality of the Washington statute. While the Court’s six opinions (three concurring, three dissenting) did not resolve the question of the constitutionality of nonparental visitation statutes, two overriding principles may be gleaned from the opinions. Specifically, the Court indicated that the class of people allowed to bring nonparental visitation actions should be limited and that courts must give parents a rebuttable presumption that their decisions with respect to visitation are in their child’s best interest.

1. The Court’s Opinions in Troxel v. Granville

The U.S. Supreme Court decision in Troxel arrived in six separate opinions. Justice O’Connor wrote the lead opinion, joined by Chief Justice Rehnquist, and Justices Ginsburg and Breyer. Justices Souter and Thomas each concurred in the judgment and filed separate opinions. Justices Stevens, Scalia, and Kennedy each dissented and filed separate opinions. The Court held Washington’s nonparental visitation statute unconstitutional only as applied to Tommie Granville in In re Visitation of Troxel. The four justices represented by the O’Connor opinion declined to reach the question of whether a finding of harm is required prior to a court’s granting visitation to non-parents. Justice Souter believed the statute was unconstitutional on its face due to its overbreadth. Justice Thomas concluded that nonparental visitation was not a compelling state interest, and therefore, beyond the state’s parens patriae power to regulate. Because these opinions sweep

132. Troxel, 120 S. Ct. at 2064.
133. Id.
134. Id.
135. See id. at 2060–61.
136. See id. at 2062.
137. Id. at 2057.
138. Id.
139. Id.
140. Id.
141. Id. at 2064.
142. Id.
143. See id. at 2065–67 (Souter, J., concurring).
144. See id. at 2067–68 (Thomas, J., concurring).
broader than the O’Connor opinion, the plurality opinion is the narrowest opinion, and therefore the holding of the court.145

2. Guidance Gleaned From the Troxel Opinions

While the Troxel Court refused to address whether Washington’s nonparental visitation statute violated the Fourteenth Amendment, the Troxel opinions provides several clues as to how the Court may treat nonparental visitation statutes in the future. Any future nonparental visitation statute must address two principle concerns raised by the U.S. Supreme Court opinions.

First, the O’Connor plurality and Souter concurrence questioned the “breathtakingly broad”146 sweep of the Washington statute. On its face, the statute allows for “any person” to petition for visitation at “any time.”147 The Washington Supreme Court interpreted this language strictly,148 seemingly construing it to mean that complete strangers could petition to visit children.149 Indeed, Washington’s statute appears to be the only statute with such broad language.150 Five Justices noted this breadth and found it troublesome.151

Second, several of the justices’ decisions emphasized the level of deference accorded to a parental decision. The O’Connor plurality opined that the record in Troxel demonstrated that the trial court merely substituted its judgment for the judgment of an otherwise fit parent.152 In fact, the trial court appeared to apply a presumption in favor of nonparental visitation.153 Citing Parham v. J.R.,154 the U.S. Supreme Court has decreed the use of the “Marks rule,” derived from Marks v. United States, 430 U.S. 188 (1977). Under the Marks rule, when a fragmented Court decides a case and no single rationale explaining the result enjoys the assent of five Justices, “the holding of the Court may be viewed as that position taken by those Members who concurred in the judgments on the narrowest grounds.” Marks, 430 U.S. at 193 (quoting Gregg v. Georgia, 428 U.S. 153, 169 n.15 (1976) (opinion of Stewart, Powell and Stevens, JJ.)). Washington courts have adopted the Marks rule. See State v. Zakel, 61 Wash. App. 805, 808, 812 P.2d 512, 514 (1991) (citing Marks, 430 U.S. at 193).

145. When lower courts face fractured decisions of the U.S. Supreme Court, the Court has decreed the use of the “Marks rule,” derived from Marks v. United States, 430 U.S. 188 (1977). Under the Marks rule, when a fragmented Court decides a case and no single rationale explaining the result enjoys the assent of five Justices, “the holding of the Court may be viewed as that position taken by those Members who concurred in the judgments on the narrowest grounds.” Marks, 430 U.S. at 193 (quoting Gregg v. Georgia, 428 U.S. 153, 169 n.15 (1976) (opinion of Stewart, Powell and Stevens, JJ.)). Washington courts have adopted the Marks rule. See State v. Zakel, 61 Wash. App. 805, 808, 812 P.2d 512, 514 (1991) (citing Marks, 430 U.S. at 193).

146. Troxel, 120 S. Ct. at 2061.

147. Id.

148. See id.

149. See id.


151. Troxel, 120 S. Ct. at 2061 (opinion of O’Connor, J.), 2066 (Souter, J., concurring).

152. Id. at 2062.

153. Id.
Nonparental Visitation Rights

Court stated that so long as a parent is fit, there is ordinarily no reason for the state to inject itself into the private realm of the family.\(^\text{155}\) The plurality noted that the problem was not the fact that the trial court intervened in the parental decision\(^\text{156}\) but that the court granted no deference to the parental decision.\(^\text{157}\) Dissenting Justices Stevens\(^\text{158}\) and Kennedy\(^\text{159}\) shared this concern, bringing the total to six justices.

Nevertheless, the *Troxel* case leaves Washington and other states without explicit guidance as to whether nonparental visitation statutes are constitutional, and if so, what elements they require.\(^\text{160}\) Thus, the Washington Supreme Court may feel constrained to follow its previous decision in *Smith*, or it may re-evaluate that holding based on the guidance (albeit limited) gleaned from the U.S. Supreme Court’s *Troxel* decision.\(^\text{161}\)

III. THE WASHINGTON LEGISLATURE SHOULD AMEND ITS NONPARENTAL VISITATION STATUTE

To ensure that Washington courts fulfill the Legislature’s intent, and, more important, to guarantee that courts apply the statute constitutionally, the Legislature should amend Washington’s nonparental visitation statute. To address the concerns that the U.S. Supreme Court voiced in *Troxel v. Granville*,\(^\text{162}\) the amendments should (1) specifically limit the class of persons permitted to petition for nonparental visitation to relatives and de facto parents, (2) codify a rebuttable presumption that a parental decision regarding visitation is in the best interest of the child, etc.

\(^{154}\) 442 U.S. 584 (1979).
\(^{155}\) *Troxel*, 120 S. Ct. at 2061.
\(^{156}\) Id. at 2062.
\(^{157}\) Id.
\(^{158}\) See id. at 2073 (Stevens, J., dissenting).
\(^{159}\) See id. at 2075–76 (Stevens, J., dissenting).

\(^{161}\) Because this Comment argues that the Legislature should amend the statute, it need neither address nor resolve this issue. However, it should be noted that every state that has considered nonparental visitation in light of *Troxel* has upheld the visitation statute as constitutional. See Jackson v. Tangreen, 18 P.3d 100 (Ariz. Ct. App. 2000); Lopez v. Martinez, 102 Cal. Rptr. 2d 71 (Cal. Ct. App. 2000); Rideout v. Riendeau, 761 A.2d 291 (Me. 2000); Gestl v. Frederick, 754 A.2d 1087 (Md. 2000); In re G.P.C., 28 S.W.3d 357 (Mo. Ct. App. 2000); Fitzpatrick v. Youngs, 717 N.Y.S.2d 503 (N.Y. Fam. Ct. 2000); Rubano v. DiCenzo, 759 A.2d 959 (R.I. 2000).

\(^{162}\) 120 S. Ct. 2054 (2000).
and (3) articulate a purpose to guide courts in interpreting the amended statute.

The proposed amendments would address the constitutional concerns of the U.S. Supreme Court while recognizing the importance of non-traditional families and protecting the best interests of the child. The rebuttable presumption that a parental decision is in the best interest of the child has already been implicitly approved by the Court. Furthermore, the amendments would properly balance the rights of parents and the \textit{parens patriae} power of the state and ensure that Washington courts would properly apply existing precedent. Should the Washington Legislature adopt the proposed amendments, the resulting statute would be facially constitutional and would provide courts the guidance necessary to apply the statute constitutionally.

A. The Washington Legislature Should Amend the Nonparental Visitation Statute To Comport with Constitutional Requirements

To bring the statute in line with constitutional requirements and to ensure that courts constitutionally apply Washington’s nonparental visitation statute, the Legislature should make three amendments. First, the Legislature should limit the class of persons who may petition for visitation. Second, the Legislature should mandate a rebuttable presumption that a parent’s determination regarding visitation is in the best interest of the child. Finally, the Legislature should add an intent section to give interpretive guidance to the courts.

1. The Amended Statute Should Limit the Class of Persons Who May Petition for Visitation

Both to address the U.S. Supreme Court’s concerns with the “breathtakingly broad” \cite{164} scope of Washington’s statute and to maintain the purpose of the nonparental visitation statute, the Washington Legislature should amend the “any person” language of the statute. It should narrowly define the phrase “any person,” which occurs twice in the statute, \cite{165} as “a relative or de facto parent.”

---

163. \textit{See} \textit{id.} at 2061.
164. \textit{Id.}
The first class of persons who should be allowed to petition for visitation is the child’s relatives. For purposes of the statute, the Legislature should define “relative” as a “grandparent, aunt, uncle, or sibling.” Most nonparental visitation statutes were originally enacted to address visitation by grandparents.166 Indeed, several state statutes address only visitation by grandparent or other relatives.167 This class of persons would include those who make up contemporary extended families, thereby recognizing the importance of family relations.

The second class of people who could seek visitation rights under the amended statute would be de facto parents. The Legislature should define a de facto parent as “any person who, while having no biological or other legal relationship to the child, has nonetheless acted as a parent to the child.”168 This category of people would include stepparents who have not adopted their stepchildren and unmarried partners who have acted as parents, including same-sex parents with no biological or legal relationship with the child. In some circumstances, these de facto parents are legally barred from adoption, either because of their same-sex status169 or because the biological parent is alive and refuses permission.170 These amendments would create two limited classes of persons with the ability to bring visitation actions,171 classes that are not “breathtakingly broad.”

2. The Amended Statute Should Require Trial Courts To Apply a Rebuttable Presumption that a Fit Parent’s Decisions Are in the Best Interest of the Child

The second proposed amendment would require trial courts to presume that a fit parent’s decision regarding visitation is in the best interest of the child. In Smith, the Washington Supreme Court did not

166. See Shandling, supra note 17, at 119.
168. See Recent Case, supra note 2; see also In re B.G. v. San Bernardino County Welfare Dep’t, 523 P.2d 244, 253 n.18 (Cal. 1974).
171. Consistent with current practice, these sections would be severable in order to protect the statute if any part of the statute was found to be unconstitutional. See, e.g., Amalgamated Transit Union Local 587 v. State, 142 Wash. 2d 183, 228, 11 P.3d 762, 791 (2000).
require trial courts to adopt the traditional common law presumption that a fit parent will act in the best interest of his or her child. In order to reverse this interpretation, the Legislature should add a new section to Washington’s nonparental visitation statute: “In any action arising under this section, the trial court shall apply a rebuttable presumption that a parent’s decisions regarding visitation are in the best interest of the child.” The Legislature should further amend the statute to provide guidance to the courts in determining the best interests of the child when the petitioner attempts to rebut the parental determination. The Legislature should provide this guidance by adopting the standards found in Washington’s other nonparental visitation statute. The proposed standards section would read:

The court may consider the following factors when making a determination of the child’s best interests:

1. The strength of the relationship between the child and the petitioner;
2. The relationship between each of the child’s parents or the person with whom the child is residing and the petitioner;
3. The nature and reason for either parent’s objection to granting the petitioner visitation;
4. The effect that granting visitation will have on the relationship between the child and the child’s parents or the person with whom the child is residing;
5. The residential time sharing arrangements between the parents;
6. The good faith of the petitioner;
7. Any criminal history or history of physical, emotional, or sexual abuse or neglect by the petitioner; and

Arguably, in doing so the court may have abrogated its duty to interpret statutes in a constitutional manner and to avoid absurd or strained results. See State ex rel. Evergreen Freedom Found. v. Wash. Educ. Ass’n, 615 Wash. 2d 615, 632, 999 P.2d 602, 612 (2000).


In adopting this section, the Washington State Legislature would be overruling the interpretation the Washington Supreme Court gave to the existing nonparental visitation statute. The Legislature would not be overruling the court’s determination that a specific interpretation of the statute is unconstitutional. The authority of the Legislature to act in this manner is well-accepted. See, e.g., Marine Power & Equip. Co. v. Wash. State Human Rts. Comm’n Hearing Tribunal, 39 Wash. App. 609, 612–17, 694 P.2d 697, 699–700 (1985).

Nonparental Visitation Rights

(h) Any other factor relevant to the child’s best interest. 176

3. The Legislature Should Add an Appropriate Purpose Section to Washington’s Nonparental Visitation Statute To Guide Courts in Interpreting the Statute

The Washington Legislature should also provide guidance to the state courts in interpreting the nonparental visitation statute by adding a purpose section to the statute. A purpose section would provide guidance to courts for interpreting any part of the statute that might be vague. The section should set forth the Legislature’s three underlying policies in enacting the statute: (1) parents have a right to control who visits their children; (2) visitation with other adults is often desirable and beneficial to children; and (3) in order to obtain visitation with a child against a parent’s will, a petitioner must affirmatively demonstrate that it would be in the best interest of the child.

B. The Proposed Amendments Would Address the U.S. Supreme Court’s Concern that the Statute Is “Breathtakingly Broad”

The proposed amendments would meet the constitutional requirements of the Fourteenth Amendment because they address the U.S. Supreme Court’s concerns that the statute is “breathtakingly broad.” 177 The Washington Supreme Court held that the nonparental visitation statute was unambiguous, and therefore read no qualifications into the “any person” language.178 While this treatment was suspect,179 under the rules of review180 the U.S. Supreme Court had to analyze this interpretation.181 Under the proposed amended statute, only two distinct classes of people would be permitted to bring visitation actions: relatives

176. This is identical to the standard found in WASH. REV. CODE § 26.09.240(6).
177. Troxel, 120 S. Ct. at 2061.
179. The court could have read the language in conjunction with the requirement that the visitation be in the best interest of the child, or in the greater context of child custody and visitation to limit the class of persons with standing to petition for visitation. This would have been consistent with the court’s duty to avoid absurd or strained consequences. See State ex rel. Evergreen Freedom Found. v. Wash. Educ. Ass’n, 615 Wash. 2d 615, 632, 999 P.2d 602, 612 (2000); see also supra Part II.B.
180. See infra Part III.C.
181. See Troxel, 120 S. Ct. at 2065.
and de facto parents. The amended statute would necessarily exclude a
stranger from petitioning for visitation.182

These classes would be in line with the cases that indicate that child
visitation is a family right.183 The U.S. Supreme Court has long rejected
the notion that a family is constitutionally defined as being only
comprised of parents and their children.184 Such cases suggest that the
Fourteenth Amendment does not favor parental claims over those of
other family members.185 This distinction is particularly significant in
the majority of nonparental visitation cases, which involve either relatives or
de facto parents, both of whom courts could consider family members.186
The U.S. Supreme Court, in other contexts, has extended family
protections to grandparents,187 aunts and nieces,188 and de facto parents.189
The proposed statute would narrow the class of people permitted to
petition for visitation consistent with the fact that a parent’s right with
respect to total strangers is not identical to the interest with respect to
others such as de facto parents.190

Furthermore, the proposed amendment would address Justice
Kennedy’s concern that a traditional family with two or even one
permanent and caring parent is not a reality for many children.191 It also
addresses Justice Stevens’s concern with protecting the child’s interest in
preserving relationships that serve the child’s welfare and protection.192
De facto parents have an important connection to their de facto children
and may have a significant positive impact on these children’s lives.193
Thus, the proposed amendment would allow for the protection of the
interests of the child in maintaining significant relationships194 and would

182. There are no reported cases in which a stranger has petitioned for visitation with a child.
185. See Toni Eddy, Comment, Grandparent Visitation Rights in Ohio When the Family is Intact,
186. Cases like Quilloin, while weakening protection given to biological parents, arguably
strengthen the protection given to the extended family. See Bean, supra note 20, at 417–18.
187. Moore, 431 U.S. at 505–06.
191. Id. (Kennedy, J., dissenting).
192. Id. at 2071 (Stevens, J., dissenting).
193. See GOLSTEIN ET AL., supra note 4, at 104–09.
194. See Troxel, 120 S. Ct. at 2071 (Stevens, J., dissenting).
Nonparental Visitation Rights

encompass a broad concept of family. Yet, by limiting the classes to those implicitly approved by U.S. Supreme Court opinions, the amendment would satisfy the “breathtakingly broad” concerns voiced by the O’Connor plurality.

C. The Proposed Amendments Would Address the U.S. Supreme Court’s Concern Regarding Deference to Parents by Codifying the Existing Common Law Presumption in Favor of Parental Decisions

The U.S. Supreme Court has previously recognized a presumption that a fit parent will act in the best interest of his or her child. The trial court’s failure to adopt this common law presumption (and the Washington Supreme Court’s failure to require the presumption) was one of the two major concerns of the justices in Troxel. The proposed amendment would require courts to apply a rebuttable presumption that the decisions of the parents are in the best interest of the child. This amendment would also bring Washington in line with the majority of other states’ nonparental visitation statutes.

After reviewing the record, the O’Connor plurality determined that the trial court merely substituted its determination of the best interest of the children without giving any deference to the parental decision. According to the O’Connor plurality, this decision directly contravened the traditional presumption that parents act in the best interest of their children. As the Court explained in Parham, it has been historically recognized that “natural bonds of affection lead parents to act in the best interests of their children.” The proposed amendment would codify the common law presumption, removing the potential for trial courts to

195. See id. at 2059.
196. Id. at 2061.
198. See Troxel, 120 S. Ct. at 2061–62.
199. See, e.g., CAL. FAM. CODE § 3104(e) (West 1994) (codifying rebuttable presumption that grandparent visitation is not in child’s best interest if parents agree that visitation should not be granted); ME. REV. STAT. ANN. tit. 19A, § 1803(3) (West 1998) (requiring that visitation not significantly interfere with parent-child relationship or with parent’s authority over child); R.I. GEN. LAWS § 15-5-24.3(a)(2)(V) (Supp. 1999) (codifying rebuttable presumption that parent’s decision to refuse visitation was reasonable).
200. Troxel, 120 S. Ct. at 2062.
201. Id.; see also, e.g., Parham, 442 U.S. at 602.
interpret the statute in the manner adopted by the Smith court. The amendment would obviate Justice Souter’s opinion that the Washington Supreme Court’s controlling interpretation rejecting deference to parental decisions rendered the statute unconstitutional.\(^{203}\) By employing a rebuttable presumption, the amendment would allow courts to constitutionally award visitation to foster parents and other caregivers, as advocated by Justice Stevens.\(^{204}\)

D. The Proposed Amendments Would Ensure Courts Will Apply Washington’s Nonparental Visitation Statute Constitutionally by Properly Balancing the State’s Parens Patriae Power with the Constitutional Rights of Parents

The Legislature should amend the nonparental visitation statute to avoid the Washington Supreme Court’s poor reasoning in Smith and provide clear guidance to lower courts applying the statute to balance constitutionally the state’s parens patriae power and parental rights. Prior to Smith, in assessing the constitutionality of an infringement on a parent’s rights, Washington courts sought the proper balance between the parent’s constitutional rights and the state’s parens patriae interest in protecting the best interests of the child.\(^{205}\) Key to this balancing test was the degree of the abridgement of parental rights at issue.\(^{206}\) The Washington Supreme Court had previously upheld the use of the state’s parens patriae power without requiring a showing of harm to the child in In re Sumey, In re Key, and State v. Steinbach.\(^{207}\) However, in Smith, the court rejected the Sumey balancing test and instead relied on U.S. Supreme Court precedent to hold that the nonparental visitation statute was facially unconstitutional because it relied on the “best-interest-of-the-child” standard rather than the “harm-to-the-child” standard.\(^{208}\) The Legislature should amend the nonparental visitation statute because the Smith court misapplied its own precedent and misinterpreted U.S. Supreme Court precedent in reaching its conclusion.

\(^{203}\) Id. at 2065–66 (Souter, J., concurring).

\(^{204}\) Id. at 2070 (Stevens, J., dissenting).


\(^{206}\) See id.

\(^{207}\) 101 Wash. 2d 460, 679 P.2d 369 (1984). For a discussion of In re Sumey and In re Key, see supra Part I.B.

Nonparental Visitation Rights

1. Washington Courts Consistently Utilized the Sumey Balancing Test Before Smith

The Legislature should amend the nonparental visitation statute because the Washington Supreme Court’s position in *Smith*—absent a threshold finding of parental unfitness or harm to the child, the state, as *parens patriae* acting on the child’s behalf, may not intrude on parental rights, no matter how slightly—cannot be reconciled with existing case law. The proposed amendments would reverse the Washington Supreme Court’s insistence in *Smith* that a showing of harm to the child or parental unfitness is required before the state may exercise its *parens patriae* power and restore the balancing test of *Sumey*. Without this amendment, lower courts will have little guidance when determining how to apply constitutionally the statute in the wake of *Smith* and *Troxel*.

In *Sumey*, the Washington Supreme Court contrasted temporary child placement with termination of parental rights. The *Sumey* court upheld the placement of a child outside the home against the parent’s wishes, reasoning that the degree of intrusion on the parent’s rights was relatively minor because the parents retain custody over the child. The substantial interests of the state and child were held sufficient to justify the limited infringement on the parent’s rights. *Key* involved a dependency proceeding, which the court held to be a more serious interference with the parent-child relationship. Yet the *Key* court also followed the *Sumey* test. In applying the *Sumey* best-interest-of-the-child test, the court explicitly rejected a requirement that there be a finding of parental unfitness or “harm to the child” in order to bring the state’s *parens patriae* power to bear.

Occasional, temporary visitation by grandparents and others, allowed only if it is in the best interest of the child, is a relatively minor interference with parental rights. Such visitation is certainly less intrusive than a dependency proceeding. Yet in *Smith*, the court required a showing of harm prior to awarding visitation, while *Key* rejected such a

209. See id.
211. See id.
212. See id.
214. Id. at 610–11, 836 P.2d at 205–06.
215. Id.
requirement in a dependency proceeding. The ruling in Smith returns the
court to the common law paradox where it was easier to terminate rights
than it was to obtain visitation rights.\(^{216}\) Thus, the amendment would
restore the Sumey test to balance parents’ constitutional rights and states’
parens patriae power by reaffirming the best-interest-of-the-child
standard. The best-interest standard is the touchstone by which all other
rights are tested and concerns addressed in various contexts dealing with
children.\(^{217}\) The decision by the Smith court calls into question all statutes
relying on the best-interest-of-the-child standard.\(^{218}\) If this result were
allowed to stand, it would radically alter the long-standing basis by
which child welfare cases are determined and overturn what was well-
settled law.\(^{219}\) As Justice Stevens pointed out, child welfare cases do not
present simply a bipolar struggle between parents and the state: they also
implicate the interests of the child in question.\(^{220}\)

\[2. \quad \text{The Smith Court Misapplied U.S. Supreme Court Precedent in}
\text{Striking Down the Best-Interest-of-the-Child Standard}\]

The Legislature should amend the nonparental visitation statute to
restore the best-interest-of-the-child balancing test of Sumey, correct the

\(^{216}\) See supra Part I.C.1.

\(^{217}\) See, e.g., Wash. State Coalition for the Homeless v. Dep’t of Soc. & Health Servs., 133
Wash. 2d 894, 923, 949 P.2d 1291, 1306 (1997) (“As in all matters dealing with the welfare of
children, the court must additionally act in the best interests of the child.”); In re Marriage
parental cooperation during post-separation action, trial court is given broad discretion to develop
and order parenting plan according to guidelines set forth in Wash. Rev. Code § 26.09.187(3)(a)
(2000) and based on best interests of children at time of trial); In re Aschauer, 93 Wash. 2d 689, 695,
611 P.2d 1245, 1249 (1978) (“This court has repeatedly said that the goal of a dependency hearing is
to determine the welfare of the child and his best interests.”); In re Sego, 82 Wash. 2d 736, 738, 513
P.2d 831, 832 (1973) (“[A] child’s welfare is the court’s primary consideration . . . when the rights
of parents and the welfare of their children are in conflict, the welfare of the minor children must
(1997) (noting that best interests of child are paramount in paternity proceedings; and upholding trial
court’s denial of putative father’s attempt to reopen and challenge paternity determination made
thirteen years prior).

\(^{218}\) Troxel v. Granville, 120 S. Ct. 2054, 2068 n.5 (2000) (Stevens, J., dissenting) (noting ten
Washington statutes invoking best-interest standard and 698 judicial opinions referencing best-
interest standard).

\(^{219}\) Every state save Georgia predicates nonparental visitation on the best interest of the child
standard. Troxel, 120 S. Ct. at 2078 (Kennedy, J., dissenting) (citing Ga. Code Ann. § 19-7-3(c)
(1999)).

\(^{220}\) Id. at 2071 (Stevens, J. dissenting).
Smith court’s misapplication of U.S. Supreme Court precedent, and ensure that Washington courts constitutionally apply the nonparental visitation statute. The line of U.S. Supreme Court cases relied on in Smith does not purport to eliminate the state’s parens patriae power or the best-interest-of-the-child standard. Instead these cases involve substantial, rather than relatively minor, infringements on parental rights or other fundamental rights.  

The restoration of the Sumey best-interest-of-the-child balancing test comports with parents’ Fourteenth Amendment liberty rights in raising their children. The Smith court cited Wisconsin v. Yoder for the proposition that “the Supreme Court cases which support the constitutional right to rear one’s child and the right to family privacy indicate that the state may interfere only if it appears that parental decisions will jeopardize the health or safety of the child, or have a potential for significant social burdens.” But this characterization goes too far. Yoder turns on the free exercise claim asserted by the child’s Amish parents and the unique facts of that case. The Court held where the interests of parents were combined with a free exercise of religion claim of the nature present in that case, the state must show a compelling interest in requiring Amish parents, contrary to their religious beliefs, to send their children to school beyond the eighth grade. However, the Yoder Court did not hold that harm is a threshold requirement for any encroachment on parental rights, as the Smith court implied in its rejection of the best-interest-of-the-child standard.


222. Smith v. Stillwell-Smith, 137 Wash. 2d 1, 17, 969 P.2d 21, 27 aff’d on other grounds sub nom. Troxel v. Granville, 120 S. Ct. 2054 (2000) (internal quotation marks omitted). In fact, the cited case does not so hold. In Yoder, the Supreme Court held that the Free Exercise Clause of the First Amendment barred the application of compulsory school attendance law to Old Order Amish who did not send their children to school after the eighth grade because “only those interests of the highest order and those not otherwise served can overbalance legitimate claims to the free exercise of religion.” Yoder, 406 U.S. at 215.

223. See Yoder, 406 U.S. at 233–34.

224. See id.

225. See id.

226. See Smith, 137 Wash. 2d at 28, 969 P.2d at 29.
The *Smith* court further erred in concluding that in *Prince v. Massachusetts* the U.S. Supreme Court required that some harm threaten the child’s welfare before the state could constitutionally interfere with a parent’s rights by exercising its *parens patriae* power. 227 Although *Prince* indicates state intervention in areas of religious practices or parental control is appropriate to prevent harm to a child, that case does not suggest harm to a child is a threshold requirement for any and all types of state encroachment of parental rights. 228 Furthermore, *Prince* was specifically limited to the facts of that case. 229 Therefore, by amending the statute to include the *Sumey* best-interest-of-the-child standard, the Washington Legislature will ensure consistent application of both Washington Supreme Court and U.S. Supreme Court *parens patriae* and Fourteenth Amendment precedent.


As the purpose section would indicate, 230 the proposed amendments would also recognize the changing nature of modern families. They would recognize that legal or biological relationships do not always address the reality of modern family life and they would protect the interests of those legally prohibited from establishing a legal relationship with a child by permitting visitation outside of the traditional, strict parental visitation rules.

The amended statute would further give flexibility to the trial courts in fashioning visitation orders in cases presenting unique family structures. Furthermore, the statute would provide needed guidance to the trial courts because its language is familiar to trial courts. Because courts have successfully applied these guidelines to nonparental visitation during and after divorce proceedings, 231 they could easily be adapted to other nonparental visitation cases. The best-interest standard would retain the flexibility required by trial courts in addressing the infinite circumstances and possibilities that surround child welfare determi-

227. See id.
229. Id. at 170–71.
230. See supra Part III.A.3.
231. See supra Part I.C.2.
nations such as the nonparental visitation at issue in these cases. The statute would also work to protect the relationships children develop with adults who play significant roles in the lives of these children. The state has a substantial interest in protecting children’s significant relationships with adults, promoting healthy family relationships, and reinforcing the positive influences of significant relationships with relatives and de facto parents.

IV. CONCLUSION

The nature of the family unit often includes adults other than biological or legal parents who provide significant positive influences on children. These benefits include stronger family bonds, moral instruction, opportunities to observe healthy adult relationships, and additional educational opportunities. Under ordinary circumstances, a child’s parents are best situated to determine the advisability of visitation with their minor children by other adults. However, in some circumstances the best interests of the child are better served by allowing visitation with certain adults against the wishes of the child’s parents.

States have recognized this reality by enacting nonparental visitation statutes. By their very nature, these statutes represent a conflict between a parent’s constitutional rights to raise a child as he or she sees fit and the state’s parens patriae power. In Troxel, the U.S. Supreme Court suggested that the Washington Legislature and Washington Supreme Court failed to appropriately resolve this conflict in the statute and its application. To address the U.S. Supreme Court’s concerns and ensure that courts constitutionally apply Washington’s nonparental visitation statute, the Legislature should make two amendments to the statute and add a purpose section. The Legislature should both limit the class of persons permitted to petition for visitation to relatives and de facto parents, and codify a rebuttable presumption that a parent’s decision regarding visitation is in the best interest of the child. In addition to

232 See Beagle v. Beagle, 678 So. 2d 1271, 1275 (Fla. 1996). While finding the Florida grandparent visitation statute violated the state constitution, the Florida Supreme Court acknowledged other states had upheld their grandparent visitation statutes against federal constitutional challenges, noting: “[i]n those cases a best interest standard was deemed to be sufficient.” Id. (citing Lehrer v. Davis, 571 A.2d 691 (Conn. 1990); Bailey v. Menzie, 542 N.E.2d 1015 (Ind. Ct. App. 1989); Spradling v. Harris, 778 P.2d 365 (Kan. Ct. App. 1989); King v. King, 828 S.W.2d 630 (Ky. 1992); Herndon v. Tuhey, 857 S.W.2d 203 (Mo. 1993); Ridenour v. Ridenour, 901 P.2d 770 (N.M. Ct. App. 1995)).
addressing the concerns of the U.S. Supreme Court, these amendments would ensure that Washington courts properly apply Washington Supreme Court and U.S. Supreme Court precedent. Parents have rights that must be respected. But children have a compelling interest in developing relationships with other caring adults. The state should act to protect all of these interests.
TWO WRONGS DO NOT MAKE A DEFENSE: ELIMINATING THE EQUAL-OPPORTUNITY-HARASSER DEFENSE

Shylah Miles

Abstract: Sexual harassment is a prevalent problem in the American workplace that accounts for nearly sixty-four percent of all gender discrimination claims under Title VII. The equal-opportunity-harasser defense allows harassers who target both males and females to escape liability. Courts have allowed the defense because they have interpreted the “because of sex” element of a sexual harassment claim to require disparate treatment or a showing that the plaintiffs would not have been harassed if they were members of the opposite sex. An equal-opportunity harasser harasses both sexes and, therefore, plaintiffs cannot prove disparate treatment. This Comment argues that the disparate-treatment requirement does not fit the sexual harassment model because it is a class-based analysis and sexual harassment is an individual-based discrimination. By limiting analysis of equal-opportunity-harasser claims to disparate treatment, courts allow sexual inequality in the workplace to continue, undermining the purpose behind sexual harassment laws. To rectify this situation, the courts should adopt an individual analysis of the “because of sex” element, which a recent U.S. Supreme Court decision allows. Adoption of an individual analysis will follow current trends in federal and Washington state courts to limit applicability of the defense and will enable courts to deny the equal-opportunity-harasser defense.

“We must be aware that whatever rules we develop will require constant reexamination, modification, and fine-tuning. The sexual, gender, and preference revolutions are not over. Society is in a constant state of evolution.”1

Steven and Karen worked in the maintenance department at the Indiana Department of Transportation.2 Their male supervisor, Gale, aimed his sexually charged actions and explicit language toward both Steven and Karen on separate occasions.3 Specifically, Gale touched Karen’s body, stood too close to her, asked her to sleep with him, and made sexist comments.4 Gale also sexually propositioned Steven while

---

2. Holman v. Indiana, 211 F.3d 399, 401 (7th Cir. 2000). The descriptions in the text are the actual facts of Holman. That court decided that the equal-opportunity-harasser defense protected the employer from liability in a sexual harassment claim. Id.
3. Id.
4. Id.
grabbing his head. Disturbed by their supervisor’s conduct, both sought legal relief through a sexual harassment claim against their employer.

After litigation, the two victims learned that Title VII harassment law offered them no protection from this harasser. Even though the law protects employees against same-sex and opposite-sex sexual harassment, federal courts have held that the harassment they were exposed to was not of either of these types. Instead, they were being harassed by an equal-opportunity harasser, whose status protects the employer from liability. Courts reason that when a supervisor harasses both a man and a woman, there is no disparate treatment and therefore plaintiffs cannot satisfy the element of a sexual harassment claim that requires proof that the harassment was “because of sex.” This legal gap in the protection of victims against sexual harassment exists in Washington law as well as in most federal circuits.

As new fact scenarios develop in sexual harassment law, law aimed to protect victims from sexual harassment must evolve in order to continue to protect all individuals from sexual harassment in the workplace. This Comment advocates that courts dismantle the equal-opportunity-harasser defense by adopting an individual analysis of the “because of sex”

5. Id.
6. Id.
7. Id. at 403.
9. Infra note 94.
10. This Comment applies equally to both the bisexual and equal-opportunity-harasser defenses. Bisexual harassment occurs when a supervisor extorts or attempts to extort sexual favors or tolerance of sexual harassment from a male and a female employee in exchange for a job, a job benefit, or protection from adverse employment actions. Barnes v. Costle, 561 F.2d 983, 990 n.55 (D.C. Cir. 1977). Another form of bisexual harassment occurs where the harasser is bisexual but does not harass members of both sexes. See Rycek v. Guest Servs., Inc., 877 F. Supp. 754, 761 n.6 (D.D.C. 1995). Where a supervisor creates a work environment that unreasonably interferes with a male and a female’s work performance or creates an intimidating, hostile, or offensive work environment for employees of both sexes, the supervisor is referred to as an equal-opportunity harasser. See Chiapuzio v. BLT Operating Corp., 826 F. Supp. 1334, 1336–37 (D. Wyo. 1993). An equal-opportunity harasser might not actually be bisexual. See id.
11. Holman, 211 F.3d at 403.
12. Id.
14. Although this Comment refers to evidence of harassment of both sexes as the equal-opportunity-harasser defense, it is not an affirmative defense. Rather, it is an evidentiary showing that prohibits satisfaction of the “because of sex” element of a sexual harassment claim. Reference to
The Equal-Opportunity-Harasser Defense

element and by disregarding the harasser’s treatment of others. Part I discusses sexual harassment, the incidence of claims, and the effects of sexual harassment in the workplace. Part II explains the status of sexual harassment law generally in the federal courts and in Washington. Part III examines how federal and Washington courts use the equal-opportunity-harasser defense. Finally, Part IV argues that evidence of disparate treatment should be replaced by an individual analysis of the “because of sex” element so that the defense may be denied.

I. THE PROBLEMS AND PERVASIVENESS OF SEXUAL HARASSMENT IN THE WORKPLACE

Sexual harassment is a pervasive problem that erects and maintains barriers to sexual equality in the workplace. Sexual harassment occurs when “submission to . . . [unwelcome sexual] conduct is made either explicitly or implicitly a term or condition of an individual’s employment.”15 In 1999, sexual harassment claims accounted for 64% of all gender discrimination claims under Title VII and the proportion filed by male employees increased from 9% in 1992 to 12% in 1997–1999.16 Approximately 50% to 85% of American females will experience some form of sexual harassment during their academic or working lives.17

Sexual harassment hinders equality of the sexes because it supports a sexual hierarchy, punishing those who do not conform to sexual stereotypes.18 A harasser may view individuals as stereotypically male or female.19 Where individuals do not conform to the harasser’s preconceptions, the harasser singles out the victim based on the victim’s

---

16. EEOC, Enforcement Statistics, at http://www.eeoc.gov/statsl (last visited Jan. 24, 2001). There are no comprehensive statistics for the total number of all workplace harassment complaints, formal and informal, because there is no central repository for the reporting of complaints that are resolved before going to the agency or judicial stage. Id.
divergence from scripted gender norms. The harasser’s “sexual misconduct” may serve to block professional advancement for men and women because such misconduct may interfere with the individual’s work performance, may create a hostile work environment, and can be directly linked to denial of job benefits. Sexual harassment qualifies as sex discrimination because individuals are treated differently because of their sex. Both Washington and federal courts have sought to prohibit sexual harassment so that these barriers to sexual equality can be eradicated.

One recent development in sexual harassment law is the emergence of equal-opportunity harassers. An equal-opportunity harasser harasses both men and women. Under the traditional analysis of the “because of sex” element of a sexual harassment claim, plaintiffs must show that if they were of the opposite sex, they would not have been harassed. Consequently, when the perpetrator harasses both men and women, this element is impossible to satisfy, leaving plaintiffs without protection under sexual harassment law. Accordingly, the status of an equal-opportunity harasser has become a defense to liability. Courts have been frustrated by this problem as they try to allow claims against equal-opportunity harassers yet still comply with the disparate-treatment requirement.

II. FEDERAL AND WASHINGTON STATE LAW REGARDING SEXUAL HARASSMENT CLAIMS

25. See infra notes 59–64 and accompanying text.
26. Glasgow, 103 Wash. 2d at 406, 693 P.2d at 712.
27. Holman v. Indiana, 211 F.3d 399, 403 (7th Cir. 2000).
28. Id.
The Equal-Opportunity-Harasser Defense

...sexual harassment claim and the evidence to satisfy them in like fashion. The only substantial difference is that, by statute, Washington explicitly states that courts are to construe the sexual harassment statute liberally in order to fulfill its purpose, while federal law does not.

A. Federal Law Against Sexual Harassment

Title VII of the 1964 Civil Rights Act\(^\text{29}\) prohibits sex discrimination, which courts have interpreted to include sexual harassment. By making sexual harassment actionable as sex discrimination, courts have required that sexual harassment claims satisfy the elements of a sex discrimination claim.\(^\text{30}\) Most important to the analysis of the equal-opportunity harasser defense is the “because of sex” element. Although courts routinely require evidence of disparate treatment to satisfy the “because of sex” element, a recent U.S. Supreme Court decision has provided plaintiffs with more flexibility in meeting this element.

1. Title VII Has Been Interpreted To Prohibit Sexual Harassment

In 1964, Congress declared that sex discrimination in employment was prohibited, but it failed either to define “sex” or list included offenses. Under Title VII of the 1964 Civil Rights Act, Congress made it unlawful for an employer “to fail or refuse to hire or to discharge any individual, or otherwise to discriminate against any individual with respect to his compensation, terms, conditions, or privileges of employment, because of such individual’s race, color, religion, sex, or national origin.”\(^\text{31}\) The prohibition against discrimination based on sex was a last-minute amendment in the House of Representatives in an effort to rouse opposition to the bill.\(^\text{32}\) Despite this effort, the bill passed as amended, with little debate on the substantive merits of prohibiting sex discrimination in the workplace.\(^\text{33}\) The Senate held no debate on the


\(^{30}\) See infra notes 59–64 and accompanying text.


\(^{32}\) See 110 CONG. REC. 2577–84 (1964).

\(^{33}\) Id.
issue, either. The lack of debate left very little legislative history to guide courts in interpreting what Congress intended sex discrimination to include.

Although the language of Title VII failed to specifically address sexual harassment, case law has established that Title VII proscribes such conduct. In *Meritor Savings Bank v. Vinson*, the U.S. Supreme Court held that sexual harassment violates Title VII. The Court reasoned that when a supervisor harasses an employee because of the employee’s sex, that supervisor discriminates on the basis of sex, and this discrimination is “every bit the arbitrary barrier to sexual equality at the workplace that racial harassment is to racial equality.”

Following the *Meritor* decision, federal courts have interpreted the term “sex” to refer to gender and have recognized two different forms of sexual harassment: hostile work environment and quid pro quo. Hostile work environment harassment occurs when offensive and unwelcome conduct is sufficiently severe to create an abusive or hostile work environment. Quid pro quo harassment is the extortion or attempted extortion of sexual favors in exchange for a job, a job benefit, or protection from adverse employment actions.

To establish a prima facie case of hostile work environment sexual harassment, the plaintiff must establish five elements. Federal courts require the plaintiff to prove that: (1) the employee is a member of a

37. Id. at 73.
38. Id. at 67 (quoting Henson v. City of Dundee, 682 F.2d 897, 902 (11th Cir. 1982)); see Katz v. Dole, 709 F.2d 251, 254–55 (4th Cir. 1983).
39. DeCintio v. Westchester County Med. Ctr., 807 F.2d 304, 307 (2d Cir. 1986); Ulame v. E. Airlines, Inc., 742 F.2d 1081, 1085 (7th Cir. 1984); DeSantis v. Pac. Tel. & Tel. Co., 608 F.2d 327, 329 (9th Cir. 1979). It thus follows that sex refers to male or female and not sexual preference.
40. See, e.g., Ellison v. Brady, 924 F.2d 872, 875 (9th Cir. 1991).
41. See, e.g., Henson v. City of Dundee, 682 F.2d 897, 910 (11th Cir. 1982). This Comment is concerned with both types of sexual harassment because a bisexual harasser exists in both quid pro quo and hostile work environment contexts.
42. 29 C.F.R. § 1604.11(a)(3) (2000).
43. Henson, 682 F.2d at 910.
protected group; (2) the employee was subjected to unwelcome harassment; (3) the harassment was based upon sex; (4) the harassment affected a term, condition, or privilege of employment and was sufficiently severe and pervasive so as to alter the conditions of employment and create an intimidating, hostile, or offensive working environment; 44 and (5) the existence of respondeat superior liability. 45 With some variation, federal courts widely follow this formulation of the hostile work environment harassment case. 46

The elements of quid pro quo harassment are nearly identical to hostile work environment harassment. Instead of a “severe and pervasive” element, quid pro quo has an element that requires receipt of an employment benefit or protection from a job detriment in exchange for the harassment by the supervisor. 47 Both types of claims require that harassment be “because of sex.”

2. The “Because of Sex” Element

a. How the “Because of Sex” Element Is Traditionally Satisfied

Courts traditionally require plaintiffs to prove the “because of sex” element with evidence of disparate treatment. 48 Thus, plaintiffs must show that they would not have been harassed if the plaintiff had been of the opposite sex. 49 The disparate treatment need not be sexual in nature to constitute sexual harassment, 50 but the harassment must be motivated by a gender-based animus. 51 Conversely, the “because of sex” element is not automatically fulfilled just because the harasser’s words have “sexual content or connotations.” 52 Courts may have created this limitation

44. This is referred to as the “severe and pervasive” element. Meritor Sav. Bank v. Vinson, 477 U.S. 57, 63–73 (1986). This element ensures that the actions by the alleged harasser substantially affected the conditions of employment and that the actions were not ordinary socializing. Oncale v. Sundowner Offshore Servs., Inc., 523 U.S. 75, 81 (1998).

45. Meritor, 477 U.S. at 63–73; 29 C.F.R. § 1604.11.


47. Henson, 682 F.2d at 909–10.

48. Id. at 904.

49. Id.


51. Id.

52. Id.
because sexually explicit conduct does not necessarily intimate a motivation to harass because of sex and, instead, could be linked to general dislike for the individual or juvenile provocation.\(^5^3\) Therefore, courts allow room to reject a sexual harassment claim where harassment is motivated by these types of considerations.\(^5^4\)

b. Why Courts Use Disparate Treatment To Satisfy the “Because of Sex” Element

Courts have traditionally used disparate treatment, a class-based analysis, to satisfy the “because of sex” element instead of an individual-based analysis. Class-based discrimination occurs when a perpetrator chooses a victim based on his or her membership in a targeted class.\(^5^5\) For example, a class-based discriminator can be presumed to discriminate against all members of a class (e.g., females).\(^5^6\) In contrast, individual discrimination occurs when a perpetrator selects a victim based on factors in addition to membership in a targeted class.\(^5^7\) An individual discriminator may not discriminate against an entire targeted class and instead may choose only one individual based on both the victim’s membership in that class and individual factors such as personality and physical characteristics.\(^5^8\)

When the U.S. Supreme Court expanded Title VII to include sexual harassment, it borrowed the elements from traditional sex discrimination claims and thereby established a class-based analysis of sexual harassment.\(^5^9\) Traditional sex discrimination occurs when a person is treated differently than other similarly situated individuals solely because of his or her sex.\(^6^0\) This is a class-based discrimination because a

---


\(^{54}\) See Doe, 85 Wash. App. at 149–50, 931 P.2d at 200.


\(^{56}\) Id.

\(^{57}\) Id.

\(^{58}\) See id.


\(^{60}\) Butler v. N.Y. State Dep’t of Law, 211 F.3d 739, 746 (2d Cir. 2000). Sex discrimination is a broader cause of action than sexual harassment because any treatment of the victim because of the
perpetrator will generally select victims based solely on their sex and can be assumed to discriminate against all members of the victims’ sex. With this in mind, courts fashioned a class-based analysis of the “because of sex” element, which required evidence of disparate treatment of the sexes. Disparate treatment requires plaintiffs to prove that, but for their sex, they would not have been discriminated against. When courts crafted elements of a sexual harassment claim, they read this element into the text of Title VII, which bans discrimination “because of sex.” Until confronted with a same-sex harassment scenario, courts relied solely on evidence of the class-based disparate-treatment requirement to satisfy the “because of sex” element.

c. Oncale v. Sundowner Offshore Services, Inc.: The U.S. Supreme Court Presents Alternatives for Satisfying the “Because of Sex” Element

The U.S. Supreme Court modified evidentiary options for satisfying the “because of sex” element when it was confronted with a fact scenario that could not fit in the traditional class-based discrimination analysis. In Oncale v. Sundowner Offshore Services, Inc., the Court was faced with a same-sex harassment action. The plaintiff worked in an all-male environment, which made disparate treatment impossible to show. The plaintiff charged that his two supervisors had “assaulted him in a sexual manner” and threatened to rape him. In determining that same-sex harassment was actionable under Title VII, the Court identified three types of evidence that might satisfy the “because of sex” element: (1) comparative evidence, (2) gender-specific conduct or actions, or (3) explicit or implicit proposals of sexual activity.

61. Locke, supra note 59, at 393–95; Paul, supra note 55, at 352.
62. Locke, supra note 59, at 393–95; Paul, supra note 55, at 352.
63. Locke, supra note 59, at 393–95.
64. Id. at 393.
66. Id. at 76.
67. Id. at 77.
68. Id.
69. Id. at 79–81.
Prior to *Oncale*, plaintiffs could only satisfy this element through proof of disparate treatment.70 *Oncale* now provides plaintiffs with three evidentiary routes in which a fact-finder may draw reasonable inferences that the harassment was because of the plaintiff’s sex.71 Commentators have recognized that *Oncale* allows greater flexibility in establishing sexual harassment claims.72

B. Washington State Law Against Sexual Harassment

The Washington Law Against Discrimination (WLAD)73 is similar to Title VII. Accordingly, Washington courts follow the reasoning of the federal courts when addressing sexual harassment claims. This reasoning includes requiring evidence of disparate treatment to satisfy the “because of sex” element.

In 1949, Washington State enacted the WLAD,74 which recognized as a civil right “the opportunity to obtain employment without discrimination because of race, creed, color or national origin.”75 The WLAD made it an unfair labor practice for employers to “discriminate against any person in compensation or in other terms or conditions of employment because of such person’s race, creed, color or national origin.”76 In 1971, the Legislature expanded the WLAD to protect all inhabitants of the state from sex discrimination and to eliminate and prevent such discrimination in employment.77

Washington courts have found that sexual harassment in the workplace violates the WLAD.78 Because the WLAD is similar in purpose and subject matter to Title VII discrimination claims,
Washington courts view cases interpreting the federal law as persuasive when construing the Washington statute.\textsuperscript{79} Like the federal courts, Washington courts divide sexual harassment into hostile work environment sexual harassment\textsuperscript{80} and quid pro quo sexual harassment claims.\textsuperscript{81} The elements of a prima facie case of hostile work environment sexual harassment\textsuperscript{82} are substantially similar to the elements under federal law.\textsuperscript{83}

Following the path of the federal courts, Washington courts adopted the traditional “but for sex” analysis from the sex discrimination model without regard to the individualized nature of sexual harassment claims. As a result, Washington adopted a class-based analysis for individual-based conduct.\textsuperscript{84} Accordingly, under Washington law a plaintiff must show that he or she would not have been harassed had he or she been of the opposite sex.\textsuperscript{85}

The WLAD does differ in one significant aspect from Title VII in that the WLAD requires that the statute be given liberal construction in order to accomplish its purpose.\textsuperscript{86} Washington courts have acknowledged this mandate. In \textit{Payne v. Children’s Home Society of Washington, Inc.},\textsuperscript{87} the court of appeals decided that conduct that was not sexual could still support a sexual harassment claim.\textsuperscript{88} Although the WLAD did not address this issue, the court reasoned that such conduct would still produce barriers to sexual equality in the workplace.\textsuperscript{89} Therefore, although federal

\footnotesize{\textsuperscript{79} Fahn v. Cowlitz County, 93 Wash. 2d 368, 376, 610 P.2d 857, 861 (1980); \textit{Payne}, 77 Wash. App. at 512, 892 P.2d at 1105.\textsuperscript{80} Glasgow, 103 Wash. 2d at 406–07, 693 P.2d at 711–12.\textsuperscript{81} DeWater, 130 Wash. 2d at 134, 921 P.2d at 1062. Quid pro quo harassment is established in the same manner as is the federal claim. Schonauer v. DCR Entm’t, Inc., 79 Wash. App. 808, 823, 905 P.2d 392, 401 (1995).\textsuperscript{82} \textit{Supra} note 45 and accompanying text.\textsuperscript{83} The federal prima facie case of hostile work environment sexual harassment claim contains one element, which Washington State’s provision does not have, that the plaintiff be a member of a protected class. Washington courts have acknowledged this requirement but have not formally included it within the enumerated provisions because the requirement is automatically met by the employee being a man or a woman. Schonauer, 79 Wash. App. at 821 n.21, 905 P.2d at 399 n.21.\textsuperscript{84} Locke, \textit{supra} note 59, at 393; Paul, \textit{supra} note 55, at 352.\textsuperscript{85} Glasgow, 103 Wash. 2d at 406, 693 P.2d at 712. Conduct does not have to be sexual in nature in order to constitute a claim of sexual harassment. Payne v. Children’s Home Soc’y of Wash., Inc., 77 Wash. App. 507, 510–12, 892 P.2d 1102, 1105 (1995).\textsuperscript{86} WASH. REV. CODE § 49.60.020 (2000); Glasgow, 103 Wash. 2d at 404, 693 P.2d at 711.\textsuperscript{87} 77 Wash. App. 507, 892 P.2d 1102 (1995).\textsuperscript{88} \textit{Id.} at 510–12, 892 P.2d at 1104.\textsuperscript{89} \textit{Id.} at 510–11, 892 P.2d at 1104.
III. FEDERAL AND WASHINGTON STATE LAW REGARDING THE EQUAL-OPPORTUNITY-HARASSER DEFENSE

Employers facing sexual harassment claims have sometimes asserted that the harasser’s conduct was not sexual harassment because the employee harassed both men and women alike. The equal-opportunity-harasser defense defeats plaintiffs’ ability to satisfy the “because of sex” element. Plaintiffs normally satisfy this element by showing that they were singled out for adverse treatment based on the plaintiff’s sex. Where men and women are being harassed alike, the plaintiff cannot offer evidence of disparate treatment and therefore cannot prove discrimination because of sex.

Many Washington and federal courts, when faced with this new dilemma, have been forced to accept the equal-opportunity-harasser defense in practice because of the disparate-treatment requirement. Some federal courts and the Washington Court of Appeals have managed to avoid the defense and still find disparate treatment by analyzing the degree of severity of the harassment. Other federal courts have rejected the defense outright by replacing disparate treatment with an individual analysis of the “because of sex” element.


Washington courts and at least six federal circuits have accepted in principle the equal-opportunity-harasser defense. These courts reason

---

91. Holman v. Indiana, 211 F.3d 399, 403 (7th Cir. 2000).
93. Holman, 211 F.3d at 403.
that men and women exposed to the same offensive working environment suffer no discrimination based on sex.\footnote{See supra note 94.} Nondiscriminatory harassment occurs when a supervisor directs abusive language, crude gestures, or sexual demands on an equal-opportunity basis.\footnote{Holman, 211 F.3d at 403.} Plaintiffs cannot satisfy the “because of sex” element because these courts require a showing of disparate treatment.\footnote{Id.}

One of the first federal cases to foresee the possibility of an equal-opportunity harasser and the problems it would create under a disparate-treatment analysis was \textit{Barnes v. Costle}.\footnote{561 F.2d 983 (D.C. Cir. 1977).} In that case, the court noted in dictum that demanding sexual favors from both men and women would not constitute sexual harassment because the conduct was not discriminatory on the basis of sex.\footnote{Id. at 990 n.55.} Following \textit{Barnes}, the Eleventh Circuit in \textit{Henson v. City of Dundee}\footnote{682 F.2d 897 (11th Cir. 1982).} also indicated that sexual overtures directed toward both sexes or conduct that equally offended both sexes would not be actionable under Title VII.\footnote{Id. at 903.} Both courts suggested that if both sexes are being treated equally, disparate treatment would be impossible to prove and plaintiffs would thus be unable to satisfy the “because of sex” element.\footnote{Id.}

The possibility of an equal-opportunity harasser became reality in \textit{Holman v. Indiana},\footnote{211 F.3d 399 (7th Cir. 2000).} where the Seventh Circuit was confronted with such a case and refused to accept the sexual harassment claim.\footnote{Id. at 403–05.} The court reasoned that inappropriate conduct directed toward both sexes falls outside Title VII’s ambit and does not satisfy the “because of sex” requirement because no disparate treatment occurs.\footnote{Id.} The court rejected a reading of \textit{Oncale} that allowed for other evidence besides disparate treatment.\footnote{Id. at 403–05.} In addition, the court warned, if plaintiffs were not required to show disparate treatment between the sexes, Title VII would become a
“code of workplace civility,” which the U.S. Supreme Court in Oncale stated it did not want to happen.\footnote{\textit{Id.} at 404 (citing Oncale v. Sundowner Offshore Servs., Inc., 523 U.S. 75, 80–81 (1998)).}

The Holman court rejected the argument that allowing such a defense will encourage harassment of both sexes in an effort to insulate harassers and employers from liability.\footnote{\textit{Id.}} The court found it difficult to imagine that harassers would understand the intricacies of sexual harassment law.\footnote{\textit{Id.}} Further, the court reasoned that employers would not manufacture the defense because employers would be creating new claims of sexual harassment and thus exposing themselves to additional liability.\footnote{\textit{Id.}}

When Washington faced an equal-opportunity-harasser scenario, the court of appeals in Herried v. Pierce County Public Transportation Benefit Authority Corp.\footnote{90 Wash. App. 468, 957 P.2d 767 (1998).} denied a claim for sexual harassment based on a facial analysis of the “because of sex” element.\footnote{\textit{Id.} at 473, 957 P.2d at 770.} Herried alleged that a co-worker created a hostile work environment by bumping Herried with his body and refusing to make room in a hallway for Herried to pass, forcing Herried to press against lockers in order to get by.\footnote{\textit{Id.} at 470–472, 957 P.2d at 768–70.} These incidents were compounded by the co-worker’s threatening manner and other conflicts between the co-worker and Herried.\footnote{\textit{Id.} at 473, 957 P.2d at 770.} The employer argued that the co-worker was hostile and intimidating to other employees as well, both male and female.\footnote{\textit{Id.} at 473, 957 P.2d at 770.} After the court found that the perpetrator harassed members of both sexes, the court ended its examination.\footnote{\textit{Id.} at 473–74, 957 P.2d at 770–71.} Based on this facial analysis, the court concluded that the plaintiff had failed to present sufficient evidence of harassment because of her sex because the alleged harasser had harassed both sexes.\footnote{\textit{Id.}} Without evidence of disparate treatment, Herried failed to satisfy the “because of sex” element of her sexual harassment claim.\footnote{\textit{Id.}}
B. Avoiding the Equal-Opportunity-Harasser Defense Without Rejecting It Outright

Some courts have sought to allow claims against equal-opportunity harassers while still satisfying the disparate-treatment requirement. These courts require trial courts to conduct a more in-depth analysis of the “because of sex” element by focusing on the differences in the severity of harassment. In an Eighth Circuit case, *Kopp v. Samaritan Health System*, the plaintiff presented evidence that her supervisor verbally assaulted approximately ten female employees. In response, the employer presented evidence that the supervisor had also verbally assaulted four men and argued that the supervisor assaulted everyone and did not discriminate against women. The court found that although the supervisor directed harsh language toward both sexes, the supervisor’s language toward women was of a more serious nature, occurred more often, and was sometimes accompanied by physical contact. By concluding that the supervisor treated women more harshly than men, the court reasoned that a fact-finder could conclude that the defendant’s treatment of women was worse than his treatment of men and hence there was disparate treatment of the sexes.

A Washington court was also able to allow a claim against an equal-opportunity harasser while still meeting the disparate-treatment requirement of the “because of sex” element by conducting an in-depth analysis of the facts. In *Kahn v. Salerno*, the defendant argued that he did not single the plaintiff out because of her sex because he had harassed other employees, both males and females. Reversing a summary judgment, the court of appeals ordered the trial court to determine whether the defendant’s harassing conduct was more abusive toward women than men, and to inquire whether the severity of the impact resulted in disparate treatment. By requiring the trial courts to analyze closely the facts and circumstances surrounding the allegations rather

119. 13 F.3d 264 (8th Cir. 1993).
120. Id. at 269.
121. See id. at 268–69.
122. Id. at 269.
123. Id. at 269–70.
125. Id. at 123, 951 P.2d at 328.
126. Id. at 124, 951 P.2d at 329.
than ending the inquiry on the facial determination that both a male and a female were harassed, the court opened the door to rejecting the equal-opportunity-harasser defense outright.

C. Courts that Reject the Equal-Opportunity-Harasser Defense

The Ninth Circuit and a district court in the Tenth Circuit have rejected the defense. In *Steiner v. Showboat Operating Co.*, the Ninth Circuit stated in dicta that evidence of an equal-opportunity harasser would not bar sexual harassment claims against that harasser by both men and women. Barbara Steiner alleged that her boss, Jack Trenkle, spoke to her in a threatening and derogatory fashion using sexually explicit and offensive terms. Trenkle’s employer argued that Trenkle harassed everyone and therefore was not harassing Steiner based on her sex. The court accepted the fact that Trenkle was abusive to both men and women; however, the court distinguished Trenkle’s treatment of Steiner from his treatment of male employees. Trenkle had referred to male employees as “assholes” but had referred to women as “dumb fucking broads” and “fucking cunts.” The *Steiner* court reasoned that Trenkle’s abuse of men did not relate to their gender and that Trenkle’s abuse of Steiner clearly did.

Although *Steiner*’s analysis was similar to the severity gradations found in *Kopp* and *Kahn*, *Steiner* went one step further in dicta and concluded that even if Trenkle had used remarks equally degrading and intense toward male employees, such conduct would not “cure” his actions toward women. The court added that it did not “rule out the possibility that both men and women . . . [would] have viable claims against Trenkle for sexual harassment.” Therefore, the court did not exclude victims of equal-opportunity harassers just because disparate treatment could not be shown.

127. 25 F.3d 1459 (9th Cir. 1994).
128. *Id.* at 1464.
129. *Id.* at 1461.
130. *Id.* at 1463.
131. *Id.* at 1463–64.
132. *Id.* at 1464.
133. *Id.*
134. *Id.*
135. *Id.*
136. *Id.*
court’s reasoning denied the equal-opportunity-harasser defense and allowed both men and women to lodge successful claims of sexual harassment against the same harasser.137

A district court in *Chiapuzio v. BLT Operating Corp.*138 explained how plaintiffs could satisfy the “because of sex” element without evidence of disparate treatment.139 In *Chiapuzio*, the plaintiffs, Dale and Carla Chiapuzio, complained that their supervisor, Eddie Bell, continuously subjected them to sexually abusive remarks.140 Specifically, Bell would comment that Dale could not satisfy his wife sexually and that he, Bell, could do a better job.141 When speaking to Carla, Bell made sexually explicit advances.142 The employer argued that Bell harassed both men and women and therefore could not have discriminated against the Chiapuzios because of either’s sex.143 The court however, refused to accept this defense.144

First, the court broadly interpreted *Meritor*.145 The U.S. Supreme Court in *Meritor* affirmed that Title VII affords employees the right to work in an environment “free from discriminatory intimidation, ridicule and insult.”146 From this affirmation, the *Chiapuzio* court reasoned that the U.S. Supreme Court was moving away from a disparate-treatment analysis and toward a view that “gender harassment occurs when unwelcome physical or verbal conduct creates a hostile work environment.”147 In other words, the court chose to focus on the “un-welcome” and “severe and pervasive” elements of the sexual harassment test with the

139. Id. at 1337–38.
140. Id. at 1335.
141. Id.
142. Id.
143. Id. at 1336.
144. Id.
145. Id.
belief that these elements would make up for any ambiguity as to the question of “but for one’s sex.”

Second, the court acknowledged that sexual harassment was an uneasy fit under Title VII and adjusted its analysis of the “because of sex” element accordingly. The court relied on Judge Bork’s often-cited dissent in Vinson v. Taylor, which recognized that there was a problem with including the traditional gender discrimination model in sexual harassment. Instead of analyzing the plaintiffs as a class, the court separated each claim and analyzed each individually. By analyzing each claim on its own merits, the court determined that both plaintiffs were harassed because of their sex. The court reasoned that Bell intended to demean and harass Dale because he was male due to remarks regarding Dale’s sexual prowess. Such remarks would not have created the same effect if they were directed toward a woman. Similarly, looking at only the circumstances and conduct directed toward Carla, the court concluded she was harassed because she was female. The key to analysis of the evidence was not whether the perpetrator harassed only members of one gender, but whether gender was a significant or motivating factor in each allegation of harassment. The court did not eliminate the “because of sex” element; rather, the court allowed the element to be satisfied by considering the sufficiency of evidence for each claim individually. Therefore, analysis of the “because of sex” element on an individual basis allowed the district court to deny the equal-opportunity-harasser defense.

148. Id.
149. Id. at 1337.
150. 760 F.2d 1330 (D.C. Cir. 1985).
151. Id. at 1333 n.7 (Bork, J., dissenting).
153. Johnson, supra note 34, at 742–43.
155. Id. at 1337–38.
156. Id. at 1338.
157. Id. at 1337.
158. Id. at 1337–38.
159. Id.
160. Id. at 1336.
IV. COURTS SHOULD REJECT THE EQUAL-OPPORTUNITY-HARASSER DEFENSE BY ADOPTING AN INDIVIDUAL ANALYSIS OF THE “BECAUSE OF SEX” ELEMENT

Analyzing each claim individually will enable courts to deny the equal-opportunity-harasser defense. Where a perpetrator harasses both men and women, the traditional interpretation of the “because of sex” element prevents plaintiffs from making the required showing of disparate treatment. Evidence of disparate treatment, however, does not exhaust the evidence of discrimination because of sex; individual analysis of the plaintiff’s claim may also expose harassment based on an individual’s sex. This analysis separates plaintiffs’ claims and analyzes conduct directed at each plaintiff individually. If that conduct is motivated by the plaintiff’s gender, then the plaintiff has satisfied the “because of sex” element. By utilizing an individual analysis, courts can deny the equal-opportunity-harasser defense, enable sexual harassment law to fulfill its purpose, and avoid absurd results that occur with the traditional disparate-treatment requirement.

A. Acceptance of the Equal-Opportunity-Harasser Defense Is Based on Misplaced Reliance on the Need for Disparate-Treatment Analysis

Evidence of disparate treatment is inappropriate in sexual harassment claims because sexual harassment is an individual discrimination while disparate treatment looks for discrimination on a class level. Courts adopted the disparate-treatment analysis used to satisfy the “because of sex” element in sexual harassment claims from the traditional sex discrimination model. The problem with including this disparate treatment within a sexual harassment framework is that sex discrimination and sexual harassment are theoretically different. While a person who discriminates based on sex presumably discriminates against all members of a particular sex, a sexual harasser may not harass all

161. See Meritor Sav. Bank v. Vinson, 477 U.S. 57, 69 (1986). The U.S. Supreme Court’s decision in Meritor acknowledged the fact that sexual harassment is not the same as class-based discrimination by its suggestion that a plaintiff’s provocative speech and dress may be admissible in sexual harassment cases. See Michelle Ridgeway Peirce, Note, Sexual Harassment and Title VII—A Better Solution, 30 B.C. L. REV. 1071, 1094 (1989). If the work environment were hostile for one gender, then characteristics of the plaintiff, one individual, should be immaterial. See id.
members of a certain sex and might harass members of both sexes.\textsuperscript{162} The decision to sexually harass a person is not based on gender alone but also on considerations such as the relationship to the victim and the victim’s physical and social characteristics.\textsuperscript{163} Therefore, in contrast to traditional class-based sex discrimination, sexual harassment is directed at individuals.\textsuperscript{164} This distinction makes sexual harassment claims an uncomfortable fit with traditional sex discrimination analysis because a court must use a class-based analysis to find individually based harassment.\textsuperscript{165}

Judge Bork recognized this uneasiness when he argued against making sexual harassment actionable in \textit{Vinson v. Taylor}.\textsuperscript{166} He noted that under the traditional discrimination framework an equal-opportunity harasser would be legally excused because a plaintiff would not be able to prove the harassment was due to the plaintiff’s sex where the harassment was directed at both genders.\textsuperscript{167} Importing the disparate-treatment interpretation of the “because of sex” element from sex discrimination cases thus created a hole in sexual harassment laws. By looking for the effects of class-based discrimination, this analysis is blind to situations where gender is just one factor in the harasser’s selection of a victim.\textsuperscript{168} Judge Bork opined that if sexual harassment claims were to be included in Title VII, they would require a separate analysis from discrimination.\textsuperscript{169} The district court in \textit{Chiapuzio} heeded Judge Bork’s warning and developed an individualized analysis of the “because of sex” element to hold equal-opportunity harassers accountable.\textsuperscript{170}

Most courts, however, have not heeded Judge Bork’s objection. By insisting on evidence of disparate treatment to satisfy the “because of sex” element of a prima facie case of sexual harassment, courts have created a doctrinal hole in sexual harassment law through which an equal-opportunity harasser might slip.\textsuperscript{171} The disparate-treatment re-

\begin{footnotesize}
\begin{enumerate}
\item[162.] Locke, \textit{supra} note 59, at 393; Paul, \textit{supra} note 55, at 352.
\item[163.] \textit{See} Paul, \textit{supra} note 55, at 352.
\item[164.] \textit{Id}.
\item[165.] \textit{Id}.
\item[166.] 760 F.2d 1330, 1333 n.7 (D.C. Cir. 1985) (Bork, J., dissenting).
\item[167.] \textit{Id}.
\item[168.] Paul, \textit{supra} note 55, at 352.
\item[169.] \textit{Vinson}, 760 F.2d at 1133 n.7 (Bork, J., dissenting).
\item[170.] \textit{See supra} notes 152–60 and accompanying text.
\item[171.] \textit{See}, e.g., Rabidue v. Osceola Ref. Co., 805 F.2d 611, 620 (6th Cir. 1986); \textit{see also} Locke, \textit{supra} note 59, at 390.
\end{enumerate}
\end{footnotesize}
requirement has forced courts to accept the equal-opportunity-harasser defense because where employees of both sexes are being harassed alike there can be no evidence of disparate treatment, and the “because of sex” element becomes impossible to satisfy. Courts should follow the district court’s example in Chiapuzio and recognize an individual analysis of the “because of sex” element to protect all victims from sexual harassment.

B. How an Individualized Analysis Operates To Satisfy the “Because of Sex” Element

An individual analysis is simply a more in-depth review of the evidence presented by each plaintiff. This analysis would more closely examine an equal-opportunity harasser’s conduct and would not allow a facial showing of a lack of disparate treatment to foreclose this examination. Under the traditional “because of sex” analysis, courts have been primarily concerned with whether or not persons of the opposite sex were harassed in a similar manner as the plaintiff. If the harasser did harass a male and a female, courts have generally looked no further than this facial analysis. Rather than review the harasser’s conduct directed toward each victim, courts have automatically assumed that such conduct was not because of either person’s gender.

An individual analysis would not stop at this facial inquiry but would accept the premise that a harasser can victimize both males and females. The shift in approach would be largely procedural. First, if there were multiple claims, the court would separate the plaintiffs’ claims. Second, when analyzing a plaintiff’s claim, the court should only examine conduct directed at that plaintiff individually regardless of

---

172. Applebaum, supra note 137, at 613.
173. Johnson, supra note 34, at 741–43.
174. This Comment does not address what should be the applicable test for the “because of sex” element for all types of sexual harassment. Instead, this individual analysis should only be applied to targeted conduct that is related to the victim’s gender. Therefore, it does not apply to facially neutral conduct, such as the telling of jokes, cartoons, and the like.
175. Locke, supra note 59, at 393; Paul, supra note 55, at 352.
176. Applebaum, supra note 137, at 613.
177. Id.
whether there were other victims of the harasser.\textsuperscript{179} Therefore, evidence of equal-opportunity harassment should not be dispositive, particularly on summary judgment.\textsuperscript{180} Finally, in order to establish a case from which a fact-finder could reasonably infer that the harassment was because of the plaintiff’s sex, a plaintiff could present three types of evidence as listed in \textit{Oncale}.\textsuperscript{181} These procedural modifications would keep the evidentiary burden on the plaintiff, who would face a showing essentially the same as in a traditional sexual harassment claim. The only difference would be that the claim would not founder on evidence unrelated to instances of discrimination against the individual.

An individual analysis would look for the same evidence of gender-specific conduct as satisfies a traditional claim. Evidence that traditionally satisfies the “because of sex” element establishes a nexus between the harasser’s conduct and the victim’s gender.\textsuperscript{182} If the conduct is related to the victim’s gender, then the fact-finder can conclude that gender animated the harasser’s behavior.\textsuperscript{183} Because such a nexus may exist regardless of the plaintiff’s conduct toward other employees, a fact-finder should be permitted to draw a similar conclusion whenever a plaintiff presents evidence of conduct related to his or her gender, notwithstanding conduct related to other employees’ genders.

The presentation of such evidence of gender-specific conduct to prove discrimination because of sex falls squarely within the evidentiary showing deemed sufficient by \textit{Oncale}.\textsuperscript{184} For example, terms such as “cunt” and “broad” are derogatory terms traditionally used to refer to women.\textsuperscript{185} Because such abusive terms are centered upon a particular sex, a fact-finder could conclude that the victim was harassed because

\begin{footnotesize}
\begin{enumerate}
\item\textsuperscript{179} \textit{Id.}
\item\textsuperscript{180} \textit{Id.} at 1338.
\item\textsuperscript{181} \textit{Id.}, \textit{Oncale} v. Sundowner Offshore Servs., Inc., 523 U.S. 75, 80–81 (1998).
\item\textsuperscript{182} Gender refers to not only whether a person is a male or a female, but also includes cultural manifestations of sex. \textit{See} Paetzold, \textit{supra} note 137, at 27 n.13. This includes personality traits and physical characteristics of a person that are considered to be tied to a particular sex and therefore are presumed to be displayed by that particular sex. Ramona L. Paetzold, \textit{Same-Sex Sexual Harassment, Revisited: The Aftermath of Oncale} v. Sundowner Offshore Services, Inc., \textit{3 Employee Rts. & Emp. Pol'y J.} 251, 257–58 (1999). For example, sensitivity and emotional behavior (personality traits) as well as certain types of clothes, makeup and hairstyles (physical characteristics) are associated with females. \textit{Id.}
\item\textsuperscript{183} \textit{Id.} at 45–46.
\item\textsuperscript{184} \textit{Oncale}, 523 U.S. at 80–81.
\item\textsuperscript{185} \textit{See, e.g.}, Steiner v. Showboat Operating Co., 25 F.3d 1459, 1464 (9th Cir. 1994).
\end{enumerate}
\end{footnotesize}
she was female. Conduct or language that would almost fail entirely of its crude purpose had the victim been of the opposite sex might show a similar nexus with the victim’s sex. For example, a harasser who draws cartoons of the victim giving birth to a baby and exaggerates the sexual characteristics of the victim will, generally, have a more profound effect upon a female victim than a male victim. By focusing on the nexus between specific conduct and the plaintiff’s gender, a court should not be able to conclude that evidence of unrelated conduct defeats, as a matter of law, plaintiff’s evidence of sexual harassment.

Applying this analysis to the facts of Holman v. Indiana, discussed in the introduction of this Comment, will further explain its operation. First, a court would separate Karen and Steven’s claims. Then, the court would look at the harassing conduct directed toward Karen individually. In Karen’s case, her supervisor, Gale, touched her body, sexually propositioned her, and made derogatory remarks about women. A fact-finder could reasonably have concluded that Gale touched Karen and sexually propositioned her because she was a woman and he was attracted to her. Similarly, a fact-finder could have concluded that the sexist remarks would not have had the same derogatory and offending effect if they were directed toward a male. Because Karen’s evidence supported the conclusion that the supervisor chose the conduct he aimed at Karen because she was a female, the court should not have granted summary judgment upon the employer’s evidence of unrelated conduct.

After rejecting summary judgment on Karen’s claim, the court would then turn to Steven’s claim. Gale repeatedly asked Steven for sexual favors. When these favors were rejected, Gale retaliated against him. Such evidence reasonably suggests that the supervisor was attracted to Steven because he was a man and that there was a quid (retaliation) based on a quo (rebuffed proposition). This evidence would have

---

186. Steiner, 25 F.3d at 1463–64.
188. Id. at 783–84.
190. 211 F.3d 399 (7th Cir. 2000).
191. Id. at 401.
192. Id.
193. In regards to same-sex harassment, this test serves to overcome orientation bias. Traditionally, under opposite-sex harassment there is a presumption or inference that sexual conduct is sex-based, whereas in same-sex harassment there is an opposite presumption that the conduct was
carried Steven’s claim past summary judgment had evidence of unrelated conduct not been given dispositive weight. As these examples show, an individual analysis would simply allow courts to reach the merits of discrete claims where plaintiffs present otherwise sufficient evidence of gender-based conduct.

C. An Individual Analysis Is an Appropriate Method for Proving the “Because of Sex” Element

Individually analyzing a plaintiff’s claim on its own merits finds support in sexual harassment case law. Adoption of an individual analysis would follow an emerging trend in federal and Washington law. Furthermore, the U.S. Supreme Court has determined that evidentiary alternatives besides disparate treatment are permissible to support a claim of sexual harassment.

1. Adoption of an Individual Analysis of the “Because of Sex” Element Would Be Consistent with an Emerging Trend in Washington and Federal Law

An individual analysis of the “because of sex” element finds support in Kopp v. Samaritan Health System and Kahn v. Salerno. Where a distinction could be made in the severity of the harassment, these courts held that the equal-opportunity-harasser defense did not apply because the plaintiff could show disparate treatment. Rather than accept the irrelevant excuse that harassment was directed at both a man and a woman, these courts required an in-depth analysis of the conduct itself.

not sex-based. Lussier, supra note 22. This is referred to as the heterosexist bias. Id. Courts view sexual harassment claims through the lens of heterosexuality. Id.; Paetzold, supra note 137, at 61–62. For opposite-sex harassment, the court views sexual conduct through this lens, which provides the nexus between the conduct and the individual’s sex. Lussier, supra note 22, at 944–45. However, when the court operates under this same view when dealing with same-sex harassment, the presumption leaves the court with the view that the conduct was not related to the sex of the victim but was rather general hostility toward the individual. Id. The proposed individual analysis would end this bias because it would suggest that sexual conduct is sex-based regardless of whether it is same-sex or opposite-sex harassment. See id.; Paetzold, supra note 137, at 61–62.

194. 13 F.3d 264 (8th Cir. 1993).
Although the *Kopp* and *Kahn* decisions illustrate successful rejection of the equal-opportunity-harasser defense, requiring evidence of disparate treatment preserves the defense in situations where the court cannot distinguish between severities of harassment. For example, by showing that victims of the harasser were exposed to varying degrees of harassment, the *Kopp* and *Kahn* courts were able to stay within the confines of a disparate-treatment requirement while allowing the victims’ claims to go forward. In cases where the harassment is equal in severity or where the court cannot make a distinction, this reasoning will force courts to deny claims of equal-opportunity-harasser victims. An alternative, and a lasting solution, would be to adopt an individual analysis not susceptible to the equal-opportunity-harasser defense.

The careful examination by the *Kopp* and *Kahn* courts of the distinction in harassment severity is just one step removed from an individual analysis. Rather than just looking to see if a difference in the severity of the conduct exists, a court could separately analyze each victim and determine whether the conduct was based on the victim’s sex. For courts like *Kopp* and *Kahn* that are already conducting an analysis of the unique context of the harassment, an individual analysis of a plaintiff’s sexual harassment claim would require no additional judicial burden.

The *Chiapuzio* court has already taken this next step and replaced evidence of disparate treatment with an individual analysis, thereby precluding the defense. This court acknowledged that sexual harassment can be directed at individuals of both sexes by the same harasser. To allow a claim of sexual harassment, the *Chiapuzio* court reviewed the conduct aimed at each plaintiff individually. Because gender was a motivating factor in the harassment, the plaintiff had satisfied the “because of sex” element. This analysis no longer requires plaintiffs to show disparate treatment. By adopting an individual analysis, *Chiapuzio* provides a foundation to deny the equal-opportunity-harasser defense. Washington and federal courts should follow this lead and deny the equal-opportunity-harasser defense.

---

201. *Id.*
202. *Id.*
2. The U.S. Supreme Court No Longer Requires Evidence of Disparate Treatment Under the “Because of Sex” Element

In Oncale v. Sundowner Offshore Services, Inc.,203 the U.S. Supreme Court paved the way for courts to accept an individual analysis of the “because of sex” element. The Oncale Court ruled that evidence of disparate treatment was sufficient but no longer necessary to establish discrimination “because of sex.”204 In Oncale, the Court was faced with a same-sex sexual harassment claim where evidence of disparate treatment would have been unavailable because the court was unable to compare conduct toward the plaintiff against conduct toward women.205

Showing that same-sex harassment was prohibited under Title VII, the Court stated in dicta that a plaintiff could show discrimination “because of sex” with three different types of evidence: (1) comparative evidence, (2) evidence of gender-specific conduct or actions, or (3) evidence of explicit or implicit proposals of sexual activity.206 Evidence of disparate treatment is thus no longer necessary to establish a prima facie case of sexual harassment.207 By adding the second and third types of evidentiary showings, the Court opened the door to an individual analysis of the “because of sex” element because such evidence concerns conduct directed toward the individual plaintiff. Hence, the Court has implicitly allowed an individual analysis to satisfy the “because of sex” element.

In Holman v. Indiana,208 the Seventh Circuit incorrectly read Oncale to require proof of disparate treatment in all cases of sexual harassment. In Holman, the court made clear that Oncale “explained that one way to ensure that Title VII does not mutate from a prohibition on sexual discrimination to a general prohibition on harassment is to . . . [require] a demonstration that there be different treatment of the sexes.”209 From this

204. See id. at 80–81; Lindemann & Kadue, supra note 46, at 52; Coombs, supra note 72, at 129–30; Lussier, supra note 22, at 957–58.
205. See Oncale, 523 U.S. at 80–81. Although not stated expressly, the facts of the case make it clear that no women were present in the workplace. See id. at 77 (referring to plaintiff’s employment on “eight-man crew”).
206. Id. at 80–81.
207. See id.
208. 211 F.3d 399 (7th Cir. 2000).
209. Id. at 404 (emphasis in original).
The Equal-Opportunity-Harasser Defense

cconcern voiced in Oncale, the Holman court reasoned that disparate treatment is required in all cases of sexual harassment.210

This analysis ignores, however, the fact that Oncale stated disparate treatment was one way to ensure protection against a civility code.211 The Court in Oncale went on to suggest two other means to ensure that the harassment was based on the victim’s sex.212 Furthermore, disparate treatment was impossible to establish in Oncale because there were no female employees against which to examine the victim’s harassment. Although the Court was not entirely clear in its reasoning, commentators agree that Oncale provided plaintiffs with three options to satisfy the “because of sex” element.213

D. Courts Must Deny the Equal-Opportunity-Harasser Defense To Avoid Absurd Results and To Fulfill the Purpose of Sexual Harassment Laws

1. Acceptance of the Defense Leads to Inequitable and Absurd Results

By allowing the equal-opportunity-harasser defense, courts create inequitable and absurd results. First, acceptance of the defense leaves an entire group of sexual harassment victims without protection simply because the perpetrator was an equal-opportunity harasser.214 For example, in Zabkowicz v. West Bend Co.,215 the plaintiff satisfied the “because of sex”216 element when she presented evidence that her supervisor made sexual innuendoes and called her “sexy bitch” and “slut.”217 Conversely, in a similar fact scenario involving an equal-opportunity harasser, the court in Holman rejected the plaintiff’s sexual harassment claim because the harasser also aimed his conduct toward a male employee.218 Thus, similarly situated plaintiffs were treated dif-

210. Id. at 403.
211. Oncale, 523 U.S. at 80–81.
212. Id.
213. LINDEMANN & KADUE, supra note 46, at 52; Coombs, supra note 72, at 129–30; Lussier, supra note 22, at 957–58.
216. Id. at 782.
217. Id. at 785.
218. Holman v. Indiana, 211 F.3d 399, 404 (7th Cir. 2000).
ferently under Title VII. Although one plaintiff’s injury was remedied while another plaintiff’s injury was not, this is not the most significant difference. The real difference is that the plaintiffs like the one in Zabkowicz are able to pursue discrete claims on their own merits while the Holman plaintiffs had their claims muted by other, unrelated claims. Because alternative causes of action, such as the tort of intentional infliction of emotional distress, assault or outrage are not viable options, this difference leaves an entire class of victims without a legal remedy.

Second, an equal-opportunity harasser is a more offensive person than a traditional harasser because an equal-opportunity harasser treats everyone abusively. This reprehensible conduct should not serve as a defense to liability for such actions. Simply put, the courts are accepting that the more people one harasses, the less susceptible one is to prosecution. This result directly contradicts the purpose of sexual harassment laws because it allows conduct that supports inequitable barriers in the workplace to go unpunished merely because the perpetrator chose to harass both a male and a female rather than only one individual.

219. Applebaum, supra note 137, at 618–19. The tort of intentional infliction of emotional distress requires that the plaintiff suffer severe or extreme emotional distress from extreme or outrageous conduct from the harasser. Id. at 619. Because many victims of harassment do not suffer severe emotional distress as a result of being harassed, this cause of action is not a viable claim for many harassment victims. Id.

220. Id. at 620. Assault requires an apprehension of a harmful or offensive contact. Id. Words alone will usually not support a cause of action for assault. Id.

221. The tort of outrage is supported by a showing that the defendant’s conduct was “so outrageous in character, and so extreme in degree, as to go beyond all possible bounds of decency, and to be regarded as atrocious, and utterly intolerable in a civilized community.” Grimbsy v. Samson, 85 Wash. 2d 52, 59, 530 P.2d 291, 295 (1975). The plaintiff cannot establish the tort of outrage based on mere insults and indignities that cause embarrassment or humiliation. Dicomes v. Washington, 113 Wash. 2d 612, 630, 782 P.2d 1002, 1013 (1989).

222. In addition, battery and constructive discharge are unavailing. Battery requires the plaintiff to prove that there was unpermitted physical contact, and this may not occur in a sexual harassment context. Applebaum, supra note 137, at 620. For constructive discharge, existence of unlawful sex discrimination alone will not support the claim. Glasgow v. Georgia-Pacific Corp., 103 Wash. 2d 401, 408, 693 P.2d 708, 713 (1985).


224. Applebaum, supra note 137, at 616.

225. Id.
2. The Equal-Opportunity-Harasser Defense Frustrates the Purpose Behind Sexual Harassment Laws

Equal-opportunity harassment creates the same hostile work environment and inequalities that traditional harassment does and thus should be prohibited by sexual harassment laws. Courts have recognized that the purpose behind prohibiting sexual harassment was to eradicate arbitrary barriers to sexual equality in the workplace.226 Like a traditional harasser, a perpetrator who harasses members of both sexes confounds this purpose in two ways.227 First, the equal-opportunity harasser’s motivation behind the harassment is the same as in traditional sexual harassment because the perpetrator singles out a victim for gendered reasons.228 Second, the effect of the harassment is the same as in traditional sexual harassment as victims are forced to experience unwelcome sexual advances, physical contact, or verbal abuse.229 Such conduct supports the continuation of sex stereotyping and the maintenance of a gender hierarchy because it serves to punish those who do not conform to scripted gender norms230 and to block professional opportunities for both sexes.231 Therefore, an equal-opportunity harasser’s conduct leads to the same arbitrary barriers to sexual equality in the workplace.232 In order to bring these harassers within the policy bounds of sexual harassment laws, courts should adopt an individual analysis of the “because of sex” element and eliminate the equal-opportunity-harasser defense.

228. Locke, supra note 59, at 407.
229. See Peirce, supra note 161, at 1096 (“[T]he offensiveness of the harassment is not lessened merely because the employer also harasses men. To the woman it is the harassment itself that offends.”).
232. See supra note 231.
Washington has an even greater responsibility to adopt an individual analysis because Washington statutory law mandates that sexual harassment laws be given liberal construction in order to fulfill their purpose of eradicating barriers to sexual equality in the workplace and protecting all employees from sexual harassment. Washington courts have used liberal construction to modify sexual harassment law so that inequities in the workplace are extinguished. For example, the Washington Court of Appeals, in *Payne v. Children’s Home Society of Washington, Inc.*, had to decide whether a sexual harassment claim existed where the conduct complained of was not explicitly sexual. Although the WLAD was silent on this issue, the court reasoned that gender-based harassment that is not of a sexual nature also creates barriers to sexual equality. Therefore, to fulfill the statute’s purpose, the court determined that conduct need not be sexual in nature to support a claim for sexual harassment under the WLAD.

Likewise, when presented with the equal-opportunity-harasser defense, Washington courts should liberally construe the WLAD and adopt an individual analysis to deny this defense. By allowing the defense, courts protect conduct that would otherwise be harassment but for a broader group of victims. Instead of fulfilling the statute’s purpose, the defense leaves a supervisor with the power to sexualize the workplace. This result contradicts the statute’s purpose because the harassing conduct serves to punish victims for not following gender norms established by the harasser and creates barriers to advancement in the workplace for both sexes.


 Rejecting the equal-opportunity-harasser defense and no longer requiring a showing of disparate treatment will not turn sexual ha-

236. *Id.* at 510, 892 P.2d at 1104.
237. *Id.* at 511, 892 P.2d at 1105.
238. *Id.* at 511–12, 892 P.2d at 1105.
240. *See* Calleros, * supra* note 231, at 79; Murphy, * supra* note 227, at 1147–49.
The Equal-Opportunity-Harasser Defense

rassment laws into a code of conduct for the workplace. Proponents of the defense fear that by discarding the disparate-treatment requirement, sexual harassment laws will become regulations of how employees should treat each other;241 however, this fear is misguided.

First, although evidence of disparate treatment would not be required under the new individualized analysis of the “because of sex” element, the plaintiff must still show that one of the motivating factors behind the harasser’s conduct was the plaintiff’s gender.242 This requirement would still guard against Title VII and state harassment laws turning into general harassment statutes in the workplace because the plaintiff would still have to show a nexus between the harassment and the plaintiff’s gender.243

Second, the U.S. Supreme Court in Oncale stated that the “severe and pervasive” element is the crucial requirement to protect against a workplace civility code.244 If the harasser’s conduct is not severe and pervasive enough to create an objectively hostile work environment, then the conduct does not rise to the level of actionable sexual harassment.245

Finally, the decision in Oncale no longer requires a showing of disparate treatment.246 If the Court was concerned that sexual harassment laws would turn into civility codes without proof of disparate treatment, the Court would not have provided plaintiffs alternatives to showing disparate treatment.247 Therefore, rejecting the equal-opportunity-harasser defense and not requiring proof of disparate treatment will not render sexual harassment laws civility codes.

V. CONCLUSION

Washington and federal courts should adopt an individual analysis of the “because of sex” element, thus abolishing the equal-opportunity-harasser defense. The defense allows outrageous harassers to go unpunished and leaves unprotected victims who are forced to endure

244. Id. at 82; see supra note 45 and accompanying text.
245. Oncale, 523 U.S. at 82.
246. See id. at 80–81; supra notes 204–07 and accompanying text.
appalling employment situations. Evidence of disparate treatment to satisfy the “because of sex” element is inappropriate in equal-opportunity-harasser cases because it is a class-based analysis attempting to detect an individual discrimination. Although Washington and federal courts have already tried to get around the defense by making a distinction in the severity of the harassment, disparate-treatment analysis will limit the ability of the courts to handle equally severe harassment. The logical progression of these courts’ efforts is to adopt an individual analysis that will provide lasting protection for all victims of equal-opportunity harassers. The U.S. Supreme Court has already paved the way for an individual analysis by no longer requiring evidence of disparate treatment to satisfy the “because of sex” element. To protect all victims of sexual harassment and to curb absurd results of the defense, Washington and federal courts should adopt an individual analysis of the “because of sex” element, thereby rejecting the equal-opportunity-harasser defense.
THE FIRST AMENDMENT VERSUS THE WORLD TRADE ORGANIZATION: EMERGENCY POWERS AND THE BATTLE IN SEATTLE

Aaron Perrine

Abstract: The 1999 World Trade Organization (WTO) ministerial meeting in Seattle was the target of highly organized, widely supported protest demonstrations. In response to the protests, city officials declared a state of emergency, ordering nighttime curfews and a daytime “no-protest zone” in downtown Seattle. They reasoned that the zone was necessary to protect the rights of WTO delegates and to restore public order. This Comment argues that mass nonviolent protests deserve more First Amendment protection than was afforded to demonstrators in Seattle. Even when violence occurs and public order is threatened, governments must narrowly tailor emergency orders to avoid trampling on peaceful protesters’ First Amendment rights. An analysis of U.S. Supreme Court and Ninth Circuit case law demonstrates that Seattle’s “no-protest zone” was unconstitutional and that courts should strike down similar restrictions on mass protests.

On November 30, 1999, nearly 40,000 people gathered in downtown Seattle in a “mass demonstration” to protest the annual ministerial meeting of the World Trade Organization (WTO). Protesters representing labor, environmental, and human and animal rights groups considered the WTO ministerial meeting a unique opportunity to express their views on the effects of globalization and on the global influence of transnational corporations. Demonstrations forced the delay of the ministerial meeting’s opening ceremony because many delegates could not reach the Washington State Convention and Trade Center from their...
The vast majority of protesters were peaceful, obeyed the police, and were not civilly disobedient; however, property damage, violence, and police provocation occurred. By mid-afternoon, Seattle Mayor Paul Schell had declared a state of emergency. Between 3:30 p.m. on November 30 and 6:00 a.m. on December 1, Mayor Schell issued three emergency orders. One of these orders, Emergency Order Number 3 (Order 3) established a limited daytime curfew zone in downtown Seattle. Order 3 did not, on its face, restrict the type of activity allowed within the zone; however, many have concluded that the zone was enforced as a “no-protest zone,” keeping demonstrators, but not shoppers, out of downtown for the rest of the conference. Although public order was restored by December 1, city officials did not lift the “no-protest zone” until after the conference ended on December 3.

A discussion of the constitutional implications of mass demonstrations is relevant and timely. Protesters continue to litigate whether their First


8. The precise moment when each order was promulgated is unknown and has been the subject of much debate; however, all three were in effect by the early morning of December 1. Id. Emergency Order Number 1 established a nighttime curfew in downtown Seattle. Local Proclamation of Civil Emergency Order Number 1, at http://www.ci.seattle.wa.us/wto/procs1a.htm (last visited May 1, 2001) [hereinafter Order 1]. Emergency Order Number 2 prohibited the use, possession, and sale of gas masks to anyone other than the police or the military (one version also allowed sales to members of the press with proper credentials). Local Proclamation of Civil Emergency Order Number 2, at http://www.ci.seattle.wa.us/wto/procs2.htm. This Comment does not address the constitutionality of these other orders.

9. Local Proclamation of Civil Emergency Order Number 3 (revised), at http://www.ci.seattle.wa.us/wto/procs3.htm (last visited May 1, 2001) [hereinafter Order 3]. For an excellent map of the restricted area and protest events, see http://www.cityofseattle.net/wtocommittee/maps/map_intro.htm. Between the time of its initial promulgation and its expiration on December 4, Order 3 was revised such that the original “panhandle” encompassing the Westin Hotel was no longer part of the restricted zone.


11. See infra notes 35–40 and accompanying text.

Amendment rights were denied by the City of Seattle and the Seattle Police Department. In addition, legal analysis of the Seattle WTO demonstrations should apply with equal force to similar demonstrations, such as the International Monetary Fund/World Bank protests in Washington, D.C., on April 16–17, 2000, the Republican National Convention protests on August 1–4, 2000, the Democratic National Convention protests on August 14–17, 2000, and the presidential inauguration protests on January 20, 2001. Courts need a consistent analytical framework for assessing government responses to mass demonstrations.

This Comment argues that Order 3 was used to regulate and restrict activities protected by the First Amendment. In the context of mass demonstrations, the First Amendment should protect peaceful protesters when their message is unpopular and their audience hostile—even when other protesters are violent or commit illegal acts. Part I details a chronology of the WTO meetings to facilitate a discussion of Order 3’s constitutionality. Part II demonstrates how the First Amendment protects protesters’ rights to express their views in public fora. Part III discusses the evolution of emergency power in the United States and the extent to which this power legitimately restricts First Amendment rights. Finally, Part IV argues that measures such as Order 3 violate the First Amendment and may not be used by local governments to restrict mass protests.

I. THE SEATTLE WTO MINISTERIAL MEETING AND THE NONVIOLENT PUBLIC PROTESTS

In spite of adequate notice, the City of Seattle and its police force were unprepared for the mass demonstrations that took place on November 30, 1999. In the months leading up to the WTO conference, Seattle Mayor Paul Schell and other city officials focused their attention

---


14. Activists who cut their teeth in Seattle have also gone on to stage highly organized, very large, peaceful protests around the world. See, e.g., Anthony DePalma, In Quebec’s Streets, Fervor, Fears and a Gamut of Issues, THE NEW YORK TIMES, Apr. 22, 2001 at A1 (describing mass demonstrations at the third Summit of the Americas in Quebec, Canada).

on promoting the potential benefits of the conference to the Puget Sound region. As November approached, city officials optimistically presumed that the city was prepared to handle the protests and the conference without additional law enforcement assistance. The Direct Action Network and other protest groups, however, were unambiguous in their intent to use nonviolent protest to shut down the WTO meeting by filling the streets and blocking access to the convention center. By 6:00 a.m. on November 30, few Seattle police officers had been deployed, even as protesters fanned out into preplanned groups to block key intersections. By 8:30 a.m., Seattle was witnessing its largest ever mass nonviolent protest.

The vast majority the protestors who filled city streets on November 30 were nonviolent; in fact, many took affirmative steps to prevent violence and to stop violence where it occurred. Media reports, however, focused attention on the so-called “Black Block” anarchists, who wore distinctive black clothing and were photographed committing vandalism on November 30. While these violent individuals represented only a small fraction of the protestors downtown, few if any were arrested on the 30th.

As the day progressed, the Seattle Police Department incrementally lost control of downtown Seattle. While the City had issued myriad

20. The Seattle City Council and the Seattle Police Department agree that only a small minority of WTO protestors were violent. Operations Panel Report, supra note 10, at 7; Police Summary of Events, supra note 1. A larger, but still relatively small, number committed nonviolent civil disobedience. Operations Panel Report, supra note 10, at 4. The majority of protestors were engaged in nonviolent, legal behavior on November 30. Id.
23. See id. By mid-morning, the police department nearly ran out of tear gas and pepper spray, and had more flown in from Wyoming. Plainclothes officers were forced to “smuggle” additional canisters of the chemical weapons through crowds of protestors. J. Martin McOmber et al., Police Caught Short at WTO: Officers Went to Wyoming for Tear Gas, SEATTLE TIMES, Mar. 8, 2000, at A1, B1.
demonstration and parade permits, the police blamed “unscheduled” events for disrupting the WTO conference. Many WTO delegates were stranded at their hotels because the City had not provided secure routes to and from the convention center. The City was similarly unprepared to arrest the tiny minority of violent protestors, whose activities intensified throughout the day. By 3:00 p.m., with more than 20,000 protestors in the streets and a labor-union march about to converge on downtown, Mayor Schell declared a state of emergency. At Mayor Schell’s request, Governor Gary Locke mobilized the Washington National Guard. By the early morning of December 1, Mayor Schell had signed Emergency Order 3, which established an “exclusionary zone,” a daytime curfew restricting access to a fifty-block area of the downtown core. Within the zone, Order 3 declared that “no person shall enter or remain in a public place.” The Order contained exceptions for WTO delegates and personnel, employees and owners of businesses within the zone, persons who resided within the zone, properly credentialed press, city officials, and emergency and public-safety personnel.

Smaller protests continued throughout the week. On December 1, some additional violence and vandalism occurred, and hundreds of nonviolent protesters were arrested. December 2 “saw a significant de-escalation in demonstration activities” and police reported only one, peaceful demonstration of approximately 1,000 protestors. On December 3, police reported no significant threats to public order.

25. Id.
27. See Police Summary of Events, supra note 1.
28. See supra note 1.
29. See Order 1, supra note 8.
30. Broom et al., supra note 7.
31. See supra note 8; cf. Kimerer Declaration, supra note 24, at 3–4.
32. See Postman & Carter, supra note 6.
33. Order 3, supra note 9.
34. Id.
35. Police Summary of Events, supra note 1.
36. Id.
37. Id.
Police officers from around the state and National Guard troops remained deployed downtown throughout the week. After the chaotic events of November 30, police never again lost control of downtown, and at all times subsequent to the afternoon of December 1 WTO delegates were able to travel safely to and from their hotels. Although public order had been restored, Order 3 remained in effect, keeping protesters far away from WTO delegates for the remainder of the conference. Order 3 expired at 7:00 a.m. on Saturday, December 4.

Protestors stated that the opportunity to interact with WTO delegates was unique, and that the City’s actions caused irreparable harm to their right of free expression. On Wednesday, December 1, the American Civil Liberties Union (ACLU) sought an injunction against Order 3 in federal court. The ACLU cited Collins v. Jordan for the proposition that daytime exclusionary zones violate the First Amendment. Citing Madsen v. Women’s Health Center, the City argued that establishing a buffer zone to protect private property and ensure the free flow of traffic was constitutional. District Judge Robert Bryan noted that each case involved distinguishable factual situations, but declined to grant the injunction. While litigation is ongoing regarding whether the City infringed on the First Amendment rights of protesters during the WTO ministerial meeting, no court has subsequently reviewed the constitutionality of Order 3.

38. Id.
39. See generally id.
40. Id.
41. Daniel Jack Chasan et al., Out of Control, Seattle’s Flawed Response to the WTO Protest, at http://www.aclu-wa.org/ISSUES/police/WTO-Report.html (last visited May 1, 2001). The ministerial meeting was the first WTO conference to occur in the United States. See http://www.wto.org/english/thewto_e/minist_e/minist_e.htm (last visited May 1, 2001); see also Operations Panel Report, supra note 10, at 4 (finding ability of anti-WTO groups to protest in visible and effective manner was “seriously impaired” by limited curfew zone).
43. 110 F.3d 1363 (9th Cir. 1996); see infra Part III.B.
45. 512 U.S. 753 (1994); see infra notes 90–95 and accompanying text.
II. MASS DEMONSTRATIONS AND THE FIRST AMENDMENT

The First Amendment to the U.S. Constitution provides that “Congress shall make no law . . . abridging the freedom of speech . . . .” 48 It is axiomatic, however, that States are not “powerless to regulate the conduct of demonstrators and picketers.” 49 The U.S. Supreme Court has articulated rules for the restriction or regulation of First Amendment activities. First, violent or threatening acts by separate groups cannot be a basis for arresting peaceful protesters. Second, even if protests are violent or some protesters break the law, the protest movement itself cannot be characterized as “violent” for First Amendment purposes unless violence is “pervasive.” Third, the Court closely scrutinizes restrictions that act as “prior restraints” on speech. Finally, content-neutral regulations are subject to reduced scrutiny, but must still meet the constitutional guidelines set forth by the Court. This part will consider each rule in turn.

A. Peaceful Demonstrators Retain Their First Amendment Rights Even when Violent Counterdemonstrations Occur

A speaker conveying an unpopular message can engender a hostile audience response that threatens public order. Conflicting U.S. Supreme Court precedent exists as to whether police can infringe on peaceful speakers’ First Amendment rights to avoid violence by the speakers’ audience. While an early case allowed this “heckler’s veto” 50 to override a speaker’s First Amendment rights, the right to deliver an unpopular message has been supported in later decisions.

In an early case, the Court held that First Amendment rights gave way when unpopular speech threatened public order. In Feiner v. New York, 51 Irving Feiner, a black man, used a bullhorn to address a small crowd in Syracuse, New York. 52 His comments included racially charged, derogatory remarks toward government officials. 53 Feiner’s speech excited white members of the crowd, and at least one crowd member

48. U.S. CONST. amend. I.
52. Id. at 316–17.
53. Id. at 317.
threatened violence if the police did not order Feiner to stop.\textsuperscript{54} Fearing a fight would break out, a police officer asked Feiner to step down from his podium, and later arrested him for disobeying that order.\textsuperscript{55} The Court found that the order to stop speaking was reasonable and that the policeman had acted out of a desire to protect public safety, not because he disagreed with Feiner or wanted to stifle his political message.\textsuperscript{56}

The \textit{Feiner} decision has been eclipsed by later case law. For example, in \textit{Gregory v. City of Chicago},\textsuperscript{57} protesters marched to Chicago Mayor Richard Daley’s home to protest the delay in desegregating the city’s public schools.\textsuperscript{58} While the march itself was peaceful and orderly, hecklers from the neighborhood were insulting and violent, hurling eggs and rocks and threatening the protesters.\textsuperscript{59} Although nearly 100 police officers accompanied the eighty-five protesters, the hostile crowd grew to more than 1200.\textsuperscript{60} The Chicago police, fearing they could not control the crowd, ordered Gregory and other marchers to disperse and arrested them when they refused to do so.\textsuperscript{61} Gregory was convicted under a Chicago disorderly conduct statute.\textsuperscript{62}

The \textit{Gregory} Court held that the marchers were protected by the First Amendment.\textsuperscript{63} The Court noted that “peaceful and orderly” marching fell within the sphere of protected First Amendment activities and that the state’s charge of disorderly conduct against the protesters was without evidentiary support.\textsuperscript{64} Finding the conviction lacked evidentiary support implied that the threatening and violent behavior of counter-demonstrators did not make the march itself “disorderly” for First Amendment purposes.\textsuperscript{65} While \textit{Gregory} did not explicitly overrule

\begin{itemize}
\item \textsuperscript{54} See id.
\item \textsuperscript{55} Id. at 317–18.
\item \textsuperscript{56} Id. at 320–21.
\item \textsuperscript{57} 394 U.S. 111 (1969).
\item \textsuperscript{58} Id. at 115 (Black, J., concurring).
\item \textsuperscript{59} Id. at 126–30 (Black, J., concurring).
\item \textsuperscript{60} Id. at 128–30 (Black, J., concurring).
\item \textsuperscript{61} Id. at 116–17 (Black, J., concurring).
\item \textsuperscript{62} Id. at 112.
\item \textsuperscript{63} Id. at 112–13.
\item \textsuperscript{64} Id. at 112.
\item \textsuperscript{65} See, e.g., Sabel v. Stynchcombe, 746 F.2d 728, 730–31 (11th Cir. 1984) (holding that arrest for failure to obey police was order unconstitutional because hostile audience, rather than demonstrators, was source of public disorder).
\end{itemize}
First Amendment and Nonviolent Protests

*Feiner*, the rule in *Gregory* is that the acts of groups separate from peaceful protesters do not diminish the latter’s First Amendment rights.67

### B. Peaceful Demonstrators Retain Their First Amendment Rights Even when Isolated Incidences of Protester Violence Occur

Although violent protest is not protected by the First Amendment, nonviolent protesters retain their First Amendment rights even if violent acts occur simultaneously with peaceful protests. In *NAACP v. Claiborne Hardware Co.*, the U.S. Supreme Court held that a seven-year NAACP boycott against businesses in Claiborne County, Mississippi, was protected by the First Amendment even though acts of violence were carried out by boycott participants. In *Claiborne*, merchants brought suit against the NAACP, the chief organizer of the boycott.69 The Mississippi Supreme Court held the NAACP liable under Mississippi tort law, and awarded damages based on the economic loss merchants suffered during the boycott.70 The U.S. Supreme Court reversed.71

The boycott at issue in *Claiborne* was not entirely peaceful. Some activists threatened black citizens of Claiborne County with violence if they patronized white-owned businesses.72 In two cases, shots were fired at homes of blacks who disregarded the boycott.73 In at least one case property damage occurred.74 Additionally, a group of protesters known as “Black Hats” stood outside boycotted businesses, recording names of those who entered and generally intimidating Claiborne citizens.75 At one point during the boycott, tension was so high that local police requested reinforcements from the Mississippi Highway Patrol76 and the mayor established a nighttime curfew.77 Still, the Court noted that the “practices

---

66. The *Gregory* Court distinguished *Feiner* by noting that the Chicago ordinance did not define “disorderly conduct” as the failure to obey a police order. *Gregory*, 394 U.S. at 112 n.*6.
67. See generally Wagner, *supra* note 50, at 219–22 (comparing *Gregory* with *Feiner*).
68. 458 U.S. 886 (1982).
69. *Id.* at 889.
70. *Id.* at 894–95.
71. *Id.* at 934.
72. *Id.* at 903–04.
73. *Id.* at 904.
74. *Id.*
75. *Id.* at 903.
76. *Id.* at 902.
77. *Id.*
generally used [by the protesters] to encourage support for the boycott were uniformly peaceful and orderly,” and refused to hold the organizers liable for economic damages. The Court found no justification for abridging the First Amendment rights of boycotters where the majority of the conduct stemming from the boycott was peaceful:

The taint of violence colored the conduct of some of the petitioners. They, of course, may be held liable for the consequences of their violent deeds. The burden of demonstrating that it colored the entire collective effort, however, is not satisfied by evidence that violence occurred or even that violence contributed to the success of the boycott. A massive and prolonged effort to change the social, political, and economic structure of a local environment cannot be characterized as a violent conspiracy simply by reference to the ephemeral consequences of relatively few violent acts.

_Claiborne_ elucidates a different facet of First Amendment law than _Gregory_. In _Gregory_, the Court held that violence by outside hecklers—counterprotesters not sharing the same purpose as those arrested—did not excuse the government’s transgression of First Amendment rights. In _Claiborne_, the Court carved out a larger space of First Amendment protection, suggesting that unless violence is “pervasive,” the rights of nonviolent protesters must be protected.

C. _Courts Strictly Scrutinize Prior Restraints on First Amendment Rights_

Courts strictly scrutinize prior restraints on expression. The U.S. Supreme Court has held that “prior restraints [on] expression [come] to this Court bearing a heavy presumption against... constitutional validity.” The early cases developing the use of prior restraints generally involved statutes or ordinances creating standardless licensing

78. _Id._ at 903 (emphasis added).
79. _Id._ at 933.
80. _See supra_ notes 57–67 and accompanying text.
81. _See Claiborne_, 458 U.S. at 923.
schemes. Modern cases, discussed here, focus on whether speech is being targeted because of its content and whether the expression of a given message is foreclosed entirely by the restriction at bar. These cases focus on injunctions granted by lower courts to prohibit expressive activity.

Injunctions against First Amendment activities may operate as prior restraints when their purpose is to restrict the content of a particular message and when expression of that content is foreclosed ahead of time. For example, in New York Times Co. v. United States, the federal government attempted to obtain an injunction barring the New York Times newspaper from publishing the “Pentagon Papers,” a classified report on government policy during the Vietnam War. The Court held that the government had not met the “heavy burden” of justifying a prior restraint on speech, and declined to grant an injunction against publication of the report.

Not all restrictions that restrain speech prior to its occurrence constitute prior restraints. In Madsen v. Women’s Health Center, for example, a Florida state judge issued an injunction establishing a thirty-six-foot buffer zone around an abortion clinic. Within the zone, anti-abortion protest was not allowed. The Court refused to analyze the injunction as a prior restraint and instead found it to be a “content-
neutral order that left open alternative methods of expression. Because protesters could still protest from the other side of the street, across from the clinic, the ban was only an “incidental” restriction on free expression and was therefore permissible.

Although the injunction in Madsen was held not to be a prior restraint, the Court noted that the First Amendment prohibited restricting any more speech than was necessary to serve the government’s interest. While the Court held that the thirty-six-foot buffer zone was justified and still left adequate additional avenues for expression, it struck down other portions of the injunction. The Madsen Court held unconstitutional the restricting on the use of “images observable” to patients within the clinic, and restricting protesters from approaching patients for the purpose of providing counseling within 300 feet of the clinic. This portion of the Madsen holding suggests that courts must analyze the specific parameters of restrictions to determine whether they are justified vis-à-vis the government’s interest.

D. Judicial Review of Content-Neutral Restrictions

Content-neutral restrictions on expressive activity are those that do not depend on the content or communicative impact of the activity, facially or as applied. Compared to prior restraints, content-neutral restrictions enjoy a substantially reduced level of judicial scrutiny. Not all content-neutral restrictions, however, survive even this level of scrutiny.

To determine whether a restriction on public expression is content neutral, courts examine whether the government “has adopted a regulation of speech because of disagreement with the message it conveys.” As long as the government articulates a legitimate interest “unrelated to the suppression of free expression,” the actual effect of the restriction is inapposite. Even where a restriction has an “incidental

92. See supra Part II.D.
93. See Madsen, 512 U.S. at 764 n.2.
94. Id. at 763.
95. Id. at 765.
96. Id. at 772–73.
97. Id.
First Amendment and Nonviolent Protests

effect” on some viewpoints but not others, the restriction is deemed content neutral if it is “justified without reference to the content of the regulated speech.”

Content-neutral restrictions may regulate expression, but may not ban protected classes of speech entirely. The Court has held that a prohibition on an entire class of speech, even though free of content or viewpoint discrimination, poses too great a danger to First Amendment rights by “eliminating a common means of speaking.” For example, in Schneider v. State, the Court struck down an ordinance prohibiting leafleting on public streets. Even though the ordinance was content neutral, the Court found that the state’s interest in preventing littering was insufficient to justify a ban on leafleting as an entire class of speech.

Content-neutral restrictions fall into two general categories. When the government is attempting to regulate the underlying conduct associated with expressive activity, and a restriction on First Amendment rights is incidental to the conduct-based regulation, the restriction is subject to the test in United States v. O’Brien. Second, when the government applies content-neutral restrictions to expressive activity in the public forum, it may regulate only the time, place, or manner of the activities, under the rule in Ward v. Rock Against Racism. The U.S. Supreme Court has applied these time, place, and manner rules to restrictions designed to protect the entry to and exit from specific locations in the context of abortion-clinic cases.

104. 308 U.S. 147 (1939).
105. Id. at 162.
106. Id. The First Amendment permits governments to require a permit for expressive activities where the permitting scheme is content neutral and serves a legitimate government interest. See, e.g., Cox v. New Hampshire, 312 U.S. 569, 574–76 (1941) (upholding statute requiring parade permits because it served legitimate government interest in maintaining public safety, traffic flow, and adequate policing during parades). But see Lovell v. City of Griffin, 303 U.S. 444, 451 (1938) (overturning city ordinance requiring permit to distribute leaflets on grounds that permitting scheme amounted to censorship).
1. **Content-Neutral Restrictions on Expressive Conduct and the O’Brien Test**

*United States v. O’Brien* sets forth the test for evaluating content-neutral restrictions on expressive activities. In *O’Brien*, the defendant was arrested for burning his draft card in violation of a federal statute. The defendant argued that burning the card was expressive activity protected by the First Amendment. The U.S. Supreme Court disagreed and articulated a test for determining the constitutionality of content-neutral restrictions on expression:

>[A] government regulation is sufficiently justified if it is within the constitutional power of the Government; if it furthers an important or substantial governmental interest; if the governmental interest is unrelated to the suppression of free expression; and if the incidental restriction on alleged First Amendment freedoms is no greater than is essential to the furtherance of that interest.

In applying this test, the *O’Brien* Court upheld the statute that banned the burning of draft cards. First, the Court found that the statute was part of a registration system for military conscription, valid pursuant to Congress’s power to raise and support armies. Second, the Court found that the statute furthered the “substantial” government interest in efficient operation of the draft. Third, the Court found that only the noncommunicative impact of the defendant’s conduct was prohibited by the statute, and that the statute did not discriminate based on the content of speech that accompanied the illegal conduct of burning a draft card. Fourth, the Court found that the statute was narrowly tailored because it prohibited only conduct that clearly frustrated the government’s interest in a smooth-running conscription system.

---

110. *Id.*
111. *Id.* at 371–72.
112. *Id.* at 377.
113. *Id.* at 377–78.
116. *Id.* at 381–82.
2. Content-Neutral Restrictions in Public Fora and the Ward Test

When restricting expressive activity in public fora, the government may regulate only the time, place, or manner of the activity. The test for time, place, or manner restrictions is very similar to the O’Brien test for expressive conduct. In *Ward v. Rock Against Racism*, the Court found constitutional an ordinance requiring concert organizers in Central Park to use sound equipment provided by the city rather than private equipment. The government’s stated purpose was to shield the public from excessive noise by controlling the volume of concerts in the park. Because the government’s interest was unrelated to the content of the music played, the Court held that the ordinance was content neutral.

The *Ward* Court also noted that time, place, and manner restrictions must be narrowly tailored and leave open alternative channels for communication. However, the Court stressed that the government’s need not employ the least restrictive method of serving the government’s interest. Instead, “the requirement of narrow tailoring is satisfied ‘so long as the . . . regulation promotes a substantial government interest that would be achieved less effectively absent the regulation.’” The Court held that the ordinance met the “alternative channels” test because users of the Central Park band shell could still adequately express their message with reduced volume.

In the Ninth Circuit, the government bears the burden of showing that time, place, and manner restrictions are narrowly tailored and provide adequate alternative methods for communication. In *Bay Area Peace Navy v. United States*, a seventy-five-yard “safety exclusion zone” in front of the VIP pier during a Fleet Week boat parade was held unconstitutional. The Bay Area Peace Navy, a local activist group,

118. *Id.* at 802.
119. *Id.*
120. *Id.* at 792.
121. *Id.* at 791 (citing Clark v. Cnty. for Creative Non-Violence, 468 U.S. 288, 293 (1984)).
122. *Id.* at 798–99 (quoting United States v. Albertini, 472 U.S. 675, 689 (1985)).
123. *Id.* at 802.
124. Bay Area Peace Navy v. United States, 914 F.2d 1224, 1227 (9th Cir. 1990) (holding that government bears the burden of proving that restriction is narrowly tailored and leaves open alternative methods for communication).
125. 914 F.2d 1224 (9th Cir. 1990).
126. *Id.* at 1231.
wished to demonstrate against the celebration of U.S. naval power by conducting its own boat parade in front of the pier. Because the group’s expressive activity included singing and other oral communication, its parade would have been ineffective from seventy-five yards away. The Ninth Circuit noted that the government “is not free to foreclose expressive activity in public areas on mere speculation about danger,” and held that the government had failed to show that a seventy-five-yard exclusion zone was not “substantially broader than necessary.”

While within the scope of Ward and O’Brien, the Ninth Circuit’s holding in Bay Area Peace Navy explicitly places the burden of proof on the government to justify restraints on expression. The government must prove both that the restriction at bar is not “substantially broader than necessary” and that ample alternatives are available for the expression being restricted.

3. Time, Place, or Manner Restrictions Protecting Entry and Exit at Specific Locations

The U.S. Supreme Court has addressed restrictions designed to protect entry and exit at particular locations in cases involving abortion clinics. Following Madsen, the Court in Schenck v. Pro-Choice Network of Western New York held that a “floating buffer zone” of fifteen feet

127. Id. at 1225.
128. Id. at 1226, 1229.
129. Id. at 1228. Subsequent to the bombing of the USS Cole in Yemen, it seems reasonable to question whether courts would still reject a security provision of this type. See Jan M. Olsen, 39 U.S. Sailors in German Hospital, The Associated Press, Oct. 14, 2000, available at 2000 WL 27906507. Order 3’s promulgation, however, was unrelated to terrorism in the sense of the USS Cole bombing; although many other security precautions surrounding the WTO conference did address terrorism, they are not the subject of this Comment.
130. Id. at 1227 (quoting Ward v. Rock Against Racism, 491 U.S. 781, 800 (1989)).
131. Id. (“The government bears the burden of proving that the narrowly tailored and alternative communication prongs are satisfied.”) (internal quotation marks omitted) (citing City of Watseka v. Ill. Pub. Action Council, 796 F.2d 1547, 1551 (7th Cir. 1986), aff’d without opinion, 479 U.S. 1048 (1987)); cf. Clark v. Cnty. For Creative Non-Violence, 468 U.S. 288, 293 n. 5 (1984) (“[I]t is common to place the burden upon the Government to justify impingements on First Amendment interests.”).
132. Id. at 1228, 1230.
133. See supra notes 90–97 and accompanying text.
134. 519 U.S. 357 (1997).
violated the First Amendment. The injunction in *Schenck* required that protesters move out of the way to stay fifteen feet from individuals entering a clinic. The Court also reaffirmed the *Madsen* rule, holding that while protecting entry into and exit from a clinic serves a substantial government interest, an injunction restricting First Amendment rights must be narrowly tailored to restrict “no more speech than necessary.”

In *Hill v. Colorado*, the Court upheld a Colorado statute making it a misdemeanor to “knowingly approach” within eight feet of a person entering or leaving a health care facility “for the purpose of passing a leaflet or handbill to, displaying a sign to, or engaging in oral protest, education, or counseling with such other person.” The Colorado Legislature found a significant state interest in “preventing the willful obstruction of a person’s access to medical counseling and treatment at a health care facility.” The Court noted that this interest was unrelated to the suppression of speech. In addition, the Court found that the restriction was narrowly tailored, because it provided only an eight-foot buffer zone and did not require a stationary protester to move out of the way if a clinic patron approached within eight feet. Finally, the statute left open adequate alternative methods for expression by allowing any protest, education, or counseling that did not entail moving within eight feet of a clinic patron.

The U.S. Supreme Court has constructed a solid foundation of constitutional rules underlying the First Amendment. First, restrictions on expressive activity cannot be justified through reference to the violence of counter-demonstrators. Second, even when certain protesters are violent, the government bears the burden of showing that the violence colored the entire protest movement before restrictions may be imposed. Third, prior restraints on speech are presumed unconstitutional, and must survive strict scrutiny by the courts. Fourth, restrictions that are facially content neutral are subject to less scrutiny, and are analyzed under the

135. *Id.* at 367.
136. *Id.* at 364.
137. *Id.* at 371 (citing *Madsen v. Women’s Health Ctr.*, 512 U.S. 753, 765 (1994)).
138. 120 S. Ct. 2480 (2000).
139. *Id.* at 2484 (quoting *COLO. REV. STAT.* § 18-9-122 (1999)).
140. *Id.*
141. *Id.* at 2491.
142. *Id.* at 2494.
143. *Id.*
tests formulated in O’Brien and Ward and explicated in the Ninth Circuit by Bay Area Peace Navy. The Court has specifically addressed restrictions designed to protect entry and exit at specific locations, and requires that they affect no more speech than necessary.

III. EMERGENCY POWERS AND THE FIRST AMENDMENT

Emergency powers allow the government to suspend constitutional rights temporarily.¹⁴⁴ In the American constitutional system, “[a] fundamental tension exists . . . between the basic premise of government constrained by law and the perceived need for unfettered, discretionary power to confront dire emergencies.”¹⁴⁵ In the context of the executive orders promulgated during the WTO protest, two features of modern emergency power¹⁴⁶ are of interest. First, the dichotomy between “normal” and “emergency” situations blurred during the twentieth century. The blurring has occurred at the federal level, but has also been exploited by state and local governments. Second, economic issues have risen to the level of an “emergency.” In reviewing modern emergency orders, many courts have applied a low level of judicial scrutiny. The Ninth Circuit, however, splits from other circuits and closely scrutinizes emergency orders promulgated in the context of mass demonstrations.

A. The Expansion of Emergency Powers

Changes in the world’s political and military structure beginning in the 20th century have fostered the gradual erosion of any meaningful dichotomy between “normal” and “emergency” power.¹⁴⁷ Increasing globalization and advancing military technology have greatly expanded the number and kinds of threats to which emergency response is necessary and perhaps reasonable.¹⁴⁸ As the pace of global politics has

¹⁴⁴ See, e.g., United States v. Chalk, 441 F.2d 1277, 1280 (4th Cir. 1971).
¹⁴⁸ See Lobel, supra note 145, at 1399–412.
increased, the political expediency of emergency powers has become more attractive to national executives.\textsuperscript{149}

As America rose to become a world military and economic leader, the President was granted increasingly broad authority by emergency legislation in Congress.\textsuperscript{150} With a few notable exceptions, the U.S. Supreme Court has acquiesced to the President’s exercise of emergency powers.\textsuperscript{151} After Japan attacked Pearl Harbor on December 7, 1941, the Court upheld the constitutionality of military orders relocating and detaining U.S. citizens of Japanese ancestry.\textsuperscript{152} Following World War Two, the “emergency” of the Korean War became the constant emergency of the Cold War.\textsuperscript{153} “Emergencies” in the last three decades of the 20th century included such national crises as the introduction of armed forces in Indo-China, Iran, Lebanon, Central America, Grenada, Libya, and the Persian Gulf.\textsuperscript{154} Congressional attempts to reform the emergency powers of the executive branch have been ineffective.\textsuperscript{155}

This blurring of the normal-emergency dichotomy at the federal level has made expanded emergency powers at the state and local levels more constitutionally palatable. State and local emergency statutes are generally written to allow the governor or mayor to respond to natural disasters or severe public disturbances.\textsuperscript{156} When the designated government official declares a state of emergency, these statutes allow, for example, the imposition of a curfew, the ban on sales of alcohol and weapons, and forced relocations to emergency shelters.\textsuperscript{157} Emergency statutes usually require that the state of emergency expire either within a certain time period absent further proclamation or as soon as the

\textsuperscript{149} \textit{Id.; see also} Gross, supra note 146, at 1858.

\textsuperscript{150} See Lobel, supra note 145, at 1400.


\textsuperscript{152} Korematsu v. United States, 323 U.S. 214, 217–18 (1944); Hirabayashi v. United States, 320 U.S. 81, 95 (1943).

\textsuperscript{153} See generally Lobel, supra note 145, at 1399–412.

\textsuperscript{154} See id. at 1414.

\textsuperscript{155} See id. at 1413–15 (describing National Emergencies Act).

\textsuperscript{156} See, e.g., \textsc{Wash. Rev. Code} § 38.52.020 (2000); \textsc{Seattle, Wash., Municipal Code} § 10.02.010 (1999); \textit{see also} \textsc{Cal. Gov’t Code} § 8625 (West 2001); \textsc{Kansas City, Mo., Municipal Code} § 50-155 (1998); \textsc{New York City, N.Y., Code} § 3-104 (2000).

\textsuperscript{157} \textit{See supra} note 156.
emergency ceases to exist.\textsuperscript{158} As discussed below, courts differ as to the stringency with which they review state and local emergency powers.\textsuperscript{159}

Emergency power in the United States has also evolved along with capitalist liberal democracy.\textsuperscript{160} As developed economies have grown more and more dependent on trade and capital flows, economic threats to the state have become more serious, rising even to the level of a military attack.\textsuperscript{161} By the 1930s, “the notion of the emergency situation was increasingly separated from any evidence of military conflict or armed rebellion whatsoever.”\textsuperscript{162} The powers granted to President Roosevelt during the Great Depression stand as “the largest single instance of delegated power in American history.”\textsuperscript{163}

The U.S. Supreme Court addressed economic emergency in an early case in \textit{Home Building & Loan Ass'n v. Blaisdell}.\textsuperscript{164} The \textit{Blaisdell} Court upheld Minnesota’s mortgage moratorium law.\textsuperscript{165} Enacted in response to the Great Depression, the law froze mortgage foreclosures in an alleged violation of the Contract Clause of the U.S. Constitution.\textsuperscript{166} While finding that “[e]mergency does not increase granted power or remove or diminish the restrictions imposed upon power granted or reserved,”\textsuperscript{167} the Court’s five-to-four decision upheld the Minnesota law.\textsuperscript{168} In a decision paradigmatic of the changing face of emergency power, Chief Justice Hughes’s opinion noted that “[t]he principle of harmonizing the constitutional prohibition with the necessary residuum of state power has had progressive recognition in the decisions of this Court.”\textsuperscript{169} \textit{Blaisdell}

\begin{footnotesize}
\begin{itemize}
\item\textsuperscript{158} See supra note 156.
\item\textsuperscript{159} See infra Part III.B.
\item\textsuperscript{160} William E. Scheuerman, \textit{The Economic State of Emergency}, 21 CARDOZO L. REV. 1869, 1875 (2000). Scheuerman suggests that the development of emergency powers has followed a four step process: (1) to impose martial law to counter labor strikes; (2) to directly manage the underlying economic crisis itself; (3) to prevent occurrence of economic emergency in the first instance; and (4) to provide a permanent measure of economic management. Id.
\item\textsuperscript{161} See id. at 1878–79; see also Gross, supra note 146, at 1863 (noting that free trade has become national-security concern).
\item\textsuperscript{162} See Scheuerman, supra note 160, at 1878.
\item\textsuperscript{163} CLINTON L. ROSSITER, \textit{CONSTITUTIONAL DICTATORSHIP: CRISIS GOVERNMENT IN THE MODERN DEMOCRACIES} 260 (Greenwood Press, Inc. 1979) (1948).
\item\textsuperscript{164} 290 U.S. 398 (1934).
\item\textsuperscript{165} Id. at 447.
\item\textsuperscript{166} Id. at 415.
\item\textsuperscript{167} Id. at 425–16.
\item\textsuperscript{168} Id. at 447–48.
\item\textsuperscript{169} Id. at 435.
\end{itemize}
\end{footnotesize}
underscored the increasing acceptance of the use of executive powers to confront economic problems.

B. Courts Apply Divergent Standards in Reviewing Local Emergency Orders

In reviewing emergency orders, courts have imposed various levels of judicial scrutiny. For example, the Fourth Circuit refrains from second-guessing emergency orders, applying only minimal due process scrutiny. In *United States v. Chalk* the Mayor of Asheville, North Carolina, imposed a nighttime curfew following race riots at Asheville High School. The plaintiffs, convicted of possessing a weapon and driving a vehicle in violation of the curfew, argued that the statute was vague and overbroad, and that the supposed threat to public safety was insufficient to justify the restrictions imposed by the mayor. While rhetorically suggesting that the use of emergency powers to curtail expressive activity is subject to judicial review, the Fourth Circuit refused to scrutinize the mayor’s order. Finding that it would be “highly inappropriate for us . . . to substitute our judgment for his,” the court limited its scope of review to determining whether the mayor acted in good faith and had some factual basis for his decision. Thus, the Fourth Circuit failed to apply even intermediate scrutiny to the order, opting instead for a kind of minimal due process scrutiny.

Other courts have cited *Chalk* in reviewing emergency orders promulgated in the wake of natural disasters and riots. While concluding that emergency orders must not be vague or overbroad, nor encourage arbitrary enforcement, the courts have upheld the laws and orders in question. The rule in Washington is functionally equiv-

---

170. The U.S. Supreme Court has not addressed this issue.
171. 441 F.2d 1277 (4th Cir. 1971).
172. Id. at 1278.
173. Id. at 1280.
174. Id. at 1280–81.
175. Id. at 1281.
176. See generally TRIBE, supra note 98.
178. See, e.g., Avino, 91 F.3d at 110; Juan C., 28 Cal. App. 4th at 1103.
When communities are threatened by natural disasters, courts are understandably hesitant to second-guess the good-faith efforts of local officials. In the context of public disturbances, however, the Ninth Circuit does not follow the “minimal due process scrutiny” rule in Chalk. In Collins v. Jordan, the Ninth Circuit refused to grant immunity to San Francisco officials who imposed a city-wide no-protest ban after violent demonstrations following the Rodney King verdict. The Collins court distinguished between a nighttime curfew and a daytime ban on expressive activities, finding that they were “entirely different matter[s]” and holding that no reasonable police officer would believe that the daytime restriction was legal.

In addition, Collins established that emergency orders must be tailored to address current crises, and cannot be used to preempt the future exercise of First Amendment rights. The Collins court held as a matter of law that “the occurrence of limited violence and disorder on one day is not a justification for banning all demonstrations, peaceful and otherwise, on the immediately following day (or for an indefinite period thereafter).” The Collins court also suggested that emergency orders should not be used as prior restraints on First Amendment activities, noting that when unlawful conduct and First Amendment activities are intertwined, the illegal activities should be punished after their occurrence so as to avoid infringing on constitutional rights.

Collins puts the Ninth Circuit at odds with the Fourth Circuit in Chalk, which examined only whether executive action was taken “in good faith.” Collins shows that it is reasonable to apply a different standard of review to emergency orders where the “emergency” is not a hurricane or a volcano, but the mass exercise of First Amendment rights. In other

180. See, e.g., Arino, 91 F.3d at 108, 110 (declining to hold that Florida curfew was overbroad, even though it was in effect for twelve weeks after Hurricane Andrew made landfall in Florida).
181. 110 F.3d 1363 (9th Cir. 1996).
182. Id. at 1366, 1378.
183. Id.
184. Id. at 1372–73.
185. Id. at 1372.
186. Id. at 1372–73.
First Amendment and Nonviolent Protests

words, *Collins* directs courts to scrutinize emergency orders that operate like restraints on the exercise of First Amendment rights.

IV. ORDER 3: AN EXAMPLE OF UNCONSTITUTIONAL RESTRICTIONS ON MASS DEMONSTRATIONS

Restrictions on mass protests—even those promulgated during a declared state of emergency—must meet certain criteria to withstand a First Amendment challenge, especially in the Ninth Circuit. When executives point to clogged streets and broken windows to justify closing the public forum to protestors, courts must determine whether First Amendment rights are being unduly trampled. Order 3, promulgated by Mayor Schell in response to the Seattle WTO protests, exemplifies a restriction that should not survive such scrutiny. An analysis of Order 3 provides a template for how courts should address executive emergency powers stemming from mass demonstrations. First, Order 3 contravenes the holding of *Gregory*, where the Court concluded that police may not abdicate their responsibility to ensure a peaceful forum for political expression through reference to potential or actual violence by third parties. Second, Order 3 fails the rule in *Claiborne*, because violence on the part of some protesters does not reduce the First Amendment protection of other, peaceful protesters. Third, Order 3 operated as a prior restraint on expression and fails the strict scrutiny test. Fourth, Order 3 is not a valid content-neutral restriction. Finally, as an emergency power, daytime protest bans fall squarely within the holding of *Collins* and are thus unconstitutional.

A. By Infringing on the First Amendment Rights of Peaceful Protesters, Order 3 Fails the Gregory Guidelines

In evaluating emergency orders that restrict First Amendment rights, courts must first consider *Gregory v. City of Chicago*. *Gregory* held that (1) to the extent violent acts are carried about by groups separate from a peaceful protest movement, such acts do not lessen the First Amendment rights of peaceful protesters and (2) the rights of peaceful protesters may not be infringed on when police are unprepared for and unable to cope with violence marginal to the protest itself.188

Although the facts of \textit{Gregory} are distinguishable, Order 3 nonetheless fails the \textit{Gregory} guidelines. In \textit{Gregory}, peaceful protesters were ringed by a hostile audience of counter-demonstrators.\textsuperscript{189} The WTO protests in Seattle were larger than those in \textit{Gregory}, and it was other anti-WTO protesters, not a hostile audience, that threatened public order.\textsuperscript{190} \textit{Gregory}, however, still applies for two reasons. First, violent protesters can reasonably be treated as counterdemonstrators, similar to those in \textit{Gregory}, because the vast majority of demonstrators held nonviolence as a fundamental tenet of their expressive purpose.\textsuperscript{191} Second, \textit{Gregory} held that the violent action of other groups provided “no evidence” that the peaceful protesters had been disorderly.\textsuperscript{192} The notion that peaceful protest does not become disorderly conduct through the actions of others applies with equal force to the Seattle demonstrations.

Even to the extent that public disorder was threatened by counter-demonstrators, \textit{Gregory} maintained that the government cannot respond by closing the public forum to peaceful protesters.\textsuperscript{193} Seattle officials were not prepared for the demonstration that confronted them on November 30.\textsuperscript{194} The violence and property destruction that did occur, however, was not completely unexpected—the police were well informed of the activities of the Black Block and other potentially violent groups.\textsuperscript{195} Under \textit{Gregory}, the Seattle Police Department had at least some responsibility to direct its attention to the sources of violence and vandalism before closing off the public forum to peaceful protesters. Order 3 allowed the police to “punt”—authorizing the use of force to clear streets and arrest any protesters downtown. Order 3 contravenes \textit{Gregory} and is an unconstitutional infringement of First Amendment rights.

Admittedly, \textit{Feiner v. New York},\textsuperscript{196} which has not been explicitly overruled, is inconsistent with \textit{Gregory}. \textit{Feiner} holds that when an

\begin{itemize}
\item \textsuperscript{189} \textit{Id.}
\item \textsuperscript{190} \textit{See id.; Police Summary of Events, supra note 1.}
\item \textsuperscript{191} \textit{See Postman & Carter, supra note 6.}
\item \textsuperscript{192} \textit{Gregory}, 394 U.S. at 112.
\item \textsuperscript{193} \textit{Id.}
\item \textsuperscript{194} \textit{See Postman & Carter, supra note 6.}
\item \textsuperscript{195} \textit{See supra Part I; cf. Operations Panel Report, supra note 10, at 23 (noting that police “failed to take adequate measures to prevent and deter . . . criminal activity”).}
\item \textsuperscript{196} 340 U.S. 315 (1951).
\end{itemize}
unpopular protest message engenders a hostile response, police may restrict protesters’ speech to maintain public order. Applying Feiner to modern protests, however, leads to an absurd result by allowing a “heckler’s veto”—a situation where police may halt any demonstration when hecklers appear. Under Feiner, speakers disobeying police orders to stop speaking may be arrested and convicted, as long as police action is in response to the crowd reaction and not from a desire to suppress the speakers’ viewpoint. The U.S. Supreme Court is unlikely to support such an absurd application of its precedent.

Gregory provides the correct, modern standard and holds that police may not abdicate their responsibility to protect the peaceful exercise of First Amendment rights through their own unpreparedness. Unlike Feiner, Gregory looks at whether the protest activity itself is protected by the First Amendment. Under Gregory, Order 3 was invalid because it took away the rights of protesters who were peacefully exercising their First Amendment rights.

B. By Infringing on the First Amendment Rights of Substantially Peaceful Protest Groups, Order 3 Fails the Claiborne Test

Next, courts reviewing emergency orders must consider NAACP v. Claiborne Hardware Co. The Claiborne Court stated unambiguously that perpetrators of violent or illegal acts were not shielded by the First Amendment. Claiborne, however, also held that even when elements of the same protest commit violent acts, those acts do not characterize the entire protest for First Amendment purposes. The Claiborne decision set a high standard for deciding when the “entire collective effort” of a protest has been tainted by violence. Under Claiborne, violent action is treated as distinct from the underlying political cause, unless the violence

197. Id. at 320–21; see supra notes 51–56 and accompanying text.
199. See Wagner, supra note 50, at 219–22.
200. See supra notes 57–65 and accompanying text.
203. Id. at 933.
204. Id.
is so pervasive as to characterize the cause itself as a “violent conspiracy.” 205

Order 3 violated Claiborne by banning demonstrations downtown for three days after the mass demonstrations on November 30. Local officials may have been justified in taking action on November 30 in direct response to violent demonstrators. 206 The violence of November 30, however, did not legitimize the foreclosure of all First Amendment rights in the downtown core for the remainder of the WTO conference. Just as the owners of businesses targeted by the Claiborne boycott could not restrict peaceful activists from standing outside their stores through reference to other, violent activists, 207 the district court should not have permitted Mayor Schell to foreclose the right of peaceful protesters to demonstrate for the remainder of the conference.

Although the boycott at issue in Claiborne remains factually distinct from the WTO mass protests, the holding should still apply. The thrust of Claiborne is to protect protest movements from being characterized as violent through reference to “relatively few violent acts.” 208 In that sense, the Claiborne rule should apply to mass demonstrations like those occurring during the WTO ministerial in Seattle, where violence was sporadic rather than pervasive. Just as the Claiborne Court refused to hold peaceful boycotters responsible for the actions of “Black Hat” enforcers, 209 the peaceful WTO protesters should not have been forced to sacrifice their rights to the “Black Block” of violent anarchists. 210 Under Claiborne, peaceful protesters may not be held guilty by association of the acts of vandals and vigilantes. 211

C. Order 3 Operated as a Prior Restraint and Fails the Strict Scrutiny Test

When an emergency order acts as a prior restraint, a reviewing court must employ the strict scrutiny standard of New York Times Co. v.
United States. Order 3 does not meet the “heavy burden” of New York Times because it restricted expressive activity on days when the City could not show the existence of a significant threat to public order. The government’s interest in restoring public order on November 30 was significant, and implicated the First Amendment rights of WTO delegates and others attempting to travel downtown. The scales tip heavily against Order 3, however, in subsequent days. The rights of delegates to travel and to assemble were not threatened during December 1–3. Ample evidence suggests that most protesters would not return on subsequent days. For example, the labor march, which brought thousands downtown, was a one-day event.

While the City of Seattle relied on Madsen v. Women’s Health Center in defending Order 3, Madsen actually shows that Order 3 operated as a prior restraint. In Madsen, the Court found that an injunction did not operate as a prior restraint because it left open myriad alternative means for communicating the respondent’s anti-abortion message. Order 3, however, was not an “incidental” burden on WTO protesters. The protesters’ anti-WTO message was completely restricted from a fifty-block area of downtown Seattle, thereby precluding their presence in any area where they could potentially interact with WTO delegates. Just as Madsen struck down a 300-foot “no-approach zone” as restricting too much speech, so the District Court should have forced Mayor Schell to restrict only as much speech as necessary.

Additionally, Order 3’s duration points toward treating it as a prior restraint. Unlike the anti-abortion protests in Madsen, the WTO-protests were inevitably transitory in nature, since the ministerial meeting lasted only until December 3. The restriction of all protest activity in the

---

212. 403 U.S. 713 (1971).
213. Id. at 714 (internal quotation marks omitted).
214. See Chasan, supra note 41.
215. Cf. supra notes 23–26 and accompanying text.
216. See Barber, supra note 17.
217. See, e.g., Sunde, supra note 18.
219. See supra notes 90–97 and accompanying text.
220. Madsen, 512 U.S. at 764.
222. Madsen, 512 U.S. at 766.
223. See Police Summary of Events, supra note 1.
downtown core for the duration of the conference did not leave open alternative methods for protest, like protests that occurred after violence had subsided. Peaceful protests should have been allowed on December 1–4, because any restriction that foreclosed protest activity for the entire week could not have been narrowly tailored.

D. Order 3 Fails the Test for Content-Neutral Restrictions on First Amendment Activities

Even where emergency orders are not treated as prior restraints, courts must apply the rules for content-neutral restrictions on speech. While courts should consider the facts of each case when evaluating emergency orders, Order 3 provides a useful template for applying the content-neutral rules. Although Order 3 is facially content neutral, it fails the O'Brien test for restrictions that are generally applicable to speech and conduct. While Order 3 did not ban protest in the entire City of Seattle, it fails the Ward test for time, place, or manner restrictions on expressive activity in the public fora. Finally, Order 3 is unconstitutional under the U.S. Supreme Court cases that focus specifically on protecting entry and exit at specific locations.

1. Order 3 Fails the O'Brien Test

United States v. O'Brien sets forth a broad test for evaluating conduct-based restrictions on First Amendment rights. Order 3 passes the first two prongs of the O'Brien test—constitutional authority and significant government interest. Order 3 is unconstitutional because it fails the fourth prong, which requires that restrictions be narrowly tailored. In O'Brien, the Court found that a prohibition on destroying

224. Order 3, supra note 9. See generally Madsen, 512 U.S. at 763 (holding that injunction was content neutral even though it tended to affect those with only one viewpoint).
226. Id. at 377.
227. See supra Part II.D.1–2. It is within the power of the Seattle City Mayor to declare a state of emergency. SEATTLE, WASH., MUNICIPAL CODE § 10.02.010 (1999); see also WASH. REV. CODE § 38.52.020 (2000). Order 3 makes reference to the imminent necessity of "extraordinary measures to protect the public peace, safety and welfare" as required by the Seattle Municipal Code and state law of Washington. See Order 3, supra note 9.
228. O'Brien, 391 U.S. at 377. Whether Order 3 passes the third prong—the requirement that its purpose be unrelated to suppressing free expression—is disputed by several legal organizations. The ACLU, National Lawyers' Guild, Trial Lawyers for Public Justice, and Direct Action Network assert
First Amendment and Nonviolent Protests

draft cards restricted only as much expressive activity as was necessary to further the government’s interest in a smooth-running conscription system. Here, however, the government’s interest was served when the association and assembly rights of WTO delegates were no longer threatened, and when public order had been restored to downtown Seattle. Those criteria were met on or before Tuesday morning, December 1. By enforcing the no-protest zone for the next three days, Order 3 restricted more speech than necessary and was unconstitutional.

2. Order 3 Is Not a Valid Time, Place, or Manner Restriction

In the narrower context of public-fora speech regulations, Ward v. Rock Against Racism requires that time, place, or manner restrictions be “narrowly tailored” and leave open “ample alternative channels” for communication. Under Bay Area Peace Navy v. United States, governments in the Ninth Circuit bear the burden of proof on both issues. Order 3 also fails the narrow-tailoring test. In Bay Area Peace Navy, the Ninth Circuit held a seventy-five-yard safety zone unconstitutional because military officials could not show that the zone was necessary to protect military officials. With regard to Order 3, the City bears the burden of producing “tangible evidence” to show that its restrictions on expressive activity were not “substantially broader than necessary.” The City cannot show that prohibiting all protest activity in the downtown core on December 1, December 2, and December 3 was necessary to preserve the safety of WTO delegates or to prevent further public disorder; therefore, Order 3 is unconstitutional.

Even aside from the narrow-tailoring test, Order 3 is unconstitutional because it fails to provide ample alternative channels for the protestors to express their views.

that police enforcement of Order 3 singled out even the most subtle anti-WTO expression. See, e.g., Chasan, supra note 41; Paul Richmond, Waging War on Dissent, a Report by the Seattle National Lawyers Guild WTO Legal Group, at http://seattle.indymedia.org/local/images/NLG-REPORT.pdf (last visited Apr. 17, 2001). The Seattle City Council’s Review Panel agreed. Operations Panel Report, supra note 10, at 31–32. As applied, Order 3 was viewpoint-specific, and thus violated the First Amendment.

230. See Chasan, supra note 41.
232. 914 F.2d 1224, 1229 (9th Cir. 1990).
233. Id. at 1227.
234. Id.
communicate their message. In *Bay Area Peace Navy*, a regulation prohibiting the “Peace Navy” flotilla from coming within seventy-five yards of the official viewing area was held unconstitutional because the protesters’ audience would be unable to hear their singing at that distance. The situation in Seattle is analogous. When WTO protesters were banned from downtown, they were unable to effectively deliver their message. The protesters needed access to the downtown core to be effective; indeed, many groups applied for parade and event permits to protect their rights to demonstrate downtown. Because Order 3 leaves no adequate alternative method for protesters to deliver their message, it fails the *Bay Area Peace Navy* test.

Order 3 is unconstitutional notwithstanding the holding in *Ward*. According to *Ward*, regulations need not be the “least restrictive” possible to serve the government’s interest; the government must show only that its stated interest would be served less effectively absent a given regulation. The facts in *Ward*, however, remain distinct, and the ruling of the *Ward* Court is not applicable in the context of mass demonstrations. *Ward* addressed an ordinance requiring concert organizers to use city-owned sound equipment and a city sound technician during outdoor concerts. While volume was controlled, the music itself was never restricted or banned. Order 3, in contrast, totally banned protest activity within the no-protest zone.

3. **Order 3 Is Unconstitutional Under the U.S. Supreme Court’s Entry-Exit Cases**

Finally, Order 3 is unconstitutional under the Court’s abortion-clinic cases—directly analogous holdings governing entry into and exit from specific locations. These cases establish that protecting entry and exit can constitute a significant government interest, such that an incidental

235. Id. at 1229.
236. Id. at 1226, 1229.
240. Id. at 784.
241. See *id*.
impact on First Amendment rights is acceptable. The Court, however, has consistently scrutinized geographic restrictions to ensure that “no more speech than necessary” is restricted. A thirty-six-foot exclusionary zone around a building was held constitutional, as was an eight-foot zone around patients and clinic staff, as long as protesters were not forced to move out of the way when approached by a protected person. On the other hand, a fifteen-foot floating buffer zone was held unconstitutional, as were all zones more than thirty-six feet—including a 300-foot “no-approach zone” and a 300-foot zone around private residences that included a prohibition against blocking street access.

Order 3 exceeds all of the acceptable rules for regulating access to a specific location. Under these rules, the Seattle Police could have provided cordoned-off passageways for WTO delegates to make their way into the convention center. Following Hill v. Colorado, the police may also have forbidden that individuals approach within eight feet of another person “for the purpose of . . . engaging in oral protest . . . with [that] person.” Closing the entire downtown core to protesters for three days, however, restricted significantly more speech than necessary. Even if the City could show that such action was the only way to ensure access, there is no precedent for the closure of such a large area. Far from providing support for Order 3’s constitutionality, Madsen, Schenk v. Pro-Choice Network, and Hill show that Order 3 cannot survive First Amendment scrutiny.

In evaluating emergency powers as content-neutral restrictions on First Amendment rights, courts must apply O’Brien, Ward, Bay Area Peace Navy, and the abortion clinic entry-exit cases. If courts find that an emergency order is not narrowly tailored, or that it restricts more speech than necessary, the emergency order should be found unconstitutional. Order 3 is an example of an emergency order that fails both the rules in

245. Id. at 759; Hill, 120 S. Ct at 2484, 2499.
246. See supra notes 134–40 and accompanying text.
247. 120 S. Ct. 2480 (2000).
248. Hill, 120 S. Ct. at 2484 (quoting COLO. REV. STAT. § 18-9-122 (3) (1999)).
249. See supra note 45 and accompanying text.
250. 519 U.S. 357 (1997).
O’Brien and Ward and the specific guidelines set out in the entry-exit cases.

E. Order 3 Is Unconstitutional as an Emergency Order

Order 3 represents the coalescence of two trends in emergency powers. On the one hand, the use of executive action to restrict the large-scale exercise of First Amendment rights shows that the normal-emergency dichotomy has blurred. Without foreign aggression, a tsunami, or even a race riot, downtown Seattle was buttoned up and cordoned off. The “emergency” in Seattle was not an invasion or a natural disaster, but the coming together of large numbers of people to exercise First Amendment rights. On the other hand, the pressure on local officials to facilitate the ministerial meeting by curtailing peaceful protests shows that economic concerns are first-order priorities for modern governments. The preeminence of the WTO in the world economy positions Order 3 as a logical next step in exploiting the notion of “emergency” for the sake of political expediency. Indeed, mass demonstrations may create emergency situations for local governments. In responding to emergencies created by the large-scale exercise of First Amendment rights, however, local governments in the Ninth Circuit must follow Collins v. Jordan.

Under Collins, daytime curfews are presumptively unconstitutional. As the Collins court held, a nighttime curfew and a daytime no-protest zone are “entirely different matter[s].” In noting that violence had occurred during evening protests, the Collins court refused to allow the extension of a protest ban to the daytime hours of the following day. Unless the City can show that such a ban was the only way to maintain public order and protect the rights of WTO delegates, Order 3 is unconstitutional under Collins.

Additionally, the Collins holding explicitly connects the duration of emergency orders with the requirement that First Amendment restrictions be narrowly tailored. An emergency on one day does not provide a

---

251. See supra Part III.A.
252. 110 F.3d 1363 (9th Cir. 1996).
253. See id. at 1371.
254. Id. at 1374.
255. Id.
256. Id. at 1372.
First Amendment and Nonviolent Protests

basis per se for banning protests on subsequent days; such a ban “is presumptively a First Amendment violation.” The City cannot show that an “emergency” existed in downtown Seattle during any daytime hours on December 1–3.\textsuperscript{258} \textit{Collins} explicitly forbids the prophylactic use of no-protest zones, and under that decision Order 3 is unconstitutional.

The rule in \textit{United States v. Chalk}\textsuperscript{259} suggests the opposite result. \textit{Chalk} indicates that emergency orders will not be scrutinized where local officials act in good faith and have some factual basis for their actions.\textsuperscript{260} \textit{Chalk} should not be applied here because that case and its progeny address riots and natural disasters, not mostly peaceful political protests.\textsuperscript{261} Also, the \textit{Chalk} line of cases does not address daytime no-protest zones or other daytime bans on expressive activity.\textsuperscript{262} \textit{Collins}, aside from being the controlling authority in the Ninth Circuit, provides a better rule because it specifically addresses daytime curfews and political protest.

In refusing to grant the ACLU’s motion for a TRO against the enforcement of Order 3, Judge Bryan effectively gave the City of Seattle the right to institute its own injunction against nonviolent protesters. Judge Bryan’s hesitancy to second-guess the government’s interest in establishing order during the ministerial meeting is understandable. However, by denying the ACLU’s motion, the court violated the holding of \textit{Collins} and sanctioned the comprehensive denial of protesters’ First Amendment rights.

V. CONCLUSION

Mass protests have reemerged in the era of globalization, highlighting a dialectic of integration and alienation among environmentalists, human rights advocates, and labor interests. The media attention and public awareness attendant to these protests is a strong incentive toward their use in furthering political goals. Without First Amendment protection, these dissenting voices will be quieted. In a media environment

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{257} Id. at 1371.
\item \textsuperscript{258} See supra notes 35–39 and accompanying text.
\item \textsuperscript{259} 441 F.2d 1277 (1971).
\item \textsuperscript{260} Id. at 1281.
\item \textsuperscript{261} See id.
\item \textsuperscript{262} See id.
\end{itemize}
\end{footnotesize}
dominated by sound bites and scandals, activists without the tool of mass protest may go unnoticed and unrecognized.

Emergency Order 3, issued by Mayor Schell in the morning of December 1, 1999, violated the First Amendment to the U.S. Constitution. Instead of promulgating narrowly tailored emergency orders to protect WTO delegates and to prevent violence, Mayor Schell forced an end to peaceful protest demonstrations. Order 3 made peaceful protesters guilty by association, punishing them for the sporadic violence of the Black Block and other fringe groups. In addition, Order 3 restricted too much speech—it closed the public forum of downtown Seattle for three days, longer than was necessary to restore order and protect the WTO delegates.

The Seattle WTO protests indicate that a more developed and aggressive jurisprudence addressing these issues is necessary and timely. At the very least, courts should protect the rights of peaceful protesters by striking down government action that goes too far in trampling First Amendment rights. Litigation regarding Order 3 should reflect the need for consistent, meaningful protection of the freedom of speech. There must be a constitutional lesson learned from the “Battle in Seattle”: When authorities respond to mass protests, they must be diligent, deliberate, and careful not to restrict too much speech. What was not demonstrated in Seattle—but what this Comment attempts to show—is that police must coexist with protestors, public order must tolerate dissent, and emergency does not justify censorship.
KIM HO MA v. RENO: CLOAKING JUDICIAL ACTIVISM AS CONSTITUTIONAL AVOIDANCE

Matthew E. Hedberg

Abstract: In Kim Ho Ma v. Reno, the Ninth Circuit rewrote the plain language of § 241(a)(6) of the Illegal Immigration Reform and Immigrant Responsibility Act of 1996 (IIRIRA) to avoid a constitutional defect in the statute. Section 1231(a)(6) of Title 8 of the U.S. Code, which codifies § 241(a)(6) of the IIRIRA, authorizes the Attorney General to detain criminal aliens, or removable aliens posing a danger to the community or a danger of flight risk, beyond the statutory removal period if they have not been removed from the country. Under the guise of constitutional avoidance, the Ma court carved out an exception to this detention authority by prohibiting the Attorney General from detaining deportable aliens beyond the statutory removal period if the aliens’ removal will not be accomplished in the reasonably foreseeable future. Although courts may use the constitutional-avoidance canon of statutory interpretation to avoid substantial constitutional questions, courts may not rely on the canon when the statutory language and legislative intent are clear. The Ma court’s statutory interpretation cannot be squared with either the plain language or the congressional intent of § 1236(a)(6) that the Attorney General’s detention authority includes the discretion to determine which criminal aliens may be released back into the community pending removal from the United States.

Court watchers bustled with anticipation as the Ninth Circuit took up the case of Kim Ho Ma v. Reno.1 The constitutionality of indefinite detention of immigrants under the Illegal Immigration Reform and Immigrant Responsibility Act of 1996 (IIRIRA)2 was finally having its day in court. For hundreds of aliens detained pending deportation, this case marked the chance for constitutional redemption. For the Immigration and Naturalization Service (INS), this was the opportunity to receive judicial recognition of the constitutionality of its actions in carrying out the country’s immigration policies. Yet, the Ninth Circuit ducked the issue and cloaked its activist decision in the chameleon-like legitimacy of constitutional avoidance.

Because indefinite detention is an important issue in this country’s immigration policy, courts need to confront directly the constitutional requirements for detention of the hundreds of aliens awaiting deportation from this country.3 In Ma, the Ninth Circuit had an opportunity to

1. 208 F.3d 815 (9th Cir.), cert. granted, 121 S. Ct. 297 (Oct. 10, 2000).
3. Compare Elizabeth Larson Beyer, Comment, A Right or a Privilege: Constitutional Protection for Detained Deportable Aliens Refused Access or Return to Their Native Countries, 35 WAKE FOREST L. REV. 1029, 1031–32 (2000) (arguing that indefinite detention is limited by constitutional
address this facet of the United States’s immigration policy but unfortunately chose to dodge the question under the shroud of constitutional avoidance. Constitutional avoidance is appropriately used to interpret ambiguous statutes in such a way that avoids deciding a substantial constitutional question, so long as the statutory construction is not plainly contrary to congressional intent. *Ma* did not present a case susceptible to interpretation using constitutional avoidance because the IIRIRA clearly permits a post-removal-period detention of criminal aliens found to pose a threat of danger or flight.

This Note argues the *Ma* court misconstrued the IIRIRA by carving out an exception to the Attorney General’s detention authority. Part I reviews the detention and removal of aliens under this country’s immigration laws. Part II explores the background of the Immigration and Nationality Act (INA) of 1952 and case law interpreting its provision authorizing the Attorney General to detain aliens. Part III explains canons of statutory interpretation in immigration law, including the canon of constitutional avoidance. Part IV analyzes the facts, procedural history, reasoning, and holding in *Ma v. Reno*. Finally, Part V argues that the Ninth Circuit incorrectly used constitutional avoidance to create a judicially crafted time limit, stretching the federal detention statute far beyond its original meaning. The Ninth Circuit’s approach is unfortunate because it encroaches on Congress’s lawmaking authority and abdicates the judiciary’s role in striking down statutes that violate the Constitution.

I. AN OVERVIEW OF THE DETENTION AND REMOVAL OF ALIENS

The INA authorizes the Attorney General to detain both deportable and excludable aliens pending removal from the United States. Aliens in the United States may be removed—sent back to their countries of origin—for committing designated crimes. The deportability of an alien is determined in removal proceedings conducted by immigration judges

---


2. The Attorney General’s statutory authority is delegated to Immigration and Naturalization Service (INS) District Directors, but this Note will refer to the Attorney General for purposes of discussing the IIRIRA’s detention authority. See infra notes 69–72 and accompanying text.


4. *Id.* § 1227(a)(2).
and the Board of Immigration Appeals (BIA) in accordance with regulations of the Attorney General.\footnote{7} After an immigration judge has ordered an alien removed from the United States, the alien is taken into INS custody for a ninety-day “removal period” during which time removal should occur.\footnote{8} Detention during the ninety-day removal period is mandatory for aggravated felons.\footnote{9} Criminal aliens or other aliens posing a risk of community danger or flight, who are not removed during the removal period,\footnote{10} may be subject to indefinite detention while awaiting removal.\footnote{11} At the beginning of 2001, there were approximately 5000 aliens in INS detention who were being detained while awaiting removal from the United States.\footnote{12}

Courts have interpreted the scope of the Attorney General’s detention and removal authority to vary depending on whether the alien is “excludable” or “deportable.”\footnote{13} Excludable aliens are aliens who have not been lawfully admitted into the United States.\footnote{14} Under the “entry fiction,”\footnote{15} the Attorney General often grants excludable aliens “parole,” allowing them physically to enter the United States pending removal to their countries of origin without lawfully admitting them into the United States.\footnote{16} Deportable aliens, on the other hand, have been lawfully admitted into the United States as resident aliens but at the time of deportation have not naturalized into the country.\footnote{17} Courts have generally held that excludable aliens have no constitutional rights to be free from indefinite detention pending removal.\footnote{18} Conversely, courts have disagreed over the extent of constitutional rights possessed by deportable

\footnotetext{7}{Id. §§ 1228, 1229a; 8 C.F.R. §§ 3.0–.65 (2000).}
\footnotetext{8}{8 U.S.C. § 1231(a)(1)(A).}
\footnotetext{9}{Id. § 1231(a)(2).}
\footnotetext{10}{There are numerous reasons why an alien may not be removed during the ninety-day removal period, including the absence of a repatriation agreement. A repatriation agreement enables a country to return aliens who have been ordered removed to their country of citizenship, provided that country is party to the agreement. See, e.g., Dinger, supra note 3, at 1552 n.14.}
\footnotetext{11}{8 U.S.C. § 1231(a)(6).}
\footnotetext{12}{Lise Olsen, Seattle ‘Lifer’ Has His Day in Highest Court, SEATTLE POST-INTELLIGENCER, Feb. 20, 2001, at A1.}
\footnotetext{13}{See Dinger, supra note 3, at 1558–59.}
\footnotetext{14}{8 U.S.C. § 1182.}
\footnotetext{15}{Shaughnessy v. United States ex rel. Mezei, 345 U.S. 206, 215 (1953) (summarizing history of entry fiction).}
\footnotetext{16}{Alvarez-Mendez v. Stock, 941 F.2d 956, 961 n.4 (9th Cir. 1991).}
\footnotetext{17}{8 U.S.C. § 1227.}
\footnotetext{18}{Shaughnessy, 345 U.S. at 215.}
aliens.\textsuperscript{19} Despite the constitutional implications that flow from an alien’s status as deportable or excludable, immigration law now subjects all aliens ordered removed to the same statutory detention provision.\textsuperscript{20}

II. DETENTION OF ALIENS ORDERED REMOVED UNDER THE INA

Immigration law regulating the detention and removal of aliens has undergone many changes over the years. Immigration law has evolved from court-imposed time limits to Attorney General discretion over alien detention. The statutory and legislative history and court decisions interpreting the current detention statute demonstrate a policy favoring detention of aliens awaiting removal, tempered only by Attorney General discretion. As the statutory amendments to the INA reveal, Congress authorized, if not required, the Attorney General to detain aggravated felons until removal is accomplished.

A. Statutory History of the INA

The INA,\textsuperscript{21} which governs the deportability of aliens, is a patchwork of legislation that has undergone numerous revisions. Before 1952, the Immigration Act of 1917\textsuperscript{22} governed the detention of aliens subject to a final order of deportation.\textsuperscript{23} Under the 1917 Act, Congress established no time limit for accomplishing deportation; rather, the Attorney General continued to detain deportable aliens not released on bond until deportation was accomplished.\textsuperscript{24} In several cases where aliens could not

\begin{itemize}
\item \textit{Compare} Zadvydas v. Underdown, 185 F.3d 279, 297 (5th Cir. 1999) (holding that deportable alien has no greater rights than excludable alien), \textit{cert. granted}, 121 S. Ct. 297 (Oct. 10, 2000), \textit{with} Binh Phan v. Reno, 56 F. Supp. 2d 1149, 1155 (W.D. Wash. 1999) (holding that deportable alien’s right to substantive due process is not extinguished by final deportation order).

\item 8 U.S.C. § 1231(a)(6) (providing that aliens who are inadmissible to United States under § 1182 or deportable under § 1227 may be detained beyond ninety-day removal period); IIRIRA § 301, 110 Stat. 3009-575 to -579 (1996), amended the INA to refer to “excludable” aliens as “inadmissible” aliens, \textit{e.g.}, 8 U.S.C. §§ 1182, 1225, and established “removal” proceedings to take the place of “deportation” and “exclusion” proceedings, \textit{e.g.}, § 1229a. Because courts continue to use the “excludable” and “deportable” terminology, this Note will refer to aliens as excludable or deportable depending on their immigration status.

\item 8 U.S.C. §§ 1101–1537.


\item Id. §§ 19, 20, 21, 39 Stat. 889–91.

\item Id. § 19, 39 Stat. 889.
\end{itemize}
be deported for various reasons, the Ninth Circuit imposed a “reasonable time” limitation on detention that generally did not extend longer than four months.\(^{25}\)

In 1952, Congress amended the provisions of the INA governing the Attorney General’s detention of aliens subject to deportation.\(^{26}\) As codified, the 1952 amendments granted the Attorney General authority to carry out an alien’s deportation within six months following a final order of deportation.\(^{27}\) At the discretion of the Attorney General, the alien could either be detained or released during the six-month period.\(^{28}\) Federal courts strictly interpreted the six-month time limit as prohibiting any detention of deportable aliens beyond six months.\(^{29}\)

The court-imposed six-month limit on detention remained in effect until 1990, when Congress barred the release of aggravated felons, with limited exceptions, under final orders of deportation.\(^{30}\) The 1990 amendment made clear that mandatory detention of aggravated felons was required both while deportation proceedings were pending and after a final order of deportation, despite the former six-month limit.\(^{31}\) Congress also included an exception to the mandatory-detention provision that allowed the Attorney General to release on bond aggravated felons lawfully admitted for permanent residence\(^{32}\) after a determination that the alien was not a threat to the community and was likely to appear for immigration hearings.\(^{33}\) Thus, for the first time, the Attorney General had the authority to detain certain aliens beyond a set removal period.

In 1996, Congress closed the door even tighter on detained aliens awaiting removal by amending the INA with the Antiterrorism and

\(^{25}\) See, e.g., Wolck v. Weedin, 58 F.2d 928, 930–31 (9th Cir. 1932); Saksagansky v. Weedin, 53 F.2d 13, 16 (9th Cir. 1931); Caranica v. Nagle, 28 F.2d 955, 957 (9th Cir. 1928); see also United States ex rel. Ross v. Walls, 279 F. 401, 403–04 (2d Cir. 1922).


\(^{28}\) Id. § 1252(c).

\(^{29}\) See, e.g., Johns v. Dep’t of Justice, 653 F.2d 884, 890 (5th Cir. 1981); Castillo-Gradis v. Turnage, 752 F. Supp. 937, 941 (S.D. Cal. 1990); cf. Oguachuba v. INS, 706 F.2d 93, 96 (2d Cir. 1983).


\(^{31}\) Id.

\(^{32}\) Id.

\(^{33}\) Id.
Effective Death Penalty Act of 1996 (AEDPA). 34 The AEDPA required the Attorney General to take into custody an expanded category of criminal aliens subject to mandatory detention. 35 Significantly, the amendment repealed a 1952 statute that had permitted the Attorney General to release criminal aliens, including lawfully admitted aliens who were determined to pose no threat of danger to the community or danger of flight. 36 Thus, after the AEDPA, the Attorney General had no discretion to release any aggravated felon from detention. Five months later, Congress again revised the INA’s detention provisions through the IIRIRA, 37 which restored the Attorney General’s discretionary relief. Rather than mandating detention following the removal period, the IIRIRA permits the Attorney General to release criminal aliens who satisfy specified statutory requirements. 38

B. Legislative History of the IIRIRA

The legislative history of the IIRIRA illustrates a compromise between mandating detention of criminal aliens and increasing the Attorney General’s discretion to make detention determinations. When reviewing legislative proposals to amend the INA, Congress aimed to ensure that aliens ordered removed did not return to their communities. 39 Congress considered several different approaches to reform the removal of illegal and criminal aliens. 40

The House of Representatives originally considered House Bill 1915, which called for stricter standards governing the release of aliens convicted of aggravated felonies during and after removal proceedings. 41

The bill’s provisions required mandatory detention during the removal

35. Id.
40. Id. at 105.
period, but called for the mandatory release of deportable aliens not
removed following the removal period. 42 Providing the impetus for
House Bill 1915 was a report from the Inspector General of the Depart-
ment of Justice that found that the vast majority of aliens not detained
following deportation proceedings absconded and were not removed
from the United States, while detention facilitated removal. 43 The stated
purpose of the statutory reform was to eliminate the incidence of aliens
absconding after deportation proceedings 44 by increasing the detention of
aliens who are ordered removed. 45

The bill that ultimately became the IIRIRA—House Bill 2202 46—was
introduced to the House of Representatives Judiciary Committee on
August 4, 1995, in place of House Bill 1915. 47 House Bill 2202 contained
the same detention provisions that were in House Bill 1915. 48 Moreover,
the Committee Report explained that, in order to ensure a criminal alien
is removed from the country, detention following entry of a final order of
removal was necessary. 49 The House of Representatives subsequently
passed the bill on March 21, 1996. 50

Meanwhile, on April 24, 1996, Congress enacted the AEDPA. 51
During a floor debate on June 7, 1995, Senator Kennedy criticized the
Senate Bill that later became the AEDPA because the Bill required the
Attorney General detain the expanded class of criminal aliens even when
those aliens could not be returned to their home countries. 52 Despite the
criticism, the AEDPA was enacted into law. 53 After enactment of the
AEDPA, INS General Counsel David Martin testified on behalf of the
Department of Justice before a subcommittee of the House of
Representatives regarding the problems the INS was having under the
AEDPA. He called for restoration of the Attorney General’s discretion to

42. Id. § 305(3) (providing that inadmissible aliens were subject to continued detention following
removal period).
43. H.R. REP. No. 104-879, at 108.
44. Id. at 107.
45. Id.
47. H.R. REP. No. 104-879, at 118.
49. Id. at 160–61.
50. H.R. REP. No. 104-879, at 121.
51. The AEDPA originated from Senate Bill 735, 104th Cong. (1995). See supra notes 34–36 and
accompanying text.
52. 141 CONG. REC. 15,068 (1995).
release aliens “who cannot, despite INS’s best efforts, be removed—
provided they meet the earlier tests regarding dangerousness and flight
risk.”  
He also informed Congress that even with restored Attorney
General discretion, the INS “fully intends to hold in custody, for as long
as necessary, those [removable aliens] who are dangerous to the
community.”

With the impetus for immigration reform renewed in Congress, the
Senate passed House Bill 2202 on May 2, 1996. Originally, the Senate
version of the IIRIRA had similar provisions as the AEDPA. The
original Senate version expanded the class of deportable criminal aliens
subject to mandatory detention, while simultaneously restricting the
Attorney General’s release authority. In conference, however, the
Senate and House compromised on the different versions. As a result of
the conference modifications, detention beyond the removal period was
made discretionary, not mandatory. In addition, the conference mod-
ifications yielded the provision allowing for detention of deportable
aliens beyond the removal period if the Attorney General determined that
the alien posed a risk to the community or was likely to abscond before
removal if released. The Senate and the House finally reached
agreement on the Conference Report and the IIRIRA was signed into law
on September 30, 1996.

C. The IIRIRA Detention Statute: 8 U.S.C. § 1231(a)(6)

The most disputed portion of the IIRIRA amendments to the INA is
§ 1231(a)(6) of Title 8 of the United States Code. Section 1231(a)(6)
governs the detention beyond the removal period of excludable and
deportable aliens ordered removed from the country. The language of

54. Removal of Criminal and Illegal Aliens: Hearing Before the Subcomm. on Immigration and
Claims of the House Comm. on the Judiciary, 104th Cong. 15 (1996) (statement of David A. Martin,
INS General Counsel).
55. Id.
56. In passing House Bill 2202, the Senate substituted the text of another Senate immigration bill.
59. Id. at 53–54, 215–16.
60. Id. at 54, 215–16.
§ 1231(a)(6) explicitly provides for a post-removal-period detention of criminal aliens. Courts have divided on the permissible duration of detention beyond the removal period.

The IIRIRA amendments to the INA mandate that, following a final order of removal, the Attorney General must detain certain aliens, including aggravated felons, during a removal period of ninety days. The ninety-day removal period marks a shift from the former six-month removal period under the 1952 amendments to the INA. If removal cannot be accomplished during the removal period, § 1231(a)(6) preserves the Attorney General’s authority to detain aggravated felons thereafter by specifying that aliens “may be detained beyond the removal period.” Unlike the AEDPA mandatory detention regime, Congress granted the Attorney General discretion to decide whether to detain aliens for an unspecified period of time beyond the ninety-day removal period.

Regulations implementing the IIRIRA allow the Attorney General to delegate the discretionary release power to INS District Directors. INS District Directors consider nine non-exclusive factors in determining whether to exercise the discretionary release authority or maintain the alien in custody. According to the implementing regulations, aliens seeking release carry the burden of demonstrating that release from custody will not pose a danger to the community or flight risk. Section 1231(a)(6) is in line with the trend of bestowing on the Attorney General the authority to make individualized determinations of who will be

64. Id. § 1231(a)(2).
65. Id. § 1231(a)(1)(A).
67. 8 U.S.C. § 1231(a)(6) (emphasis added). The section provides:
An alien ordered removed [who is excludable or deportable based on criminal grounds, . . . or who has been determined by the Attorney General to be a risk to the community or unlikely to comply with the order of removal, may be detained beyond the removal period and, if released, shall be subject to [supervision].
Id.
68. Id.
70. Id. § 241.4(a)(1)–(9).
71. Id. § 241.4(a).
subject to continued detention based on community safety and flight risk.  

D. Cases Interpreting the Attorney General’s Detention Authority Under § 1231(a)(6)

The overwhelming majority of courts that have addressed the issue have interpreted the language of § 1231(a)(6) as explicitly authorizing the Attorney General to detain deportable aliens beyond the ninety-day removal period. Representative of these courts’ decisions is *Duy Dac Ho v. Greene*. In *Ho*, the Tenth Circuit found that § 1231(a)(6) “is not ambiguous,” and “places no time limit” on the Attorney General’s authority to detain aliens beyond the ninety-day removal period. The Tenth Circuit refused to read a time limit into the statute because to do so would override Congress’s will as expressed in the plain language of the statute. The Tenth Circuit reasoned that by expressly declining to limit the Attorney General’s detention authority in § 1231(a)(6), Congress unambiguously authorized the Attorney General to detain indefinitely beyond the removal period certain removable aliens who cannot be removed within the removal period.

The Fifth Circuit, in *Zadvydas v. Underdown*, also upheld the continued detention of a deportable alien whose deportation was indefinitely delayed pending acceptance by another country. The court

---

72. *See generally id. §§ 241.1–33* (providing regulations governing post-hearing detention and removal pursuant to 8 U.S.C. § 1231(a)(6)).


74. 204 F.3d 1045 (10th Cir. 2000).

75. *Id.* at 1056–57.

76. *Id.* at 1057.

77. *Id.* After concluding that the Attorney General possesses the authority under § 1231(a)(6) to indefinitely detain deportable aliens pending removal, the Tenth Circuit rejected the alien’s due process claims. *Id.* at 1059–60.

78. 185 F.3d 279 (5th Cir. 1999), *cert. granted*, 121 S. Ct. 297 (Oct. 10, 2000).

79. *See id.* at 287, 291.
did not question the district court’s holding that Congress clearly would have placed a time limit on the period of detention if it had so intended and, therefore, the continued detention of a deportable alien was consistent with the Attorney General’s statutory authority. The court of appeals upheld the finding that mandatory detention language coupled with the absence of a time limit on detention necessarily means that indefinite detention is authorized. These cases are representative of other courts’ decisions that § 1231(a)(6) explicitly authorizes post-removal-period detention of deportable aliens.

III. APPROACHES TO STATUTORY INTERPRETATION

Courts employ a number of methods when interpreting statutes. Chief among the statutory interpretation tools is the plain-meaning approach, whereby courts seek to give full effect to the statutory language. Where the statutory language is ambiguous, courts will construe statutes in such a way that avoids deciding substantial constitutional questions. Additionally, as a result of congressional and executive plenary power over immigration law, courts will defer to the reasonable statutory interpretation made by those coordinate branches of the government. Finally, courts seek to give statutes their intended effect by supplying meaning to each statutory word and refusing to render any part of a statute superfluous.

A. The Traditional Approach to Statutory Construction: The Plain-Meaning Rule

Statutory construction necessitates a full accounting of a statute’s text. When interpreting statutes, courts attempt to bring to fruition the intent of the enacting legislative body. Although determining intent is

81. Caplinger, 986 F. Supp. at 1025. Though it found indefinite detention to be statutorily authorized, the district court ultimately held that the alien’s detention violated his constitutional right to due process. Id. at 1027. The Fifth Circuit, however, rejected the substantive due process challenge to the alien’s detention. Zadvydas, 185 F.3d at 296–97.
82. See supra note 73.
often difficult, courts generally regard the plain meaning of the statute’s language as the primary indicia of intent. Only if the statute’s plain meaning is ambiguous do courts look to the statute’s history, underlying policy, structure, and other canons of statutory interpretation. Moreover, abiding by the plain-meaning rule necessarily means that courts should not alter a statute by reading words or elements into the statutory language that are not present on its face.

The plain meaning of the statutory language guides a court’s approach to statutory construction unless legislative history clearly dictates a different result. The underlying goal of this approach is to respect and enforce the plain meaning of the entire language composition in the statute, including underlying policy. To this end, the U.S. Supreme Court has cautioned that when interpreting statutes, courts are bound by the purpose Congress sought to achieve and the means it has selected in carrying out that purpose. The fact that the plain language of a statute appears to produce harsh results is not an invitation for courts to redraft the statute. Furthermore, in the specific context of immigration law, the Ninth Circuit has previously held that where the plain meaning of the statute is unambiguous, that meaning is to be given full effect.

B. The Canon of Constitutional Avoidance

Among the judicial tools used to interpret statutes is the canon of constitutional avoidance. This canon seeks to prevent interpreting

85. Id.
90. AT&T Corp. v. City of Portland, 216 F.3d 871, 876 (9th Cir. 2000).
91. MCI Telecomm. Corp. v. Am. Tel. & Tel. Co., 512 U.S. 218, 231 n.4 (1994); see also Daniel A. Farber, Statutory Interpretation and Legislative Supremacy, 78 GEO. L.J. 281, 292 (1989) (“When statutory language and legislative intent are unambiguous, courts may not take action to the contrary.”).
92. United States v. Locke, 471 U.S. 84, 91, 95 (1985) (holding that mining claims had been extinguished because claimants recorded their claims on December 31 instead of “prior” to December 31).
93. Coronado-Durazo v. INS, 123 F.3d 1322, 1324 (9th Cir. 1997).
94. For a discussion on the U.S. Supreme Court’s historical use of the canon of constitutional avoidance, see Adrian Vermeule, Saving Constructions, 85 GEO. L.J. 1945, 1948 (1997).
ambiguous statutes in ways that would create constitutional infirmity.\textsuperscript{95} Constitutional avoidance is founded on the prudential concern that constitutional issues should only be addressed when necessary and that because Congress is bound by its duty to uphold the Constitution, courts should presume that legislative enactments adhere to constitutional mandates.\textsuperscript{96}

Constitutional avoidance is used to avoid substantial constitutional questions only when the plain language\textsuperscript{97} of the statute or congressional intent is ambiguous.\textsuperscript{98} When the plain language of a statute collides with the Constitution, constitutional avoidance cannot save the statute.\textsuperscript{99} The U.S. Supreme Court has warned that the constitutional-avoidance canon “is not a license for the judiciary to rewrite language enacted by the legislature,”\textsuperscript{100} and “disingenuous” attempts to avoid a constitutional question are impermissible.\textsuperscript{101} In \textit{United States v. Locke},\textsuperscript{102} the Court eschewed the use of constitutional avoidance and confronted the constitutional question head-on because to do otherwise would contort the statute to the point of “disingenuous evasion.”\textsuperscript{103}

*C. Judicial Deference to Executive Branch Statutory Interpretation in Immigration Law*

1. *General Tenets of Judicial Deference*

When interpreting a statute that is silent or ambiguous, the role of the judiciary is not to impose its interpretation of a statute over that of another governmental branch’s reasonable interpretation, especially in an

\textsuperscript{97} DeBartolo, 485 U.S. at 575.
\textsuperscript{98} Miller v. French, 530 U.S. 327, 336 (2000).
\textsuperscript{99} Benjamin v. Jacobson, 124 F.3d 162, 177 (2d Cir. 1997).
\textsuperscript{101} George Moore Ice Cream Co. v. Rose, 289 U.S. 373, 379 (1933) (Cardozo, J.) (“Avoidance of a [constitutional question] will not be pressed to the point of disingenuous evasion . . . . [When Congress’s intent is clear, t]he problem must be faced and answered.”).
\textsuperscript{102} 471 U.S. 84 (1985).
\textsuperscript{103} Id. at 96, 110 (holding that due process was not violated when holders of unpatented mining claims who failed to comply with annual filing requirements forfeited their claims); see Ernest A. Young, \textit{Constitutional Avoidance, Resistance Norms, and the Preservation of Judicial Review}, 78 TEX. L. REV. 1549, 1550 (2000) (noting criticism of constitutional avoidance for ignoring congressional intent).
area that the coordinate branch is entrusted to administer. Moreover, when Congress is aware of the executive’s interpretation of a statute and does not alter this interpretation, while amending other portions of the same statute, courts infer that the executive’s interpretation is consistent with congressional intent and thus warrants judicial deference.

When deciding whether to defer to an executive agency’s interpretation of a statute, a court’s first step is to determine whether Congress has expressed its intent on the issue. If Congress has, “that is the end of the matter,” and both courts and agencies must defer to congressional intent. Courts employ the “traditional tools” used for interpreting the plain meaning of statutes in determining legislative intent for purposes of this inquiry.

If Congress has not directly expressed its intent on the question at issue, courts may not impose their own statutory interpretation over an agency’s reasonable interpretation. Pursuant to Chevron, U.S.A., Inc. v. Natural Resources Defense Council, Inc., courts seek to avoid the danger of venturing into areas of agency expertise, to which they owe special deference. Judicial deference is justified because agencies are in a better position than the courts to handle the exigencies of competing policy choices dealing with the public interest. Deference is also justified because of the agency’s greater familiarity with the ever-changing facts and circumstances surrounding the subjects regulated.

The U.S. Supreme Court has made clear that when given the choice between competing statutory interpretations, an agency’s statutory interpretation that is reasonable in light of conflicting policies warrants judicial deference unless the agency’s accommodation is clearly contrary to congressional intent.

104. Aragon-Ayon v. INS, 206 F.3d 847, 851 (9th Cir. 2000).
105. Greenhorn Farms v. Espy, 39 F.3d 963, 965 (9th Cir. 1994).
107. Id. at 842–43.
110. Id. at 844.
111. Id. at 866.
113. Chevron, 467 U.S. at 844–45.
2. Judicial Deference in Immigration Law

Case law interpreting the scope of the government’s immigration power recognizes the authority of the political branches to craft broad policy decisions over all matters relating to admission, exclusion, and deportation of aliens, subject only to limited judicial review. The U.S. Supreme Court has long recognized that the political branches generally enjoy broad power over immigration, including the “power to expel or exclude aliens.” One rationale for this recognition is that the executive branch is especially entitled to judicial deference in the immigration context, where sensitive foreign relations questions are implicated. In the specific context of the IIRIRA, the Court has found that many of the Act’s provisions are “aimed at protecting the Executive’s discretion from the courts—indeed, that can fairly be said to be the theme of the legislation.” Accordingly, the Ninth Circuit has recently held that, despite the seemingly severe result, executive action increasing the burden needed to stay a removal order was consistent with the policy goals of the IIRIRA to vest the Attorney General with broad discretion in carrying out the country’s immigration policy.

The Ninth Circuit has also interpreted Congress’s lack of restraint on the Attorney General’s detention authority to require that courts defer to the Attorney General’s decision to prolong detention for excludable aliens. In Barrera-Echavarria v. Rison, the Ninth Circuit refused to read a time limit into a detention statute that did not explicitly provide for the authority to detain indefinitely. The court based its decision on the fact that Congress was aware of the impediments to removal for some aliens yet refused to restrict the Attorney General’s detention authority. In fact, that panel examined the interplay of the INA’s various sections and held that in the absence of a statutory limitation “the statutory

120. Andreiu v. Reno, 223 F.3d 1111, 1116 (9th Cir. 2000).
121. 44 F.3d 1441 (9th Cir. 1995).
123. Id. at 1446.
scheme implicitly authorizes prolonged detention." 124 The \textit{Barrera-\textit{Echavarria}} court attributed its construction, in part, to the deference owed the Attorney General’s interpretation. 125 Nearly every circuit court to reach the issue has agreed with the result in \textit{Barrera-\textit{Echavarria}}. 126

In immigration law, the “plenary power” doctrine provides another source of judicial deference to the legislative and executive branches. Plenary power over immigration enables congressional and executive flexibility in adjusting policy choices to changing political and economic circumstances. 127 The U.S. Supreme Court adheres to the guideline that when a collateral branch of government has plenary power over an issue, courts must defer to the exercise of that authority so long as the bounds of the Constitution are not transgressed. 128 Both the Third and Fifth Circuits have emphasized that the government’s plenary power over immigration renders it “largely immune from judicial inquiry or interference.” 129 Thus, absent a finding of constitutional violation, courts should not overturn proper exercises of plenary power in the immigration context. 130

\section*{D. Presumption Against Superfluous Language in Statutes}

Courts presume that every provision in a statute is intended to have independent effect. 131 Courts seek to construe statutes so that the language Congress selected is not weakened through a construction that would render any word void or superfluous. 132 Instead, courts seek to

\begin{thebibliography}{99}


125. \textit{Id.} at 1444–48. In \textit{Barrera-\textit{Echavarria}}, the Ninth Circuit also relied on the “entry fiction” in upholding the constitutionality of indefinite detention of excludable aliens. \textit{Id.} at 1450; \textit{see also supra} note 18 and accompanying text.

126. \textit{See, e.g.}, Guzman v. Tippy, 130 F.3d 64, 65 (2d Cir. 1997); Gisbert v. U.S. Attorney General, 988 F.2d 1437, 1446, amended by 997 F.2d 1122 (5th Cir. 1993); Fernandez-Roque v. Smith, 734 F.2d 576, 582 (11th Cir. 1984); Palma v. Verdeyen, 676 F.2d 100, 103–04 (4th Cir. 1982).


\end{thebibliography}
adhere to the basic principle that each statutory word should, if possible, be read to preserve its operative effect in the statutory scheme. This presumption against surplusage helps to ensure that Congress’s intent is fully realized. For example, in Ratzlaf v. United States, the U.S. Supreme Court reversed a Ninth Circuit decision that essentially read one word out of the statute. In Ratzlaf, the court of appeals erroneously treated the “willfulness” requirement in a criminal statute as of no consequence. As the Court noted, “[j]udges should hesitate . . . to treat statutory terms [as surplusage] in any setting.” Similarly, in Gustafson v. Alloyd Co., the Court held that courts should avoid interpreting a statute in a way that “renders some words altogether redundant.” In Gustafson, the Court avoided interpreting the word “communication” in the Securities Act of 1933 in such a way that would effectively render other terms in the statute redundant and eliminate their meaning within the statute. Instead, the Gustafson Court held that the statute must be read in its entirety, thereby giving effect to the language Congress intended to include in the statute.

IV. **KIM HO MA v. RENO**

In Kim Ho Ma v. Reno, the Ninth Circuit was asked to construe the IIRIRA’s detention provision. The District Court for the Western District of Washington, while not questioning that § 1231(a)(6) authorized the continued detention of aliens, had held that indefinite detention violated the substantive due process rights of deportable aliens. In contrast, the court of appeals did not reach the constitutional grounds but determined that the Attorney General lacked authority under § 1231(a)(6) to detain deportable aliens beyond the removal period if removal will not occur in a “reasonable time.”

133. Id.
135. Id. at 149.
136. Id. at 140.
137. Id. at 140.
139. Id. at 574.
140. Id. at 574–75.
141. Id. at 574–75.
142. 208 F.3d 815 (9th Cir.), cert. granted, 121 S. Ct. 297 (Oct. 10, 2000).
143. Id. at 830–31.
A. Facts and Procedural History

Kim Ho Ma is a native and citizen of Cambodia. He lawfully entered the United States in 1985 as a refugee and became a lawful permanent resident in 1987. In 1996, he was convicted in a Washington state court of first-degree manslaughter as a result of participating in a gang-related shooting. His conviction made him removable as a "deportable" alien convicted of an aggravated felony.

Pursuant to Attorney General regulations, the INS took Ma into custody following his release by state authorities on June 6, 1997. Once in custody, the INS commenced removal proceedings against Ma. An immigration judge found Ma removable because of his conviction. Ma appealed this ruling to the BIA, and the BIA affirmed the immigration judge’s decision. The immigration judge concluded that Ma’s detention was authorized under the IIRIRA because he would be a danger to the community if released. Ma’s order of removal became final on October 26, 1998. The final order of removal extinguished Ma’s status as a lawful permanent resident and any legal right to remain in the country. Despite the final removal order, the Attorney General could not remove Ma within the ninety-day period because the United States had no repatriation agreement with Cambodia. Following the ninety-day removal period, during which Ma’s detention was mandatory, he was detained pursuant to § 1231(a)(6).

Ma filed a petition in the U.S. District Court for the Western District of Washington for a writ of habeas corpus challenging the con-

144. See id. at 818.
145. Id. at 819.
146. See supra note 17 and accompanying text.
147. Ma, 208 F.3d at 819; see 8 U.S.C. § 1101(a)(43)(F) (Supp. V 1999) (defining “aggravated felony” to include crime of violence for which term of imprisonment imposed is one year or more); Administrative Record at 48, Kim Ho Ma v. Reno, 208 F.3d 815 (C99-15/WD)
148. Ma, 208 F.3d at 819.
149. Ma, 208 F.3d at 819.
151. Ma, 208 F.3d at 819.
152. Administrative Record at 41.
153. Ma, 208 F.3d at 819.
155. Ma, 208 F.3d at 819. See supra note 10 for discussion regarding repatriation agreements.
157. See supra Part II.C.
Kim Ho Ma v. Reno

stitutionality of his continued detention. Following an evidentiary hearing, the district court concluded that continued detention violated Ma’s Fifth Amendment substantive due process rights. Although the district court did not dispute that § 1231(a)(6) authorizes continued detention, it granted Ma’s petition for a writ of habeas corpus on September 29, 1999, and ordered his immediate release. The INS appealed the district court judgment to the Ninth Circuit Court of Appeals, which affirmed the district court’s judgment.

B. Holding and Reasoning

Although the Ninth Circuit affirmed the district court’s judgment granting Ma habeas corpus relief, the court did not reach the merits of the district court’s due process ruling. The court of appeals noted that the central issue addressed by the parties’ arguments, as well as the basis for the district court’s decision, was indeed the constitutionality of the INS’s detention policy. Yet without either the benefit of briefing by the parties or consideration by the district court on the statutory question, the court of appeals based its ruling on a construction of the statute. The court disagreed with the vast majority of other federal courts that considered the issue and concluded that because § 1231(a)(6) does not specify a time limit for continued detention beyond the removal period, a “reasonable time” limitation on detention should be read into § 1231(a)(6).

The Ma court determined that § 1231(a)(6) was ambiguous with regard to the length of time detention was authorized beyond the removal period and that this ambiguity raised a substantial constitutional

160. Kim Ho Ma v. Reno, No. C99-151L (W.D. Wash 1999), aff’d, 208 F.3d 815, 818 (9th Cir.), cert. granted, 121 S. Ct. 297 (Oct. 10, 2000) (granting Ma habeas relief). The court followed a previously established legal framework, which addressed the substantive and procedural due process challenges brought by aliens subject to § 1231(a)(6) detention following final orders of removal. Id. at 3; see also Binh Phan v. Reno, 56 F. Supp. 2d 1149, 1156–58 (W.D. Wash. 1999).
161. Ma, No. C99-151L.
163. Id. at 820.
164. See id.
165. See supra note 73 and accompanying text.
166. Ma, 208 F.3d at 830.
question. Specifically, the court reasoned that § 1231(a)(6) was ambiguous because if Congress had intended to authorize indefinite detention of deportable aliens, it would have made a “clear statement to that effect.” Because Ma was a deportable alien entitled to Fifth Amendment due process protection, the court reasoned that this ambiguity might have constitutional implications. Therefore, the court sought to read the statute in such a way as to avoid considering the constitutionality of indefinite detention. The Ma court acknowledged that § 1231(a)(6) was like the similarly worded detention statute in *Barrera-Echavarria* in that they both unambiguously authorized the Attorney General to continue detention of certain aliens beyond the removal period. However, the Ma court tried to distinguish *Barrera-Echavarria* as applicable only to excludable aliens while trying to avoid ruling on the constitutionality of indefinite detention of deportable aliens.

The court tried to navigate these concerns by applying the constitutional-avoidance canon and reading into § 1231(a)(6) a reasonable time limitation. The court reasoned that Congress would have included express language authorizing “indefinite detention” if it had so intended. Moreover, the court believed that reading a “reasonable time” limitation into § 1231(a)(6) was consistent with its earlier interpretation of the differently worded 1917 Act, despite the unclear reasoning of earlier cases interpreting the 1917 Act. Although it noted that the 1917 Act was not exactly like § 1231(a)(6), the court found the absence of an express time limitation in both statutes analogous. Finally, in dicta, the court explained that its interpretation was consistent with international law.

Despite INS protest to the contrary, the court of appeals concluded that no reasonable likelihood existed that Ma would be removed to

---

167. *Id.* at 821–22, 827.
168. *Id.* at 822.
169. *Id.* at 825.
170. *Id.* at 827.
171. 44 F.3d 1441 (9th Cir. 1995).
172. *Ma*, 208 F.3d at 824–25 n.20; *see also Barrera-Echavarria*, 44 F.3d at 1445.
174. *Id.* at 828 n.25.
175. *Id.* at 822, 829; *see supra* notes 22–25 and accompanying text.
176. *Ma*, 208 F.3d at 829.
177. *Id.* at 830 (noting that courts generally construe statutes to avoid violating international laws that prohibit prolonged and arbitrary detention unless Congress has enacted law to contrary).
Kim Ho Ma v. Reno

Cambodia. The court of appeals reasoned that the absence of a repatriation agreement between the United States and Cambodia providing for the return of each country’s nationals was sufficient proof to demonstrate that Ma’s removal would not occur in the reasonably foreseeable future. Therefore, the court ruled that the INS could no longer detain Ma.

V. THE KIM HO MA v. RENO COURT IMPERMISSIBLY INVOKED THE DOCTRINE OF CONSTITUTIONAL AVOIDANCE

The Ninth Circuit’s holding that the Attorney General’s detention authority under § 1231(a)(6) cannot exceed the ninety-day removal period when it appears that an alien cannot be removed in the “reasonably foreseeable future” is unjustified. The court of appeals’s decision is contrary to the text of § 1231(a)(6), the structure of the INA, and the legislative and statutory history of the IIRIRA. Furthermore, the executive branch’s reasonable interpretation of § 1231(a)(6) warrants judicial deference, and the court’s unwarranted intrusion on the executive branch’s reasonable statutory interpretation contradicts the policy considerations underlying the IIRIRA. By ignoring the plain meaning of § 1231(a)(6), the Ma court has sheltered a constitutional deficiency at the heart of indefinite detention.

A. The Ma Court Improperly Construed § 1231(a)(6)

The Ninth Circuit’s construction of § 1231(a)(6) rewrites unambiguous statutory language that has historically permitted the Attorney General to detain criminal aliens under final orders of removal who threaten community safety or present a flight risk. Nearly every court that has interpreted § 1231(a)(6) has concluded that the plain language of the statute unambiguously authorizes the Attorney General to detain beyond the removal period those aliens he or she deems unfit for release. The court’s ruling cannot be reconciled with the congressional intent manifested in the text of § 1231(a)(6), the structure of the INA, and the statutory and legislative history of the IIRIRA.

178. Id. at 831.
179. Id.
180. Id.
181. See supra note 73 and accompanying text.
1. The Ninth Circuit’s Construction of § 1231(a)(6) Conflicts with the Text of the Statute

The Ninth Circuit’s construction of § 1231(a)(6) in Ma cannot be reconciled with the plain meaning of the statute. The court’s interpretation of § 1231(a)(6) conflicts with the literal text of the statute and creates a requirement not found in the statute’s plain language. In addition, the court’s construction fails to give every provision of § 1231(a)(6) independent effect and introduces a contradiction into the statute.

The court’s restriction on indefinite detention fails to recognize the plain language of the statute. Although § 1231(a)(6) does not use the language of “indefinite detention,” the statute contemplates the prospect of indefinite detention by the use of the phrase “may be detained beyond the removal period and, if released, shall be subject to [supervision].”182 The Fifth and Tenth Circuits have concluded that § 1231(a)(6)’s “may be detained beyond the removal period” language places no time limit on the Attorney General’s detention authority and unambiguously authorizes detention beyond the removal period.183 Moreover, because “if” is defined as “in the event that,”184 the statute considers release from detention an alternative that may not occur. Indeed, the detention power is only qualified by the twin considerations of community safety and flight risk.185 In contrast to the plain language, the Ma court’s interpretation requires removal within the “reasonably foreseeable future,” and thereby ignores the substantial discretion that Congress vested in the Attorney General. The fact that such discretion poses a risk of constitutional violation suggests a problem with the statute itself, not merely with the way in which it might be read.

By reading an implicit limitation into the Attorney General’s detention authority, the Ma court violated the U.S. Supreme Court’s warning not to create statutory requirements that do not appear in the plain language of the statute.186 As written, § 1231(a)(6) does not place a limit on the length of time beyond the removal period that aliens may be detained. Instead, § 1231(a)(6) commits to the Attorney General’s discretion the

183 Duy Dac Ho v. Greene, 204 F.3d 1045, 1057 (10th Cir. 2000); Zadvydas v. Underdown, 185 F.3d 279, 287, 297 (5th Cir. 1999), cert. granted, 121 S. Ct. 297 (Oct. 10, 2000).
decision to continue detention beyond the removal period.\textsuperscript{187} The \textit{Ma} court, however, created a new requirement that limits the Attorney General’s discretion and requires release when there is a reasonable likelihood that an alien will not be removed in the reasonably foreseeable future.\textsuperscript{188} Congress, however, specified that the exercise of this discretion should be guided only by concerns for community safety and flight risk. As a result, the \textit{Ma} holding creates considerable tension with the text of § 1231(a)(6) by adding words to the statute that do not appear on its face.

The Ninth Circuit’s statutory interpretation also violates the longstanding rule that courts should seek to give every statutory provision independent effect.\textsuperscript{189} First, § 1231(a)(6) provides that those aliens who have not been removed during the removal period and who the Attorney General determines pose a community danger or flight risk “may be detained beyond the removal period.”\textsuperscript{190} By ordering the release of aliens who cannot be removed in the “reasonably foreseeable future,” the \textit{Ma} decision abrogates this explicit authorization for the Attorney General to detain beyond the removal period. Second, § 1231(a)(6) commands that “if released, [those aliens not yet removed] shall be subject” to supervision.\textsuperscript{191} Once again, however, the \textit{Ma} decision denies the independent effect of “if released” by rewriting the statute to read “when released.” The \textit{Ma} court’s statutory interpretation controverts the longstanding principle that statutory construction should not render any statutory word superfluous.

By stretching § 1231(a)(6) beyond its plain meaning, the \textit{Ma} court introduced a contradiction into the statutory scheme. On the one hand, the \textit{Ma} court recognized the Attorney General’s statutory authority to detain an alien during the ninety-day statutory removal period, but on the other hand \textit{Ma} requires release from detention when removal is not on the foreseeable horizon. A contradiction left unexplained in the Ninth Circuit’s reasoning is why detention is permitted at all in the case of an alien whose removal is not likely in the “reasonably foreseeable future.” Why should an alien who is ordered removed be subject to the mandatory ninety-day detention if there is not a repatriation agreement in place and at the end of the removal period release will be required? For

\textsuperscript{187} See supra note 73 and accompanying text; see also Part II.D.

\textsuperscript{188} Kim Ho Ma v. Reno, 208 F.3d 815, 818–19 (9th Cir.), cert. granted, 121 S. Ct. 297 (Oct. 10, 2000).

\textsuperscript{189} See supra Part III.D.

\textsuperscript{190} 8 U.S.C. § 1231(a)(6).

\textsuperscript{191} Id. (emphasis added).
the class of aliens who cannot be removed in the “reasonably foreseeable future,” the Ninth Circuit’s decision appears to transform the ninety-day removal period into nothing more than detention for detention’s sake. Such a punitive transformation of the removal period is clearly contrary to the statutory scheme’s proposition that immigration detention is not punishment but is instead an administrative incident to deportation proceedings.\textsuperscript{192} Furthermore, this punitive transformation of the removal period contradicts the Ninth Circuit’s avowed intention to avoid arbitrary detention.\textsuperscript{193} The Ninth Circuit failed to explain the incongruous results of its statutory interpretation.

2. \textit{The Ninth Circuit’s Construction of § 1231(a)(6) Conflicts with the Structure of the INA}

The Ninth Circuit’s construction of § 1231(a)(6) ignores the framework Congress established for carrying out the country’s immigration policies. The enactment of other INA provisions governing detention of criminal aliens establishes that Congress intended detention, not mandatory release, to be the general aim of the INA. In carving out an exception to the Attorney General’s detention authority under § 1231(a)(6), the \textit{Ma} court assumed Congress did not take into account the prospect that removal may not occur in the reasonably foreseeable future.\textsuperscript{194} This assumption is unfounded. The structure of the INA reveals Congress’s awareness that some countries may refuse to accept the return of their nationals.

Within the structure of the INA, Congress explicitly recognized the possibility that some aliens may not be removed during the removal period. For example, in the subsection immediately following § 1231(a)(6), Congress enacted a law that restricts the employment opportunities of aliens ordered removed.\textsuperscript{195} Despite this prohibition, Congress enacted an exception for aliens the Attorney General finds “cannot be removed due to the refusal of all countries designated by the alien or under [§ 1231] to receive the alien,” or when “the removal of the alien is otherwise impracticable or contrary to the public interest.”\textsuperscript{196} Thus, when Congress intended different treatment for this class of aliens,

---

\textsuperscript{193} \textit{Ma}, 208 F.3d at 830.
\textsuperscript{194} \textit{Id}. at 827–28.
\textsuperscript{195} 8 U.S.C. § 1231(a)(7).
\textsuperscript{196} \textit{Id}.
Kim Ho Ma v. Reno

it explicitly did so. Accordingly, Congress would have made an exception to § 1231(a)(6) similar to the exception in § 1231(a)(7) if it had intended to restrict the Attorney General’s post-removal-period detention authority in § 1231(a)(6) with respect to aliens who cannot be removed in the “reasonably foreseeable future.”

Other sections of the INA reveal Congress’s recognition that some aliens ordered removed will not be returned immediately to their original country due to the refusal of some countries to accept the return of their citizens.197 For example, Congress passed transitional rules,198 following enactment of the IIRIRA, that restricted the release of criminal aliens including those who “cannot be removed because the designated country of removal will not accept [their return].”199 In addition, the IIRIRA amended a section of the INA authorizing the Secretary of State to handle the contingency of visa requests from citizens of countries that “den[y] or unreasonably delay[]” accepting the return of its own citizens.200 The clear recognition of, and provision for, aliens who cannot be removed due to another country’s refusal to accept the return of its nationals is evident throughout the INA. If Congress had intended to make similar exceptions in the post-removal-period context of § 1231(a)(6), it would have done so. Although Congress’s choice not to make such exceptions may have created a statute that produces due process violations, the Ma court should have addressed that choice, not the one it would have made.

3. The Ninth Circuit’s Construction of § 1231(a)(6) Conflicts with the Legislative History of the IIRIRA

The Ma decision deviates from explicit legislative history revealing congressional intent to grant the Attorney General broad discretion to determine which aliens to detain past the removal period.201 Congress’s stated intent for the current statutory regime is to increase detention of criminal aliens prior to removal.202 Given the explicit legislative purpose

197. See supra Parts II.A & B.
199. Id.
201. See supra Part II.B.
202. H.R. REP. NO. 104-879, at 107–09 (1997); see supra Part II.B.
in enacting the IIRIRA, the Ma court’s contention that Congress *implicitly* intended to place time limits on the Attorney General’s detention authority 203 is not borne out by the IIRIRA’s legislative history.

The Ma court ignored the legislative history demonstrating Congress’s knowledge of the potential for indefinite detention and its chosen course of action. A central theme to the Ma court’s analysis is that if Congress had intended “indefinite detention” it would have so stated.204 The Ma court ignored the contrary assertion that Congress could just as easily have limited the Attorney General’s detention authority if it so desired.205 In fact, the shift from the AEDPA’s mandatory detention regime to the IIRIRA’s restoration of discretion to the Attorney General reveals Congress’s awareness and acceptance of the possibility of prolonged detention following post-removal-period detention.206

The legislative compromise that spawned the IIRIRA also militates against the Ma court’s brand of judicial activism. The original House version of the IIRIRA required release of deportable aliens on expiration of the ninety-day removal period,207 while the Senate version mandated detention even after the removal period.208 Congress compromised on the post-removal-period detention by committing the detention determination to the Attorney General’s discretion.209 If Congress had intended to place a definite time limit on the detention of deportable aliens who cannot be immediately returned to their original country, then it would have adopted the original House bill. In construing § 1231(a)(6), the Ma court renegotiated a legislative compromise.

204. *Id.* at 828 n.25.
4. The Ninth Circuit’s Construction of § 1231(a)(6) Conflicts with the Statutory History of the IIRIRA

Ma’s reliance on decisions flowing from the early versions of the INA is misplaced, as the court overlooked the significance wrought by post-1990 INA amendments. In arriving at its interpretation of § 1231(a)(6), the Ma Court relied, in part, on an analogy between case law emanating from the Immigration Act of 1917 and § 1231(a)(6).\(^{210}\) The 1917 Act simply provided that aliens should be “taken into custody and deported.”\(^{211}\) Courts interpreting the 1917 Act did not make their reasoning clear; nevertheless, the statute was interpreted as requiring release from custody if an alien could not be removed within a reasonable period of time.\(^{212}\) Courts interpreting subsequent amendments to the INA established a six-month limit on detention.\(^{213}\) Beginning with the 1990 amendments to the INA, however, Congress clarified that detention of aggravated felons following a final order of deportation was not subject to the previous six-month limit.\(^{214}\)

The reliance on case law interpreting the INA’s 1917 version is misplaced because even the Ma court acknowledged that “these older cases [interpreting the 1917 Act] did not interpret a statute exactly like [§ 1231(a)(6)].”\(^{215}\) Section 1231(a)(6) is not subject to earlier limits on detention and explicitly authorizes detention “beyond the removal period.”\(^{216}\) Indeed, the series of amendments preceding § 1231(a)(6) have rejected court-imposed time limits on detention by mandating detention for criminal aliens.\(^{217}\) Given the clear distinction between the language of the 1917 Act and that of § 1231(a)(6), the court of appeals erred in drawing support for its interpretation from that line of older Ninth Circuit cases restricting detention under the 1917 Act to a reasonable time.\(^{218}\)

\(^{210}\) Ma, 208 F.3d at 828–29.
\(^{212}\) Ma, 208 F.3d at 829.
\(^{213}\) See, e.g., Oguachuba v. INS, 706 F.2d 93, 96 (2d Cir. 1983); Johns v. Dep’t of Justice, 653 F.2d 884, 890 (5th Cir. 1981); Castillo-Gradis v. Turnage, 752 F. Supp. 937, 941 (S.D. Cal. 1990).
\(^{214}\) See supra note 30 and accompanying text.
\(^{215}\) Ma, 208 F.3d at 829.
\(^{217}\) See supra Part II.A. (discussing AEDPA and IIRIRA).
\(^{218}\) See supra note 25 and accompanying text.
B. The Ma Court Should Have Deferred to the Attorney General’s Reasonable Interpretation of § 1231(a)(6)

The Ma court failed to adhere to well-settled principles of judicial deference, especially in immigration law, toward reasonable executive branch statutory interpretations.219 In the context of § 1231(a)(6), the Attorney General has interpreted the statute to authorize the detention of deportable aliens beyond the removal period without limits.220 The vast majority of courts that have reviewed the statute have shared the Attorney General’s interpretation.221 The Ninth Circuit, however, has carved out an exception to the Attorney General’s detention authority by giving the statute a limiting construction.222 Because immigration matters involve serious questions affecting international relations and foreign policy,223 a reasonable interpretation of § 1231(a)(6) by the Attorney General should have been afforded deference.224 This deference to the Attorney General’s statutory interpretation should have led the court instead to consider whether this interpretation was constitutional.

The court’s decision runs counter to its earlier decisions that refused to shackle the Attorney General’s detention authority.225 In refusing to substitute its judgment for that of the executive, the Ninth Circuit previously observed that “[r]eadings a time limit on detention [of excludable aliens] would risk frustrating the government’s ability to control immigration policy and relations with foreign nations.”226 The foreign policy considerations are the same whether the alien is excludable or deportable. The only difference is whether the alien has constitutional rights while in the United States.227

---

219. See supra Part III.C. The Ma court determined that the plenary-power doctrine’s applicability varies on a case-by-case basis and was subject to constitutional restraints. The court avoided the constitutional question in Ma, yet did not specify which constitutional restraints precluded the doctrine’s application in this case. 208 F.3d at 826 n.24.
221. See supra note 73 and accompanying text.
222. Ma, 208 F.3d at 830.
223. See supra Part III.C.
224. See supra note 118 and accompanying text.
225. See supra notes 121–26 and accompanying text.
226. Barrera-Echavarria v. Rison, 44 F.3d 1441, 1448 (9th Cir. 1995).
227. See supra notes 14–19 (discussing excludable and deportable aliens).
The *Ma* court eschewed the *Chevron* doctrine, which requires courts to defer to agency interpretations of statutes.\(^{228}\) Judicial deference to administrative interpretations, such as the INS’s interpretation of § 1231(a)(6), has been consistently employed when the scope of statutory authority involves reconciling conflicting policy choices that rest within the specialized knowledge of the agency.\(^{229}\) Given such an explicit guideline for the application of judicial deference to administrative agency statutory interpretation, the Attorney General’s and INS’s interpretation should not have been regarded lightly. The *Ma* court should have deferred to the Attorney General’s reasonable interpretation of § 1231(a)(6) and then exposed that interpretation to full constitutional scrutiny.

**C. The *Ma* Court’s Use of Constitutional Avoidance Ignored Congress’s Policy Preferences Contained in § 1231(a)(6) and Abdicated the Judiciary’s Role of Enforcing the Constitution**

By imposing its judgment over that of the Attorney General, the *Ma* court attempted to soften this nation’s decidedly harsh immigration policy. This attempt infringes on the executive branch’s ability to make foreign policy decisions and usurps Congress’s lawmaking authority. Moreover, this attempt compromised judicial authority in construing the Constitution.\(^{230}\) Congress clearly chose detention as a necessary part of the country’s immigration policy.\(^{231}\) The *Ma* court should have respected that choice and measured it against the Constitution.

Contrary to the *Ma* court’s characterization of its use of constitutional avoidance as an exercise of “judicial restraint,”\(^{232}\) the court used constitutional avoidance as a vehicle for its own brand of judicial activism. Constitutional avoidance cannot be used to reinterpret an unambiguous statute nor to trample the clear intent of Congress.\(^{233}\) Given

\(^{228}\) See supra note 114 and accompanying text. The *Ma* court found the *Chevron* doctrine inapplicable because substantial constitutional questions were raised. Kim Ho Ma v. Reno, 208 F.3d 815, 821 n.13 (9th Cir.), cert. granted, 121 S. Ct. 297 (Oct. 10, 2000).


\(^{230}\) See Marbury v. Madison, 5 U.S. (1 Cranch) 137, 176 (1803).


\(^{232}\) Ma, 208 F.3d at 822.

\(^{233}\) See supra Part III.B.
the unambiguous text and legislative intent of §1231(a)(6), the Ma court’s use of constitutional avoidance is untenable.\footnote{Whether §1231(a)(6) presents a substantial constitutional question is beyond the scope of this Note’s statutory argument.}

The Ma court should have analyzed §1231(a)(6) in the same manner that it analyzed the detention statute in Barrera-Echavarria. In Barrera-Echavarria, the Ninth Circuit refused to read a time limit into the detention statute for excludable aliens and ruled that indefinite detention was constitutionally permissible.\footnote{Barrera-Echavarria v. Rison, 44 F.3d 1441, 1444–50 (9th Cir. 1995); see also supra notes 121–24 and accompanying text.} This approach preserves both Congress’s role in writing the country’s immigration laws and the judiciary’s role in deciding the constitutionality of those laws. The Ma court, however, found §1231(a)(6) to be ambiguous despite the text and legislative intent of the statute and invoked the canon of constitutional avoidance to avoid determining the constitutionality of indefinitely detaining deportable aliens.\footnote{Ma, 208 F.3d 815 at 821–22.} The Ma court improperly distinguished its decision from Barrera-Echavarria because §1231(a)(6) unambiguously subjects both deportable and excludable aliens to the same detention statute. Far from “judicial restraint,” this judicial activism compromised both the legislative and judicial functions.

The Ma court disingenuously evaded the constitutionality of indefinitely detaining deportable aliens. By reading a reasonable time limit into §1231(a)(6) for deportable aliens, the court’s decision permits “excludable” aliens to be indefinitely detained under Barrera-Echavarria while carving out an exception for “deportable” aliens.\footnote{Id. at 825.} This result squarely conflicts with the IIRIRA’s unambiguous text that subjects all aliens ordered removed—“excludable” and “deportable”—to the same detention provisions of §1231(a)(6).\footnote{See supra note 20 and accompanying text.} If Congress intended to provide disparate treatment of “excludable” and “deportable aliens,” then it presumably would have so specified.\footnote{See Greenhorn Farms v. Espy, 39 F.3d 963, 965 (9th Cir. 1994).} The fact that §1231(a)(6)’s application may be harsh is not a license for the Ma court to disingenuously avoid a constitutional question by rewriting the statute’s plain language.\footnote{See United States v. Locke, 471 U.S. 84, 95 (1985).} As a result of the Ma court’s use of constitutional avoidance, the Attorney General’s detention authority under §1231(a)(6)
varies depending on whether an excludable or deportable alien is being detained. The Ma court’s statutory interpretation is a disingenuous attempt to avoid the constitutional question raised in Barrera-Echavarria\textsuperscript{241} and therefore does not justify the use of constitutional avoidance.\textsuperscript{242}

The fact that the court of appeals relied on the canon of constitutional avoidance is proof that the court lacked textual support for its interpretation of § 1231(a)(6). After all, if the text of § 1231(a)(6) ruled out the post-removal-period detention at issue in \textit{Ma}, then the court appeals could have simply rested its decision on a plain reading of the statute. Because § 1231(a)(6)’s unambiguous language\textsuperscript{243} and clear congressional intent\textsuperscript{244} definitively set the boundaries of the scope of the Attorney General’s detention authority, the \textit{Ma} court incorrectly used the canon of constitutional avoidance.\textsuperscript{245}

VI. CONCLUSION

Section 1231(a)(6) is properly viewed by its lack of limitations on the length of time the Attorney General may detain removable aliens. Based on its text, structure, and history, § 1231(a)(6) contemplates the prospect of indefinite detention. Indeed, the statute clearly states that the Attorney General may detain removable aliens beyond the removal period, and “if released,” the alien will be subject to the regulations provided. Given such unambiguous language, the \textit{Ma} court’s use of constitutional avoidance is unjustified. Because Congress chose indefinite detention as a part of the immigration policy of the United States, it was the court’s role to decide the constitutionality of that choice. Ultimately, the \textit{Ma} court’s interpretation of § 1231(a)(6) had the effect of a sieve, draining the statute of its intended meaning by passing it through a porous judicial authority.

This Note aimed to encourage judicial scrutiny of the grave consequences of indefinite detention. Lacking an authoritative voice in court, an alien will be subject to the patchwork caprice of various

\textsuperscript{241} See supra note 125.
\textsuperscript{242} See George Moore Ice Cream Co. v. Rose, 289 U.S. 373, 379 (1933).
\textsuperscript{243} See supra Part V.A.1.
\textsuperscript{244} See supra Parts V.A.1–4.
interpretations of § 1231(a)(6). Displacing the issue is not a responsible way to manage such an important aspect of American foreign and domestic policy. Unlike the Ninth Circuit in \textit{Ma}, courts confronting the issue of indefinite detention should refuse the cloak of constitutional avoidance and confront indefinite-detention statutes head-on.
THE DECLARATION OF INDEPENDENCE: A 225TH ANNIVERSARY RE-INTERPRETATION

Carlton F.W. Larson *

Abstract: The importance of the Declaration of Independence to American law has been obscured by dubious associations with natural rights jurisprudence. Legal scholars have therefore overlooked the numerous ways in which the Declaration is relevant to a host of legal issues. Ample textual and historical evidence demonstrates that the Declaration, not the Articles of Confederation or the Constitution, legally constituted the United States of America as a distinct nation in the world community. The Declaration was not the act of thirteen states declaring their individual independence, but the act of one American people announcing the birth of an American nation. Nor is the Declaration merely an abstract treatise on individual natural rights. The Declaration displays a deep concern for which forms of government will most effectively allow self-government to flourish. As such, the Declaration speaks in constitutional language that is far more precise than we are often led to believe.

The Declaration of Independence is 225 years old this month. All across the country, Americans will celebrate the Fourth of July with trumpet and song, with fireworks and good cheer. In the rarefied halls of the legal academy, however, such celebrations will undoubtedly be greeted with little more than an air of amused detachment. After all, we all know that the Continental Congress actually voted for independence on July 2, not on July 4. We all know that the Declaration was merely a propaganda paper of no legal significance whatsoever. We all know that its animating principle—natural law—is about as intellectually respectable as alchemy. We all know that there is no more certain way to get thrown out of court than to file a claim based on the Declaration of Independence. We all know, in short, that the Declaration is utterly irrelevant to our role as lawyers and as legal scholars.  

Note:


I am grateful to Akhil Reed Amar for his encouragement of this project and for his many thoughtful comments on my arguments. Pauline Maier graciously read the entire manuscript and I am thankful for her suggestions and support. Special thanks to David Armitage for lending me a copy of his forthcoming article and to the Historical Society of Pennsylvania for permission to quote from its manuscript collections. My thoughts on this subject have benefited from conversations with Steven Engel, Drew Hansen, Carl Larson, Robert Spoo, and John Turner. All errors, of course, are mine alone.

An earlier version of the Article was the co-recipient of the Benjamin Scharps Prize for best paper by a third-year law student at Yale Law School.

1. See, e.g., infra notes 15–20 and accompanying text.
This Article is an invitation to re-think our entire approach to the Declaration of Independence. I contend that a thorough understanding of the Declaration is critical to understanding major issues in American public law, and not simply in the trivial sense that the Declaration is some sort of “philosophical background” to the Constitution. The Declaration matters, and it is important that we bring to it the same level of critical analysis that we apply to the Constitution and to other legal texts.

This Article advances three main points. First, our view of the Declaration’s place in American law has been obscured by dubious judicial adventures in “natural rights.” The *Lochner* era produced a host of decisions linking the Declaration of Independence to extreme views of economic liberty. The lingering distrust of those decisions, and of the impulses that motivated them, has made legal scholars less willing to notice the numerous ways in which the Declaration has direct relevance to American law. These legal issues range from the determination of American citizenship to ownership of the seashore to the conditions under which the western states entered the union. Indeed, in at least two states, the “principles of the Declaration of Independence” appear to have been adopted as positive law.2 As a survey of the case law will demonstrate, the legal issues raised by the Declaration are numerous and complex and have been the subject of litigation for well over 200 years.

Second, the Declaration was not, as is often asserted, a declaration of thirteen states declaring their individual independence. Rather, it was the declaration of one American people declaring the existence of one American nation. It is therefore entirely appropriate to date the legal existence of the American nation from July 4, 1776, and not from the date of the Articles of Confederation or of the ratification of the Constitution. The American nation preceded both the Articles, which were merely declarative of that nation’s existence, and the Constitution, which “perfected” that nation. The Declaration alone legally constituted the United States of America as a distinct nation in the world community.

Third, the Declaration is not simply an abstract and irrelevant treatise on the purposes of government. Rather, it is deeply concerned with what I term “constitutional formalism,” that is, a careful and detailed attention to the structure of government and to the rule of law. Most significantly, the Declaration addresses at length what sort of “forms” are most likely

---

2. See infra notes 73–75 and accompanying text.
Declaration of Independence
to allow self-government to flourish. A close reading of the *whole* text of
the Declaration suggests that the “rights” about which the Declaration is
most concerned are not so much individual natural rights, but the
collective rights of the American people to self-government. George III’s
violation of these rights—the basic rights of self-government—is what
justified American independence from Britain. This reading rescues the
Declaration from the *Lochner*-esque oblivion to which it has been
consigned and restores self-government to its proper place as the
Declaration’s fundamental principle.

This Article is divided into four parts. Part I explores the
Declaration’s place in American law and in the legal academy. Part II
argues, based on substantial textual and historical evidence, that the
Declaration is best read as the act of one American people creating one
American nation. Part III explains the Declaration’s roots in
“constitutional formalism.” Part IV offers a brief summary and
implications for modern legal doctrine.

I. THE DECLARATION OF INDEPENDENCE AND AMERICAN
LAW

A. Prologue: A Nineteenth-Century Perspective

Four score and seven years after the adoption of the Declaration of
Independence, Abraham Lincoln rose at Gettysburg to praise the
Declaration as the governing statement of American political
philosophy. It is easy to forget that Lincoln’s enthusiasm for the
Declaration was far from unprecedented. Two years earlier, Lincoln’s
arch-nemesis, Chief Justice Roger Brooks Taney, had sternly lectured
Lincoln on the importance of the principles of the Declaration. In the
aftermath of the surrender of Fort Sumter, Lincoln had suspended the

3. On the Gettysburg Address and Lincoln’s views on the Declaration of Independence, see
Wills’s more extravagant claims, such as his assertion that Lincoln’s speech was “one of the most
daring acts of open-air sleight-of-hand ever witnessed by the unsuspecting,” *id.* at 38, have been
effectively rebutted by Pauline Maier. PAULINE MAIER, AMERICAN SCRIPITURE: MAKING THE
DECLARATION OF INDEPENDENCE xix–xx, 189–208 (1997) (tracing the ways in which Lincoln’s
speech tapped into larger cultural currents). If Lincoln’s invocation of the Declaration was such a
novelty, one wonders why, for example, in their oral arguments in the *Amistad* case twenty-two
years earlier, Roger Baldwin and John Quincy Adams repeatedly invoked the principles of the
Declaration of Independence as the founding principles of the American nation. *See* The *Amistad,*
40 U.S. 518, 549, 552, 556, 557 (1841).
writ of habeas corpus along the route between Washington and Philadelphia, and the army had imprisoned a Confederate sympathizer, John Merryman, in Baltimore’s Fort McHenry\(^4\) (the site of the battle that inspired Francis Scott Key to write “The Star-Spangled Banner”). In an opinion denying Lincoln’s power to suspend habeas corpus, Taney admonished the President that “[t]he constitution of the United States is founded upon the principles of government set forth and maintained in the Declaration of Independence.”\(^5\) One such principle was civilian control of the military, as evidenced by the Declaration’s charge that George III “had affected to render the military independent of, and superior to, the civil power.”\(^6\)

Taney’s invocation of the Declaration was not merely a clever riposte to Lincoln, who had himself invoked the principles of the Declaration throughout his political career;\(^7\) rather, the Declaration figured prominently in Taney’s view of the constitutional order. The most extensive discussion of the Declaration in a Supreme Court opinion is in none other than Taney’s infamous opinion in *Dred Scott v. Sanford*,\(^8\) in which Taney explained at great length why the Declaration did not resolve the issue of Dred Scott’s citizenship. Taney insisted that the words “all men” in the phrase “all men are created equal” simply could not be understood to apply to blacks.\(^9\) A contrary interpretation would have rendered the conduct of the men who drafted the Declaration “utterly and flagrantly inconsistent with the principles they asserted.”\(^10\) This was impossible, Taney argued, because “the men who framed this declaration were great men—high in literary acquirements—high in their sense of honor and incapable of asserting principles inconsistent with those on which they were acting.”\(^11\)

To Lincoln, the *Dred Scott* decision was an abomination. Taney’s opinion, Lincoln argued, did “obvious violence to the plain

---

5. *Ex parte* Merryman, 17 F. Cas. 144, 152 n.3 (C.C.D. Md. 1861) (No. 9487).
6. Id.
7. See, e.g., Donald, supra note 4, at 269, 277; Wills, supra note 3, at 100 (discussing Lincoln’s views on the Declaration).
9. See id. at 410.
10. Id.
11. Id.
unmistakable meaning of the Declaration’” such that if the Declaration’s framers were to “‘rise from their graves, they could not at all recognize it.’”12 Taney’s interpretation of the Declaration has been quite properly consigned to the dustbin of history, but it is worth noting the efforts he expended in trying to distinguish away the language of the Declaration. Taney could have easily argued that the Declaration had no relevance to this issue. Justice Curtis said as much in his dissent,13 and Justice McLean, in a separate dissent, ignored the Declaration completely.14 But for Taney, as for Lincoln, it was inconceivable to reach a decision contrary to the Declaration of Independence. Despite all their differences, Lincoln and Taney shared one deeply held belief—that a proper understanding of the Declaration of Independence was essential to understanding the Constitution and laws of the United States of America.

B. The Declaration at the Millennium

To modern constitutional scholars, this notion seems as quaint as one of Matthew Brady’s faded daguerreotypes. As the introduction to this Article suggests, the position of the Declaration of Independence in recent constitutional thought is one of utter and complete irrelevance. Benno Schmidt, the former dean of Columbia Law School and former President of Yale University claims, “American constitutional law is positive law, and the Declaration of Independence has no standing in constitutional interpretation whatsoever.”15 The most recent volume of Laurence Tribe’s treatise on constitutional law mentions the Declaration only twice over the course of 1381 pages.16 One might expect that Bruce Ackerman’s comprehensive study of American constitutional law

12. DONALD, supra note 4, at 201 (quoting 2 COLLECTED WORKS OF ABRAHAM LINCOLN 404 (Roy P. Basler ed., 1990)).

13. See Dred Scott, 60 U.S. (19 How.) at 574–75 (Curtis, J., dissenting) (arguing that although Taney’s interpretation of the Declaration was probably wrong, the issue was beside the point because the real issue was the status of free blacks at the time of the ratification of the Constitution).

14. See id. at 529 (McLean, J., dissenting).


development would refer to the document that affected the greatest constitutional transformation of all—American independence from Britain. But Ackerman cites the Declaration only five times over the course of two volumes, and never for any substantive point.\textsuperscript{17} For most legal academics, the Declaration is little more than a political puff piece,\textsuperscript{18} or a “propaganda manifesto,” as Richard Hofstadter described it.\textsuperscript{19} It is of historical interest,\textsuperscript{20} perhaps, but of no more use to the world of law than the Pledge of Allegiance or the diary of Paul Revere.

The Declaration’s disappearance from the world of legal scholarship contrasts sharply with its continued place of prominence in our nation’s political life. One hundred years after the Gettysburg Address, Martin Luther King, Jr., stood on the steps of the Lincoln Memorial in Washington, D.C., and called on the nation to live up to its “creed” expressed in the Declaration of Independence.\textsuperscript{21} President Clinton has noted that he “reread[s]” the Declaration “on a regular basis”\textsuperscript{22} and has stated that “if you believe in the Declaration of Independence and the Constitution . . . then you are an American.”\textsuperscript{23} On the opposite end of the

\begin{itemize}
\item[18.] See Dan Himmelfarb, Note, The Constitutional Relevance of the Second Sentence of the Declaration of Independence, 100 YALE L.J. 169, 169 n.2–5 (1990) (surveying the literature); see also WALTER BERNS, TAKING THE CONSTITUTION SERIOUSLY 16 (1987)(“It has become common to hear scholars dismiss the Declaration as mere propaganda, or a clever lawyer’s brief with a dash of natural law added for spice, a convenient weapon to wield against the British, and no more.”).
\item[19.] RICHARD HOFSTADTER, THE PROGRESSIVE HISTORIANS 269 (1968).
\item[20.] Even this might be doubted. The two greatest historical studies of the American Revolution, as fresh and compelling today as when they were first published over thirty years ago, ignore the Declaration almost entirely. See BERNARD BAILYN, THE IDEOLOGICAL ORIGINS OF THE AMERICAN REVOLUTION (1967); GORDON S. WOOD, THE CREATION OF THE AMERICAN REPUBLIC, 1776–1787 (1969). Such treatment from our most eminent historians helps to confirm the legal academy’s view of the Declaration as trivial and unimportant.
\item[22.] President William J. Clinton, Remarks to Senior Citizens Council (July 28, 1998), available at 1998 WL 425314.
\item[23.] President William J. Clinton, Remarks at Portland State Univ. Commencement (June 18, 1998), available at 1998 WL 321814. Clinton’s grasp of the Declaration, however, is far from firm. In one speech, Clinton observed that “we are bound together by a written Constitution that’s 220 years old, going back to the Declaration of Independence,” thus confusing the Declaration and the Constitution. President William J. Clinton, Remarks to the U.S. Agricultural Communicators Congress (July 16, 1996), available at 1996 WL 399864. In another speech, he remarked, “The last time I checked, the Constitution said, ‘of the people, by the people and for the people.’ That’s what the Declaration of Independence says,” thus confusing the Declaration, the Constitution, and the
political spectrum, Republican presidential candidates Alan Keyes and Gary Bauer repeatedly argued that the practice of abortion violates the Declaration of Independence. Keyes declared that he stood “without apology for the principles articulated in the Declaration of Independence, and in particular for the unalienable right to life of all human offspring.” Bauer contended that he would “put no justice on the court that does not understand the clear, moral idea found in the Declaration of Independence that is the basis of this country.” In one presidential debate, all but one of the Republican candidates mentioned the Declaration of Independence as an essential item for any time capsule prepared for the twenty-first century. Representative Henry Hyde of Illinois, in closing arguments in the impeachment trial of President Clinton, appealed to “our country’s birth certificate, our charter of freedom, our Declaration of Independence.” A 1989 federal act described the Declaration as “one of the greatest documents in human history.” Few Americans, I suspect, can recall any phrases from the Constitution, but almost everyone knows that “all men are created equal” and that they have a right to “life, liberty, and the pursuit of happiness.” Indeed, in 2000 the New Jersey Senate passed a bill requiring schoolchildren to recite that portion of the Declaration of Independence each day along with the Pledge of Allegiance. And a rare

25. Excerpts from GOP Debate, ASSOCIATED PRESS ONLINE, Jan. 6, 2000, available at 2000 WL 3303828. Bauer observed that “the first right [the Declaration] lists is the right to life.” Id.
29. In the interest of our nation’s forests, I have departed from the strict requirements of the Bluebook when citing to the Declaration of Independence. Adherence to the relevant citation rule would swell the footnotes to an intolerable length, with only an imperceptible benefit to the reader. As form should always follow function, even in the arcane world of legal citation, I have not provided footnotes for every quotation from the Declaration. Instead, the Declaration is reproduced in full in the Appendix, where the interested reader can peruse it at leisure. In cases where I have added emphasis to the text, however, a traditional footnote so noting has been included. I am grateful to the editors of the Washington Law Review for accommodating this approach.
30. Ron Marsico, Senate Clears Recitation of Declaration in School, STAR-LEDGER (Newark, N.J.), June 27, 2000, at 13, available at 2000 WL 23585492. Some opponents of the measure felt it “could send a wrong message to girls.” Id. The measure should nonetheless please Gary Bauer; on
original printing of the Declaration recently sold for $8.1 million at a Sotheby’s auction. The buyer, television and movie producer Norman Lear, intends to make the printing the focus of a traveling exhibit and patriotic show.\textsuperscript{31}

\section*{C. Noscitur a Sociis: The Taint of Lochner}

Given the Declaration’s deep resonance in American public life, why do legal academics view it with such suspicion? The answer, I believe, is that the Declaration has been tainted by judicial decisions of dubious constitutional merit. The celebration of the Declaration in \textit{Dred Scott} was followed by similarly fulsome praise in the decisions of the \textit{Lochner} era. One of the earliest decisions in this mold was \textit{Butcher's Union Slaughter-house & Live-Stock Landing Co. v. Crescent City Live-Stock Landing & Slaughter-house Co.},\textsuperscript{32} in which two concurring justices invoked the Declaration to protect economic liberty. Justice Bradley argued that the “right to follow any of the common occupations of life is an inalienable right” secured by the Declaration of Independence.\textsuperscript{33} Justice Field was even more exuberant, announcing that the Declaration was a “new evangel of liberty to the people,” and explaining that the phrase “pursuit of happiness” meant “the right to pursue any lawful business or vocation, in any manner not inconsistent with the equal rights of others, which may increase their prosperity or develop their faculties, so as to give to them their highest enjoyment.”\textsuperscript{34} These sentiments soon worked their way into governing law. In the 1897 decision of \textit{Gulf, Colorado & Santa Fe Railway Co. v. Ellis},\textsuperscript{35} the Supreme Court ruled that the state of Texas could not require railroads

---


\textsuperscript{32} 111 U.S. 746 (1884).

\textsuperscript{33} \textit{Id.} at 762 (Bradley, J., concurring); \textit{see also} \textit{Slaughter-House} Cases, 83 U.S. (16 Wall.) 36, 116 (1873) (Bradley, J., dissenting) (noting that the phrase “pursuit of happiness” in the Declaration of Independence is equivalent to “property”).

\textsuperscript{34} \textit{Butcher's Union}, 111 U.S. at 756–57 (Field, J., concurring).

\textsuperscript{35} 165 U.S. 150 (1897).
to pay attorneys’ fees in tort actions. The Court, in an opinion by Justice Brewer, noted that the Declaration is the “thought and the spirit” of the Constitution and observed that “it is always safe to read the letter of the Constitution in the spirit of the Declaration of Independence.”\(^{36}\) In this case, the Court read the Declaration’s statement that “all men are created equal” to mean that such tort legislation violated the “equality of rights which is the foundation of free government.”\(^{37}\) Four years later, the Court relied on this language to strike down a statute governing the operations of large stockyards.\(^{38}\) Lower courts, too, wielded the Declaration in defense of laissez-faire constitutionalism. A district court in South Carolina proclaimed in 1907 that “the right of every citizen to work where he will, . . . to select not only his employer, but his associates . . . is one of those inalienable rights formulated in the Declaration of Independence.”\(^{39}\) The Ninth Circuit ruled in 1937 that the National Labor Relations Board could not use its powers to protect unionized employees, because the term “liberty” as used “in the Fifth Amendment to the Constitution and in the Declaration of Independence included the right to freely contract.”\(^{40}\) As late as 1943, the Supreme Court of North Dakota stated that a statute requiring the licensing of professional photographers was inconsistent with an individual’s right to the “pursuit of happiness” under the Declaration of Independence.\(^{41}\)

This tight intertwining of the Declaration with the most extreme forms of *Lochner*-ism suggests why academics fear giving significant weight to the Declaration. The Declaration’s sonorous phrases seem to provide little guidance in determining the scope of the liberty of which it speaks, and the document can be cited for the most dubious of causes. Indeed, many of the Declaration’s most fervent supporters are precisely those people with whom the legal academy is least likely to sympathize. For years, tax protestors have unsuccessfully argued that the federal

---

36. *Id.* at 160.
37. *Id.* In an 1891 address at Yale Law School, Justice Brewer (who was Justice Field’s nephew) contended, “When, among the affirmatives of the Declaration of Independence, it is asserted that the pursuit of happiness is one of the unalienable rights,” it means that government cannot interfere with the “acquisition, possession and enjoyment of property.” Justice David J. Brewer, Address at Yale Law School (1891), quoted in Owen M. Fiss, *David J. Brewer: The Judge as Missionary, in Philip J. Bergan et al., The Fields and the Law 53, 59* (Federal Bar Council 1986).
40. *NLRB v. Mackay Radio & Tel. Co.*, 87 F.2d 611, 615 (9th Cir. 1937).
income tax laws violate the Declaration of Independence. In 1972, a New York Court of Appeals judge contended that abortion was illegal under the Declaration of Independence, because the Declaration has the “force of law” and abortion violates the “natural law” that the Declaration invokes. Pro se litigants turn to the Declaration with disturbing regularity, invoking it as a palliative for almost every conceivable injury. Lacking any determinate or precise meaning, the

42. See, e.g., United States v. Collins, 920 F.2d 619, 623 (10th Cir. 1990) (noting that defendant in tax case filed an 84-page motion to dismiss, which was “lavishly larded with citations to the Declaration of Independence” and which argued that federal criminal jurisdiction was unconstitutional outside of the District of Columbia); Burroughs v. Wallingford, 780 F.2d 502, 503 (5th Cir. 1986) (rejecting argument that tax liens violated “unalienable” rights secured by the Declaration of Independence); Newman v. Schiff, 778 F.2d 460, 461 (8th Cir. 1985) (noting that tax protestor had written book entitled “The Tax Rebel’s Guide to the Constitution and the Declaration of Independence”); United States v. Farber, 679 F.2d 733, 735 (8th Cir. 1982) (noting that tax protestor included copies of the Declaration of Independence and the Mayflower Compact in his return); United States v. Schmitz, 542 F.2d 782, 783, 785-786 (8th Cir. 1976) (rejecting argument that 1040 forms “interfere with and obstruct . . . duties and rights under the Declaration of Independence”); Koll v. Wayzata State Bank, 397 F.2d 124, 125 (8th Cir. 1968) (dismissing claim that bank had engaged in conspiracy to deprive claimant of “‘rights, privileges and immunities’ secured by the Declaration of Independence” and noting that the complaint consisted of “16 printed pages of disconnected, incoherent, and rambling statements” challenging the constitutionality of the income tax); Mckenney v. Blumenthal, 1979 WL 1342, at *1 (N.D. Ga. Feb. 23, 1979) (rejecting claim that IRS procedures violated the Magna Carta and the Declaration of Independence); Steinbrecher v. Comm’r of Internal Revenue, 1977 WL 1287, at *1 (W.D. Tex Aug. 15, 1977) (rejecting challenge to tax laws that rested jurisdiction on, inter alia, the Declaration of Independence and the Northwest Ordinance of 1787).


Declaration of Independence

Declaration becomes little more than a Rorschach test, into which anything and everything might be read. The meaning of the Declaration is therefore best left to the elected branches of government, to the Lincolns of the world, rather than to the Taneys. The understanding of “liberty” in the Due Process Clause of the Fourteenth Amendment has created enough trouble; why encumber judges with the additional duty of figuring out, for example, what constitutes “the pursuit of happiness?”

The few academics who do emphasize the Declaration do nothing to dispel these concerns. These scholars focus almost exclusively on the second sentence of the Declaration, and they conclude that the Declaration is primarily about natural law and the protection of natural rights. Accordingly, the Declaration should be read in light of its natural law origins. “[T]he natural-rights principles embodied in the Declaration,” one scholar tells us, “are at the heart of the Constitution.” Not surprisingly, arguments of this sort have been a resounding failure in the legal academy. Invoking “natural rights” in a modern law school is about as persuasive as citing Cotton Mather’s treatise on witchcraft. Legal Realism has effectively banished “natural rights” to a distant cage in Felix Cohen’s “menagerie of metaphysical monsters.” Natural law has mystical overtones that seem out of place in a modern society, and cases that have invoked or implied natural law arguments are widely regarded as embarrassments.

that the Declaration of Independence reads in part: ‘He (the King of Great Britain) has erected a multitude of New Offices, and sent hither swarms of Officers to harass our people and [e]at out their substance”).

45. See, e.g., Derden v. McNeel, 978 F.2d 1453, 1456 n.4 (5th Cir. 1992) (noting that it is unhelpful to cite the Declaration of Independence since “general statements about inalienable rights . . . tell us little about the prerogatives of an individual in concrete factual situations”).


47. GERBER, supra note 46, at 3. Philip Hamburger argues that “[i]n the eighteenth century, however, American ideas of natural rights and natural law” were “relatively precisely defined” and “circumscribed by their very character as natural rights.” Philip A. Hamburger, Natural Rights, Natural Law, and American Constitutions, 102 YALE L.J. 907, 908 (1993).


50. See, e.g., Bradwell v. Illinois, 83 U.S. (16 Wall.) 130, 141 (1873) (“[N]ature herself . . . has always recognized a wide difference in the respective spheres and destinies of man and woman.”)
Even scholars who attempt to minimize the Declaration’s purported roots in natural law do little to alleviate academics’ inherent suspicion of the document. In a recent book, Charles Black, Jr., argues that the Declaration protects what he terms “human rights.”51 Black believes that a complete theory of human rights law can be based solely on the Declaration, the Ninth Amendment, and the Privileges or Immunities Clause of the Fourteenth Amendment.52 Although Black assiduously avoids the term “natural rights,” his jurisprudence of “human rights” is largely indistinguishable from a jurisprudence of “natural rights,” and it is doubtful that his argument does much to resolve any concrete dispute over the scope of these “human rights.” Mark Tushnet also repeatedly invokes the Declaration (appropriately rephrasing certain terms to reflect modern sensibilities). In a recent book, Tushnet argues that the Declaration and the Preamble to the Constitution form what he terms the “thin Constitution.”53 From this, he draws the startling conclusion that a proper adherence to the principles of the Declaration and the Preamble requires the abolition of judicial review.54

D. The Law of the Declaration

The almost exclusive emphasis that has been placed on the Declaration’s famous second paragraph55 has assured that the Declaration will always be seen as, well, a bit fluffy—fine for Fourth of July orations, but useless for any serious analysis of legal issues. Certain phrases of the Declaration are so familiar that it is easy to think that they are synonymous with the whole. Thus, an imagined familiarity with the

This is the law of the Creator.”) (Bradley, J., concurring); Sodero v. Sodero, 56 N.Y.S.2d 823, 827 (N.Y. Sup. 1945) (stating that natural law is “codified in the Ten Commandments”).


52. See id. at ix.


54. See id. at 129–76. A more sensible view of the Declaration is found in J.M. Balkin, *The Declaration and the Promise of a Democratic Culture*, 4-SPG WIDENER L. SYMP. J. 167 (1999). Balkin sees the Declaration as consisting of “promises,” which are the “soul” of the Constitution. Id. at 169. One such promise is the promise of a democratic culture, made possible by the American Revolution’s attack on monarchy. Id. at 170. However, the document seems more concerned with “tyranny” than with monarchy, and it never suggests that the two are synonymous. See infra Part III.C; see also Himmelfarb, supra note 18, at 175–76.

55. See, e.g., BLACK, supra note 51, at 5 (noting that he will consider only the Declaration’s opening paragraphs); Himmelfarb, supra note 18 (arguing the constitutional significance of the Declaration solely on the basis of first sentence of the second paragraph).
Declaration of Independence

Declaration has bred a very real contempt for its supposed vagueness and malleability. Yet President Lincoln and Chief Justice Taney were both convinced that a proper understanding of the Declaration was desperately important to the world of law. Why might they have done so, and why might we still find the Declaration relevant to our role as lawyers? Consider the following ways in which the Declaration and law are intricately intertwined:

**Justification for State Constitutions.** By justifying the independence of the American colonies, the Declaration legitimated the drafting of new state constitutions. Many states accordingly cited the Declaration in the preambles to their constitutions, and New York included the complete text of the Declaration in its constitution of 1777.

**Severance of English Common Law.** Most American jurisdictions recognize that English statutes and common law decisions prior to July 4, 1776, are part of American common law. Courts have repeatedly held that the Declaration severed English common law from its American counterpart. English law after the Declaration has no force in America and must be pled as foreign law.

56. See, e.g., Saikrishna B. Prakash, America’s Aristocracy, 109 YALE L.J. 541, 553–54 (1999) (reviewing Tushnet, supra note 53) (“The Declaration has a little more substance [than the Preamble to the Constitution], but that is not saying much. Though it gets the patriotic juices flowing . . . [its] principles tell us nothing concrete.”).

57. See, e.g., Md. Const. of 1776, preamble; N.C. Const. of 1776, preamble; Pa. Const. of 1776, preamble; S.C. Const. of 1777, preamble; Vt. Const. of 1777, preamble; see also Trs. of Phillips Exeter Acad. v. Exeter, 27 A.2d 569, 584 (N.H. 1940) (stating that “full statehood dated from the Declaration of Independence”).

58. N.Y. Const. of 1777, preamble.

59. See, e.g., Hilton v. Guyot, 159 U.S. 113, 180 (1895); United Copper Sec. Co. v. Amalgamated Copper Co., 232 F. 574, 577 (2d Cir. 1916); Greear v. Paust, 279 N.W. 568, 570 (Minn. 1938); Gwathmey v. North Carolina, 464 S.E.2d 674, 679 (N.C. 1995) (stating that the “common law” to be applied in North Carolina is the common law of England to the extent it was in force and use within the State at the time of the Declaration of Independence); Richards v. Redelheimer, 36 Wash. 325, 328, 78 P. 934, 935 (1904); see also R.I. GEN. LAWS § 43-3-1 (2000); Lowe v. Kansas, 163 U.S. 81, 85 (1896) (“[What is due process of law] depends upon the question whether it was in substantial accord with the law and usage of England before the Declaration of Independence, and in this country since it became a nation.”); Manoukian v. Tomasion, 237 F.2d 211, 215 (D.C. Cir. 1956) (“British statutes antedating the Declaration of Independence have almost universally been regarded as having the effect of judicial precedent.”). But see Penny v. Little, 4 Ill. (3 Scam.) 301, 302–03 (1841) (noting that the Illinois legislature dated American common law to the fourth year of the reign of James I when the first territorial government was established in America).

**Title to Land.** The Declaration of Independence transferred title to public lands from the Crown to the states, and fixed the boundaries of the original thirteen colonies. The Declaration also transferred western lands to the thirteen states. An 1827 Supreme Court decision held that title to the trans-Appalachian lands did not pass by cession from Great Britain in the 1783 Treaty of Paris; rather, these lands transferred by right to the United States with the signing of the Declaration of Independence.

**Creation of American Citizenship.** American law dates American citizenship to the Declaration of Independence. By contrast, English law holds that Americans remained British subjects until the Treaty of Paris in 1783. A person in the United States on July 4, 1776, is

---

61. *See, e.g.*, Phillips v. Delaware, 330 A.2d 136 (Del. 1974) (holding that title to lands held by William Penn terminated with the Declaration and transferred by sovereign succession to the state of Delaware); Town of Oyster Bay v. Commander Oil Corp., 677 N.Y.S.2d 746, 748 (N.Y. Sup. Ct. 1998) (“by the Declaration of Independence title to the lands under tidewaters which had been vested in the King passed to the states in which they were situated”).

62. *See New Jersey v. Delaware, 291 U.S. 361, 370 (1934) (“The Declaration of Independence had made Delaware a state with boundaries fixed as of that time.”); Howard v. Ingersoll, 54 U.S. (13 How.) 381, 398 (1852) (“It is well known to all of us, when the colonies dissolved their connection with the mother country by the Declaration of Independence, that it was understood by all of them, that each did so, with the limits which belonged to it as a colony.”); Rhode Island v. Massachusetts, 37 U.S. (12 Pet.) 657, 736–37 (1838).

63. *See Harcourt v. Gaillard, 25 U.S. (12 Wheat.) 523, 527 (1827); see also Downes v. Bidwell, 182 U.S. 244, 302 (1901) (noting that the Declaration gave the United States power to acquire territory). Whether the western lands passed directly to individual states or to the Congress was a divisive issue, which was finally resolved with Virginia’s cession to the United States of its remaining western claims. *See Peter Onuf, The Origins of the Federal Republic* 17 (1983). That the Declaration passed title to crown lands to the United States was well-recognized by contemporaries. *See, e.g.*, *To the Printers of the Pennsylvania Gazette, PA. GAZETTE*, Aug. 15, 1781 (“By virtue of our Declaration of Independence, we stand possessed of all that property of the King of Britain in America, which he held by purchases from the Indians. . . .”).


65. *See Inglis, 28 U.S. (3 Pet.) at 121. A Pennsylvania law of Dec. 5, 1778, required officeholders to swear that they had never “since the Declaration of Independence, directly or indirectly aided, assisted, abetted, or in any wise countenanced the King of Great Britain,” and that they had
presumed to be an American citizen, unless he actively took steps to retain his allegiance to Britain.\textsuperscript{66} Although these distinctions are of little consequence today, they figured in numerous eighteenth- and nineteenth-century inheritance disputes, because only citizens could inherit real property in America.\textsuperscript{67} The Declaration is also relevant to naturalization proceedings, as some courts have held that an understanding of the Declaration is an essential prerequisite to citizenship.\textsuperscript{68}

\textit{Treason.} An important corollary to the law of citizenship is the law of treason. The Declaration made official what colonial Americans had been suggesting for some time—that adherence to the king of Great Britain was treason.\textsuperscript{69} The Declaration states that “we hold [our British brethren], as we hold the rest of Mankind, Enemies in War, in Peace, Friends.” “Enemies,” of course, has a technical meaning within the law of treason. The British were now officially the enemies, and the states would quickly respond with a host of treason statutes designed to punish those who aided the British cause.\textsuperscript{70}

\textit{Requirement for Admission to the Union.} Since the latter half of the nineteenth century, Congress has required that territories seeking statehood create constitutions that are “not repugnant to... the principles of the Declaration of Independence.” Such provisions were part of the Enabling Acts for the territories that became Hawaii, Alaska, New Mexico, Arizona, Oklahoma, Utah, North Dakota, South Dakota,

\begin{quote}
“ever since the Declaration of Independence... demeaned [themselves] as faithful citizen[s] and subject[s] of this, or some one of the United States.” PA. GAZETTE, May 17, 1780.
\textsuperscript{66} See Inglis, 28 U.S. (3 Pet.) at 123–24; see also Kilham v. Ward, 2 Mass. 236 (1806) (discussing the relation between the Declaration and citizenship).
\textsuperscript{67} See, e.g., Munro v. Merchant, 28 N.Y. 9 (1863); see also Lamoreaux v. Ellis, 50 N.W. 812, 815–16 (Mich. 1891) (evaluating citizenship of an individual’s grandfather at time of Declaration to determine individual’s eligibility for the office of sheriff); State ex rel. Phelps v. Jackson, 65 A. 657, 658 (Vt. 1907) (evaluating citizenship of an individual’s great-grandfather at time of Declaration to determine individual’s eligibility for the office of state’s attorney).
\textsuperscript{68} See, e.g., In re Goldberg, 269 F. 392, 397 (E.D. Mo. 1920) (holding that “an acquaintance with, and working knowledge of, the principles of the Declaration of Independence” is “an indispensable prerequisite to naturalization”).
\textsuperscript{69} See, e.g., Shanks v. Dupont, 28 U.S. (3 Pet.) 242, 255 (1830) (Johnson, J., dissenting) (“The courts of this country all consider this transfer of allegiance as resulting from the declaration of independence.”).
\textsuperscript{70} The Declaration confirmed, rather than created, the idea that treason could be committed against America. Although the treason statutes enacted subsequent to the Declaration were enacted at the state level, they are nationalist in flavor and draw upon the experience of colonists prior to the Declaration in punishing adherents to Britain as “traitors to their country.” Carlton F.W. Larson, Constructing Treason in Revolutionary Pennsylvania 107 (1997) (unpublished A.B. thesis, Harvard University) (on file with the Harvard University Archives).
Montana, Washington, Colorado, and Nevada. In an early decision, the Oklahoma Supreme Court held that the Enabling Act has no force subsequent to statehood and was not intended to be “an irrevocable limitation on the sovereign powers of the state,” but other states have rejected this narrow reading. The Supreme Court of Arizona has held that the Enabling Act is “the fundamental and paramount law” of Arizona; nothing in the Arizona Constitution can be “inconsistent with the Enabling Act.” In other words, a constitutional amendment in Arizona is void if it conflicts with the “principles of the Declaration of Independence.” In similar fashion, the Supreme Court of Colorado has explicitly ruled that various amendments to the Colorado constitution do not violate the principles of the Declaration, which suggests that the Declaration is also binding in Colorado.


72. Atwater v. Hassett, 111 P. 802, 813 (Okla. 1910); cf. Frantz v. Autry, 91 P. 193, 203–04 (Okla. 1907) (holding that whether proposed state constitution conformed with principles of Declaration was a political question for the President of the United States to determine). For the contrary view, see McCabe v. Atchison, Topeka & Santa Fe Ry., 186 F. 966, 984 (8th Cir. 1911) (Sanborn, J., dissenting) (citing the Declaration requirement in the Oklahoma Enabling Act and arguing that a state cannot violate with impunity the terms of its admission to the Union).


Justice Spencer of the Nebraska Supreme Court similarly argued that the Declaration is binding under the Nebraska Enabling Act. See DeBacker v. Sigler, 175 N.W.2d 912, 914 (Neb. 1970) (Spencer, J., dissenting); Nebraska ex rel. Belker v. Bd. of Educ. Lands & Funds, 175 N.W.2d 63, 69 (Neb. 1970) (Spencer, J., dissenting); State ex rel. Morris v. Marsh, 162 N.W.2d 262, 277 (Neb. 1968) (Spencer, J., dissenting). In none of these cases did his colleagues address this issue.

Any force the Declaration may have in these states is most likely a matter of state constitutional law. In Coyle v. Smith, 221 U.S. 559 (1911), the Supreme Court held that an Enabling Act provision
Creation of a “Negative Constitution.” The bulk of the Declaration of Independence consists of specific allegations against the King of Great Britain. In the *Merryman* case, Chief Justice Taney cited one of these allegations (that the King had attempted to “render the military independent of, and superior to, the civil power”) to support his conclusion that neither President Lincoln nor the military could suspend the writ of habeas corpus. Taney assumed that the Constitution would not permit those acts that the Declaration had specifically cited as justifying revolution. In other words, the Declaration was a sort of “negative constitution,” a source of structural limitations on the power of government. The charges against the King can be easily converted to positive restrictions on the government: “Thou shalt not render the military independent of, and superior to the civil power,” for example. These commands are illustrative of the understanding of the proper role of government that animated the Constitution and can therefore have exceptionally persuasive weight, even if they technically lack binding authority.

This mode of reasoning has been a powerful force in American law since at least 1799. The charges against the King have been regularly invoked by American judges to limit the powers of government. Judges have argued that the charge, “He has refused his Assent to Laws, the most wholesome and necessary for the public Good,” limits executive
veto power; that the charge, "He has affected to render the Military independent of and superior to the Civil Power," limits the power of the military; that the charge, "For depriving us, in many Cases, of the Benefits of Trial by Jury," limits the government’s ability to restrict jury trials; and that the charge, "For transporting us beyond Seas to be tried for pretended Offences," limits the government’s power to change venue in criminal prosecutions. Our judiciary has never tired of reminding us that one of the causes of the American Revolution was that George III had "made Judges dependent on his Will alone, for the Tenure of their Offices, and the Amount and Payment of their Salaries." Judicial opinions have been illuminated by even the most obscure charges of the Declaration: "He has called together Legislative Bodies at Places unusual... and distant"; “He has endeavoured to prevent the Population of these States”; “He has... sent hither Swarms of Officers..."


82. See, e.g., Weiss v. United States, 510 U.S. 163, 198–99 (1994) (Scalia, J., concurring in part and concurring in the judgment); United States v. Will, 449 U.S. 200, 219 (1980); O’Donoghue v. United States, 289 U.S. 516, 531 (1933); Harline v. DEA, 148 F.3d 1199, 1203 (10th Cir. 1998); In re Clay, 35 F.3d 190, 191 (5th Cir. 1994); Oates v. Rogers, 144 S.W.2d 457, 459 (Ark. 1940); Lee v. Bd. of Pension Trs., 739 A. 2d 336, 341 (Del. 1999); Gordy v. Dennis, 5 A.2d 69, 72 (Md. 1939); Grimbald v. Beattie, 177 S.E. 667, 676–77 (S.C. 1934) (describing the charge "[a]s almost a controlling influence on the decisions reached herein").

to harass our People, and eat out their Substance"; 85 "For quartering large Bodies of Armed Troops among us"; 86 "For taking away our Charters"; 87 "For... abolishing our most valuable laws"; 88 and "[He has incited] the merciless Indian Savages, whose known Rule of Warfare, is an undistinguished Destruction, of all Ages, Sexes and Conditions." 89 This use of the Declaration to limit the powers of government was aptly summarized by a federal appellate court in 1899: "The Declaration of Independence is not, as has sometimes been flippantly asserted, a mere string of glittering generalities. It is a bill of rights which enters fundamentally into the structure of our Government." 90

Sovereignty. The Declaration unambiguously ended British sovereignty over the American colonies. What the Declaration then did with that sovereignty is much less clear. Did it create thirteen independent, sovereign nations, or did it vest sovereignty in the Continental Congress, or the people of the states, or in some combination of both? The question is significant, because major issues in American federalism hinge on the status of the states at the time of the Constitution’s ratification. 91 The case law in this area, however, is

89. See Coleman v. United States, 715 F.2d 1156, 1157 (7th Cir. 1983).

An Ohio Supreme Court Justice, convinced that the Declaration must somehow be relevant but not sure precisely why, simply included the entire Declaration of Independence in a dissenting opinion in a recent train accident case. See Gladon v. Greater Cleveland Reg’l Transit Auth., 662 N.E.2d 287, 308–10 (Ohio 1996) (Douglas, J., dissenting). Similarly, Justice Black once suggested that the Supreme Court’s jurisprudence under section 5 of the Fourteenth Amendment violated at least four of the Declaration’s charges against the king. Perkins v. Matthews, 400 U.S. 379, 407 n.7 (1971) (Black, J., dissenting).

91. For example, the Supreme Court’s recent sovereign immunity decisions seem implicitly to assume that the states were once thirteen independent nations. See Alden v. Maine, 527 U.S. 706, 712 (1999) (holding that Congress may not subject a state to suit in its own courts without the state’s consent); Coll. Sav. Bank v. Fla. Prepaid Postsecondary Educ. Expense Bd., 527 U.S. 666, 687 (1999) (holding that states are immune from suit under the Trademark Remedy Clarification Act); Fla. Prepaid Postsecondary Educ. Expense Bd. v. Coll. Sav. Bank, 527 U.S. 627, 630 (1999) (holding that Congress may not abrogate state sovereign immunity under the Commerce and Patent Clauses and establishing strict requirements for abrogation under section 5 of the Fourteenth Amendment); Seminole Tribe v. Florida, 517 U.S. 44, 47 (1996) (holding that Congress may not abrogate state sovereign immunity under the Indian Commerce clause).
incoherent and contradictory, and the Supreme Court has done nothing to clarify it.

The Declaration states that “these United Colonies are, and of Right ought to be, Free and Independent States . . . and that as Free and Independent States, they have full power . . . to do all . . . Acts and Things which Independent States may of right do.” A number of courts have reasoned that the Declaration thereby created thirteen sovereign nations.\(^92\) Chief Justice Taney adhered to this view, arguing that “by the Declaration of Independence, [the colonies had] become separate and independent sovereignties, against which treason might be committed.”\(^93\) However, in a 7-1 decision in 1936, the Supreme Court articulated a different conclusion, stating that as a result of the Declaration of Independence, “the powers of external sovereignty passed from the Crown not to the colonies severally, but to the colonies in their collective and corporate capacity as the United States of America.”\(^94\) The Second Circuit echoed this view in a recent case, noting that beginning with the Declaration of Independence . . . the member states had no more power to make war or enter into treaties of peace or alliance than they had had as colonies under the British Crown. The Articles of Confederation appear merely to have


\(^{93}\) Kentucky v. Dennison, 65 U.S. (24 How.) 66, 101 (1861); see also GARRY WILL, INVENTING AMERICA: JEFFERSON’S DECLARATION OF INDEPENDENCE 332 (1978). (“Not one country, but thirteen separate ones, came into existence when the Declaration was at last made unanimous on July 19, 1776.”).

\(^{94}\) United States v. Curtiss-Wright Export Corp., 299 U.S. 304, 316 (1936); see also State ex rel. Mills v. Dixon, 213 P. 227, 230 (Mont. 1923) (the states were “never in their individual capacity strictly” sovereign); Maynard v. Newman, 1 Nev. 271, 273 (1865) (noting that the Declaration of Independence was the act of “one nation, or one people”). The Curtiss-Wright Court’s holding that certain foreign affairs powers are inherent in the federal government is criticized in Charles A. Lofgren, United States v. Curtiss-Wright Export Corporation: An Historical Reassessment, 83 YALE L.J. 1 (1973).
confirmed, rather than to have originated, Congress’ peace-treating power.95

* * *

In all of these ways the Declaration is relevant to the world of law. Although the Declaration is reproduced in the United States Code as part of the “Organic Laws” of the United States, and although courts have occasionally held or suggested that the Declaration is “law,”96 the Declaration is not really “law” in the way that term is traditionally understood.97 Rather, the best understanding of the Declaration is that it is a legal document that has continuing significance in American law. As the examples above have demonstrated, this significance is pervasive and cuts to the core of some of our nation’s most difficult legal issues. It is therefore vitally important that we have a proper understanding of what the Declaration of Independence actually accomplished, and of what theories of government underlie it. It is to the first of those issues that I now turn.

II. THE DECLARATION OF INDEPENDENCE AND THE CREATION OF AN AMERICAN NATION

This Part argues that the Declaration of Independence was the act of one American people, creating an American nation. After the Declaration, America presented itself as one nation to the world. Well

95. Oneida Indian Nation v. New York, 691 F.2d 1070, 1088 (2d Cir. 1982).
96. See, e.g., Kowall v. United States, 53 F.R.D. 211, 214 n.2 (W.D. Mich. 1971) (stating that the Declaration “has never been repealed, and, is valid and binding on all governments of the United States, federal and state, as a primary obligation” and citing for support Lincoln’s Gettysburg Address and Second Inaugural as carved on the Lincoln Memorial); Perez v. Lippold, 198 P.2d 17, 30 (Cal. 1948) (Carter, J., concurring) (“The Declaration of Independence is a part of the law of our land.”); Fid. & Cas. Co. of N.Y. v. Union Sav. Bank Co., 163 N.E. 221, 222 (Ohio Ct. App. 1928) (“[The] declaration is a part of the law of this state, as much so as its Constitution and statutes.”).

Courts have often used legal metaphors to describe the Declaration. It has been termed a “charter of our liberties,” Cooper v. Hindley, 70 Wash. 331, 336, 126 P. 916, 919 (Wash. 1912); a “national Magna Charta,” State v. Cutshall, 15 S.E. 261, 263 (N.C. 1892); a statement of the “American creed,” Bell v. Maryland, 378 U.S. 226, 286 (1964) (Goldberg, J., concurring); and as “the Magna Charta of our republican institutions,” United States v. Morris, 125 F. 322, 325 (E.D. Ark. 1903).

97. Cf. Troxel v. Granville, 530 U.S. 57, 91 (2000) (Scalia, dissenting) (“The Declaration of Independence . . . is not a legal prescription conferring power upon the courts.”); Morehouse v. United States Dep’t of Justice, 1998 WL 320268, at *3 (N.D. Tex. June 8, 1998) (“The Declaration of Independence does not directly create rights that can be enforced through the judicial system.”); City of Anniston v. Court of County Comm’rs, 48 So. 605, 605 (Ala. 1909) (“The Declaration of Independence is in the printed volumes called the Code, and has been in the printed copies of every previous Code; but, of course, it is not a part of the Code proper.”).
before the Articles of Confederation were even drafted, much less ratified, the Continental Congress operated as a national government that possessed all the powers of external sovereignty. What the Declaration most certainly did not do was create thirteen separate nations, completely independent of each other.

My argument, however, is not that the Declaration created a unitary nation-state. If the definition of a nation is that of an entity with exhaustive power over all subjects, and in which a majority can bind a minority, we still do not live in an American nation. A minority of the American people govern in the Senate, and, as we are too well aware, a minority of the people can elect a President through the electoral college. Likewise, there are a variety of subjects over which the federal government has no constitutional authority to act.98 As James Madison recognized, the Constitution of 1789 was “neither a national nor a federal constitution; but a composition of both.”99 The nation that was formed on July 4, 1776, may best be described as some sort of “confederate republic.”100 It held all the powers of external sovereignty, although the details of its internal organization were less clear.


100. In the debates surrounding the Constitution, both Federalists and Anti-Federalists used Montesquieu’s term “confederate republic” to describe both the proposed government under the Constitution and the old government under the Articles of Confederation. Douglas G. Smith, An Analysis of Two Federal Structures: The Articles of Confederation and the Constitution, 34 SAN DIEGO L. REV. 249, 260 (1997).

In his speech to the Pennsylvania ratifying convention, James Wilson stated:

The United States may adopt any one of four different systems. They may become consolidated into one government, in which the separate existence of the states shall be entirely absorbed. They may reject any plan of union or association, and act as separate and unconnected states. They may form two or more confederacies. They may unite in one federal republic.

James Wilson, Speech Delivered on 26th November, 1787, at the Convention of Pennsylvania, in 2 THE WORKS OF JAMES WILSON 759, 766 (Robert Green McCloskey ed., 1967). This suggests that Wilson viewed both the Articles and the proposed Constitution as types of “federal republics,” and not as fundamentally different entities.

Alexander Hamilton, relying on Montesquieu, described a confederate republic as an “assemblage of societies,” or an association of two or more states into one state.” THE FEDERALIST No. 9, at 122 (Alexander Hamilton) (Isaac Kramnick ed., 1987). According to Hamilton, “the extent, modifications and objects of the federal authority are mere matters of discretion.” Id. Thus, for Hamilton, there was no fundamental difference in the nature of the union between the government under the Articles of Confederation and the proposed Constitution.
Nor do I argue that the states necessarily lacked the legal authority to leave the American union after the Declaration. Article VII of the Constitution, for example, certainly contemplates that individual states might go their own way, and several of *The Federalist Papers* address the fear that the union might split into three or four separate confederacies. The important point is not so much when each state lost its power to leave the Union, but when the Union legally acquired all the great incidents of national sovereignty. These powers of sovereignty did not arise from the Articles of Confederation, ratified in 1781. Nor did they arise from the Constitution of 1789. Although the Constitution certainly created a “more perfect” American nation from an internal perspective, it was almost entirely irrelevant from an external perspective. By the time the Constitution was ratified, the Continental Congress had a lengthy history of conducting wars, negotiating treaties, governing vast federal territories, and adjudicating interstate disputes. The states did none of these things, and made no attempt to do so. If a state had wished to exercise any of the powers of external sovereignty, it would have had to withdraw affirmatively from the American union and declare itself a completely independent state.

In short, the Declaration was an act of all the American people, creating an entity, the United States of America, which presented itself as one nation to the world. This Part is dedicated to an exploration of this theme. Section A offers a brief discussion of the interpretive method employed in this Part and in Part Three. Sections B and C focus closely on the Declaration’s text. Section B considers the problem of the voicing of the Declaration and argues that the American people as a whole declared American independence. Section C addresses what sort of entity or entities the Declaration created and argues that the Declaration...
created one American nation, at least with respect to the rest of the world. Section D argues that ample historical evidence supports the textual reading offered in Sections B and C.

A. A Preliminary Note on Method

The argument in the next two Parts will require a close attention to the text of the Declaration. Despite the recent publication of several important works on the Declaration, it remains true, as Gary Wills pointed out twenty years ago, that the Declaration has not been subject to the rigorous “construction” of legal analysis. It has been too easy to move quickly from the text of the Declaration to vague generalities about the nature of liberty or into the subtleties of eighteenth-century intellectual life. Yet the delegates to the Continental Congress knew they were composing a document that would justify their actions to the ages, and they accordingly spent an extraordinary amount of time editing Thomas Jefferson’s draft of the Declaration. They excised sections, added new ones, and made minute adjustments to language. As Pauline Maier has put it, “By exercising their intelligence, political good sense, and a discerning sense of language, the delegates managed to make the Declaration at once more accurate and more consonant with the convictions of their constituents, and to enhance both its power and


Garry Wills’s controversial Inventing America, supra note 93, relates Jefferson’s draft of the Declaration to the trans-Atlantic intellectual world of the Enlightenment. Although many of Wills’s observations continue to be useful, his conclusion that Jefferson’s primary intellectual debt was to the Scottish Enlightenment has not withstood scholarly criticism. See, e.g., Ronald Hamowy, Jefferson and the Scottish Enlightenment: A Critique of Garry Wills’s Inventing America: Jefferson’s Declaration of Independence, 36 WM. & MARY Q. 503 (1979).

105. WILLS, supra note 93, at xxv.

106. See MAIER, supra note 3, at 143–50.
its eloquence.” Thus, it is important that we pay careful attention to the precise language the Congress employed—and not just in the famous second paragraph, but also in the more substantial charges against the King. As contemporaries recognized, these charges were the heart of the document, the legal basis that justified revolution.

A focus on the text of the Declaration requires, of course, that we have the correct text in front of us. There at least two documents with a strong claim to being the “official” text of the Declaration, and they differ from each other in caption, signature, capitalization, and punctuation. The first text is the printing prepared by Philadelphia printer John Dunlap at the direction of the Continental Congress on the night of July 4, 1776. It was through this printing, known as the “Dunlap broadside,” that most Americans became acquainted with the Declaration. Congress sent copies of the Dunlap broadside to all of the states, and ordered that it “be proclaimed in each of the United States, and at the head of the army.” The second text is the parchment copy now on display at the National Archives. This copy was prepared pursuant to a July 19, 1776, resolution of the Continental Congress, and, unlike the Dunlap broadside, it includes the signatures of all the delegates. However, this parchment copy was effectively a secret document, and it was not until January 18, 1777, after military success at Trenton and Princeton, that Congress sent copies to the states.

The Dunlap broadside has the better claim to our attention. It is the nearest text to the actual events of July 4, 1776. Moreover, as the widely disseminated public version of the Declaration, it is the text that most Americans came to know and celebrate. The subsequent publication history of the Declaration unfortunately tended to blend elements of the Dunlap broadside with elements of the parchment copy, with the result

---

107. Id. at 150. For this reason, it is proper to say that Thomas Jefferson drafted the Declaration, but that the Continental Congress is its author.

108. MAIER, supra note 3, at 105. In the works of Harry Jaffa and Walter Berns, two scholars for whom the Declaration is a fundamental source of constitutional principles, the charges against the King seldom, if ever, make an appearance. See BERNS, supra note 18; JAFFA, supra note 46. The same can be said of Scott Gerber’s recent work on the Declaration. See GERBER, supra note 46.


110. MAIER, supra note 3, at 130–31, 159.

111. Id. at 130.

112. Id. at 150–51.

113. Id. at 153.
that many modern printings of the Declaration confuse elements of the two texts.114 For the purposes of this Article, I will rely on the Dunlap broadside as the authoritative text and refer to the parchment only for secondary points. For the reader’s convenience, the Dunlap broadside is reproduced in the Appendix.

This Article employs many of the interpretive devices that Akhil Amar describes as “intratextualism.”115 Intratextualism is specifically concerned with how words and phrases in a document interact with and illuminate each other.116 As Amar puts it, intratextualism allows us to “squeeze more meaning” from a document than would an interpretive method that is strictly clause-bound.117 In a thoughtful article, Robert Spoo has observed that Amar’s brand of intratextualism has an important aesthetic component and is a legal counterpart to the New Criticism that flourished in literary studies in the 1940s and 50s.118 Aesthetic


The Dunlap broadside is reprinted in JULIAN P. BOYD, THE DECLARATION OF INDEPENDENCE: THE EVOLUTION OF THE TEXT 78 (1945), and is transcribed in DUMBAULD, supra note 109, at 157–61. The parchment copy is transcribed in CARL L. BECKER, THE DECLARATION OF INDEPENDENCE 185–93 (1922).


116. See id. at 791–95.

117. Id. at 826–27.

118. See Robert Spoo, “No Word is an Island”: Textualism and Aesthetics in Akhil Reed Amar’s The Bill of Rights, 33 U. RICH. L. REV. 537, 573–76 (1999). Christopher Eisgruber has suggested that this sort of interpretation can quickly lead to what he terms the “Aesthetic Fallacy,” that is, the belief that “the Constitution is like a poem, a symphony, or a great work of political philosophy,” and that “[e]ach word and every phrase must come together to form a harmonious and pleasing composition.” Christopher L. Eisgruber, The Living Hand of the Past: History and Constitutional Justice, 65 FORDHAM L. REV. 1611, 1617 (1997). A similar point is made in Adrian Vermeule & Ernest A. Young, Hercules, Herbert, and Amar: The Trouble with Intratextualism, 113 HARV. L. REV. 730, 770 (2000). Of course, taken to extremes, any form of constitutional interpretation can become ridiculous, as Amar well recognizes. See Amar, supra note 115, at 799 (“Carried to extremes, intratextualism may lead to readings that are too clever by half.”). The solution, of course, is not abandonment of an interpretive method, but rather the use of good judgment in its application.

The connection between legal and literary interpretation has a durable pedigree. For example, James Wilson, a principal architect of the Constitution, signer of the Declaration, and Justice of the Supreme Court, observed in his 1793 opinion in Chisholm v. Georgia, 2 U.S. (2 Dall.) 419, 454
sensibilities are even more useful in interpreting the Declaration of Independence, a self-consciously rhetorical document, than they are for the Constitution, and much of my interpretation will be conducted in the spirit of the New Critics. My goal, though, is not simply to offer a pleasing and elegant interpretation of the Declaration. Unlike the New Critics, an interpreter of the Declaration cannot consider the text in splendid isolation from the historical reality that surrounds it. The challenge is to provide an interpretation that makes sense historically and that conforms to the best understanding of what contemporaries took the Declaration to mean. Accordingly, I often turn to historical evidence to demonstrate the soundness of my interpretations.

B. Who Declared American Independence?

The first question to ask of any declaration is, “Who is doing the declaring?” That is, in whose voice does the document purport to speak? For the Declaration of Independence, this is a surprisingly difficult question to answer. In its voicing, the Declaration is a rhetorical mirror of the Constitution. The first paragraph of the Constitution begins in the first person (“We the People of the United States”), but then shifts abruptly to the third person (“All legislative powers herein granted”) and remains in the third person for the rest of the document. The voicing of the Declaration is exactly the reverse. It opens in the third person (“When in the Course of human Events, it becomes necessary for one People to dissolve the Political Bands which have connected them with another”), but then shifts abruptly to the first person (“We hold these truths to be self-evident”) and remains in the first person for the rest of the document. The insistent use of the first person is one of the Declaration’s most conspicuous rhetorical devices:

We hold these Truths to be self-evident . . . .

(1793), “[F]or, in an instrument well drawn, as in a poem well composed, silence is sometimes most expressive.” (The text in the U.S. Reports is corrupt at this point. For the correct text, see 5 THE DOCUMENTARY HISTORY OF THE SUPREME COURT OF THE UNITED STATES 196 (Maeva Marcus ed., 1994)).

Vermeule and Young raise the additional criticism that judges are intellectually and practically ill-equipped to carry out the intratextualist project. See Vermeule & Young, supra, at 759–77. Whatever the force of this argument as an empirical description of judicial competence, it does not detract from, and arguably justifies, the continuation of the intratextualist project in the academy.

119. U.S. CONST. pmbl.
He has kept among us . . . Standing Armies, without the consent of our Legislatures . . .

He has . . . subject[ed] us to a Jurisdiction foreign to our Constitution, and unacknowledged by our Laws . . .

For quartering large Bodies of Armed Troops among us:

For cutting off our Trade with all Parts of the World:

For imposing Taxes on us without our Consent:

For depriving us . . . of . . . Trial by Jury:

For transporting us beyond Seas to be tried for pretended Offences . . .

For taking away our Charters, abolishing our most valuable Laws and altering fundamentally the Forms of our Governments . . .

Nor have we been wanting in Attentions to our British Brethren.

We have warned them . . .

We have reminded them . . .

We have appealed to their native Justice . . .

[W]e have conjured them . . .

We must, therefore, acquiesce . . .

We . . . Declare, That these United Colonies are, and of Right ought to be, FREE AND INDEPENDENT STATES . . .

[W]e mutually pledge to each other our Lives, our Fortunes, and our sacred Honor.121

The Declaration is far from explicit, though, about precisely to whom or to what this pronoun “we” refers. What noun should we read into its place? The text admits of at least four possibilities. First, it may refer to the American states. Second, it may refer to the delegates to the Continental Congress. Third, it may refer to the people of the individual

121. (emphasis added).
Declaration of Independence

states as separate political entities. Finally, it may refer to the American people acting as a whole. The following Subsections evaluate the arguments in favor of each of these alternatives, and conclude that the best reading is that “we” refers to the American people as a whole.122

I. The States

The best argument for a “We, the states” reading of the Declaration is based on the caption to the parchment copy. This heading describes the Declaration as “The unanimous Declaration of the thirteen united States of America.”123 Certainly this seems like powerful evidence that the Declaration is purporting to speak in the voice of the states, a conclusion which finds some support in other parts of the text. For example, the Declaration accuses George III of sending “Swarms of Officers to harrass our People,” and of destroying “the Lives of our People.” The phrase “our People” most plausibly implies the voice of a state. Representatives might use this phrase, but it makes little sense for the people themselves, either in states or in the aggregate, to refer to “our people.”

The bulk of the text, however, dramatically undermines a “We, the states” reading, and reveals the sheer implausibility of the caption to the parchment copy. The Declaration repeatedly uses the third person to refer to the states. For example: “Such has been the patient sufferance of these Colonies; and such is now the Necessity which constrains them to alter their former Systems of Government”;124 “the Establishment of an

122. Another possibility is that there is no consistent subject in the Declaration, and that the meaning of the first person pronouns changes from sentence to sentence. Such a lurching quality is inconsistent with everything we know about the effort that went into drafting the Declaration, see MAIER, supra note 3, at 97–153, and is so highly unlikely that it does not merit further discussion.

123. BECKER, supra note 114, at 185.

124. This intriguing phrase might also be pointed to by a defender of a “We, the States” reading. The use of the term “Necessity” relates the phrase back to the preamble (“it becomes necessary for one people”), and thus links “these Colonies” with the act of separation. An important distinction must be made, however, between creating new state governments and effecting independence from Britain. On May 10, 1776, the Continental Congress recommended that the states form new governments, a step which marked the beginning of state constitution-making. 4 JOURNALS OF THE CONTINENTAL CONGRESS, 1774–1789, 342 (Worthington C. Ford ed., 1906). Although this had many of the practical effects of a declaration of independence, it did not formally sever the connection with Britain. It is this process of state-level constitution-making to which the Declaration refers with the phrase “such is the Necessity which constrains them [the Colonies] to alter their former Systems of Government.” The reference is to a process that is already occurring, not to the formal act of separation that the Declaration effects.
absolute Tyranny over these States”; “HE has endeavoured to prevent the
Population of these States”; “introducing the same absolute Rule into
these Colonies”; “as Free and Independent States, they have full Power
to...” Likewise, in many places a “We, the states” reading is
nonsensical: “For depriving us [the states], in many Cases, of the
Benefits of Trial by Jury”; “For transporting us [the states] beyond Seas
to be tried for pretended Offences.” Nor can states really pledge “to each
other our Lives, our Fortunes, and our sacred Honor.”

So why does the parchment copy claim that the Declaration is “The
Unanimous Declaration of the thirteen united States of America?” The
original caption, which was widely disseminated in the first printed
broadsides of the Declaration, read, “A Declaration by the
Representatives of the United States of America in General Congress
Assembled.” There is no direct evidence of why the Continental
Congress ordered this change when it directed the preparation of the
parchment copy. However, the change is most likely a result of the
peculiar circumstances surrounding the Declaration’s adoption. New
York initially abstained from voting for independence, and its delegates
were not given permission to approve the Declaration until a week after
its adoption. By this time, however, the text of the Declaration had
already been issued. The Congress almost certainly wanted to emphasize
the unanimity of the states in some fashion on the new parchment copy.
However, since the body of the Declaration’s text could not be readily
altered, the only real alternative was to tinker with the caption. One
option would have been simply to insert the word “unanimous”: “The
Unanimous Declaration by the Representatives of the United States of
America...” But that was not strictly accurate, as two delegates
abstained from voting. The only real option was to say something
about a unanimous declaration of the states, even if that meant directly
contradicting the voicing of the Declaration’s text.

Rhetorical considerations may also have played a role. The initial
phrasing, “A Declaration by the Representatives of the United States of

125. (emphasis added).
126. But see WILLS, supra note 93, at 340 (“These new states pledge to each other their honor,
that honor accruing to sovereignties as they take their ‘free and equal station’ with other nations.”).
Although this might perhaps explain how states could have “honor,” a traditionally personal virtue,
Wills does not explain why states would pledge to each other their “lives,” nor does he address the
many other places in which a “we, the states” reading is simply implausible.
127. MAIER, supra note 3, at 45.
128. Id.
America in General Congress Assembled,” is a bit clumsy, piling clause upon clause (by . . . of . . . of . . . in) to no real effect. The opening indefinite article is weak and does nothing to distinguish this declaration from any other the Congress might issue. That the representatives are in “General Congress Assembled” conveys a legalistic point about the legitimacy of the Declaration, but has little rhetorical power. The altered caption is shorter and more direct: “The unanimous Declaration of the thirteen united States of America.” This emphasizes the critical point of unanimity, a point heightened by the resonance between the words “unanimous” and “united.”

Nonetheless, the parchment caption is wildly misleading as a description of the Declaration’s text. The states are simply not the active voice in the Declaration. The delegates may have recognized this in the change from “A Declaration by” to “The Unanimous Declaration of.” “By” implies agency; “of” has a more passive connotation. The Declaration may be “of” the thirteen states, without really being “by” them.

Rhetorical considerations also explain the two uses of the phrase “our people,” which most strongly suggest the voice of the states. In both instances, a plural first person pronoun would have been awkward and inappropriate in a document that was meant to be read aloud. The formulation, “He has . . . sent hither Swarms of Officers to harrass us” would be an ungainly tongue-twister. “Harrass our People” is much better. Likewise, only someone with a tin ear for language would write, “He has plundered our Seas, ravaged our Coasts, burnt our Towns, and destroyed our Lives.” Not only do persuasive documents rarely purport to speak from beyond the grave, the rhythm of this clause demands a longer concluding phrase. Again, “destroyed the Lives of our People” is much better.

2. The Delegates to the Continental Congress

A more plausible argument can be made that the persistent “we” of the Declaration is the voice of the delegates to the Continental Congress. The Declaration concludes with the majestic phrase, “we mutually
pledge to each other our Lives, our Fortunes, and our Sacred Honor.” On the parchment copy, the delegates signed their names immediately beneath this phrase. Moreover, in the one place in which the Declaration provides a clear appositive phrase for the pronoun “we,” it identifies the pronoun with the delegates: “WE, therefore, the Representatives of the UNITED STATES OF AMERICA, in GENERAL CONGRESS, Assembled . . . .” This echoes the caption to the Dunlap broadside: “A Declaration By the Representatives of the United States of America, In General Congress Assembled.”

These arguments, although forceful, are not ultimately persuasive. The identification of “we” with the delegates is restricted by a critical modifying phrase: “WE, therefore, the Representatives of the UNITED STATES OF AMERICA, in GENERAL CONGRESS, Assembled . . . do, in the Name of, and by the Authority of the good People of these Colonies, solemnly Publish and Declare . . . .”¹³⁰ The key phrase is “in the Name of.” In this context, “in the Name of,” is not simply rhetorical reiteration of “by the Authority of.”¹³¹ Rather, it states that the Declaration is the Declaration of “the good People of these Colonies.” This clause specifically names the voice that has been speaking throughout. The Declaration was not made in the “Name of” the delegates, but of the “good People of these Colonies.”

Moreover, identifying the delegates as the “we” of the Declaration is inconsistent with both the preamble of the Declaration and the theory of popular sovereignty that the Declaration lays down in its famous second paragraph. The preamble states, “WHEN in the Course of human Events, it becomes necessary for one People to dissolve the Political Bands which have connected them with another . . . a decent Respect to the Opinions of Mankind, requires that they should declare the causes which impel them to the Separation.”¹³² This sentence is immediately followed by the first use of the pronoun “We”: “We hold these Truths . . . .” The implication could not be more clear. The voice that announces itself in the second paragraph is the voice of the one people who are required by “a decent Respect to the Opinions of Mankind” to issue a declaration

¹³⁰. (emphasis added).
¹³¹. For example, the Constitution twice employs the phrase “ordain and establish,” but it is unlikely that any meaningful distinction is intended between “ordaining” and “establishing.” U.S. CONST. pmbl. & art. III, § 1. Likewise, the Declaration refers to both “the Laws of Nature and of Nature’s God,” a rhetorical flourish that is certainly not intended to suggest the existence of two separate bodies of law. MAIER, supra note 3, at 132–33.
¹³². (emphasis added).
justifying independence. Of course, practicality will dictate that such a declaration be made through representatives, but that does not make it any less a declaration by the people.

This interpretation is bolstered by the political theory of the second paragraph, which states “that whenever any Form of Government becomes destructive of these Ends, it is the Right of the People to alter or to abolish it.” In other words, it is the people, and only the people, who have the right to overthrow a government. Independence from Britain could only be effected by the people themselves; thus, it is entirely appropriate that the document justifying the exercise of this right be in the voice of the people.

3. One People or Many?

Determining that the Declaration is in the voice of “the good People of these Colonies” only partly resolves the issue. What does “good People” mean in this context? Does it mean the people of each colony acting in their capacity as members of that colony, or does it refer to the American people en masse? I believe that the best interpretation is that the voice is that of all the American people, acting together, to announce themselves as “one people,” at least with respect to the rest of the world.

We can best approach this subject by recognizing that the term “people” had a very distinct meaning in the eighteenth-century. To the twentieth-century ear, “people” often suggests the undifferentiated mass of mankind. The Continental Congress, by contrast, would have understood the term as applying only to members of a distinct political community. This long-standing sense of the term is captured in the first definition offered in the Oxford English Dictionary: “A body of persons

133. (emphasis added).
134. Stephen Lucas identifies another reason for rejecting a “We, the delegates” reading. He notes that the Declaration’s statement, “In every stage of these Oppressions we have Petitioned for Redress in the most humble Terms,” clearly refers to events that happened before the Continental Congress convened in 1774, which suggests that “we” refers to the colonists in general. Lucas, Justifying America, supra note 104, at 109.
135. This point is a corollary to the arguments advanced in Part II.B, which describe the Declaration as creating one American nation, at least with respect to the rest of the world.
136. See 11 THE OXFORD ENGLISH DICTIONARY 505 (1989) (noting “men or women indefinitely” as sixth definition of “people”).
composing a community, tribe, race or nation.” The Declaration employs the term “people” in this sense. When the Declaration refers to other groups of individuals, it uses different terms and uses them precisely: “Men,” “mankind,” and “human,” refer to all persons throughout the earth; the term “inhabitant” refers to any person who happens to be within a particular geographical area.

The preamble states, “WHEN in the Course of human Events, it becomes necessary for one People to dissolve the Political Bands which have connected them with another . . . a decent Respect to the Opinions of Mankind requires that they should declare the causes which impel them to the Separation.” Here the Declaration clearly distinguishes “one People” from “another” people (the British) and from the larger “Mankind” to whom independence must be justified. This distinction is amplified in the second paragraph: “WE hold these Truths to be self-evident, that all Men are created equal . . . that to secure these Rights, Governments are instituted among Men . . . [and] it is the Right of the People to alter or abolish [a form of government that is destructive of these ends].” The phrase “among Men” echoes the preamble’s phrase “among the Powers of the Earth.” In this context, “among” implies a clear separation of entities, just as the Constitution allows Congress to regulate commerce “among the several States.” By instituting governments, “Men” divide themselves into “Peoples.” It is these “People,” not “Men” in general, who have the right to alter their form of government. While all men can claim the “inalienable rights” of “life, liberty, and the pursuit of happiness,” only members of a particular political community can exercise the more specific right of revolution. Exercising this right meant that the British were no longer part of the same “People” as the Americans, and accordingly the Declaration states

137. Id. at 504. It is noteworthy that the Oxford English Dictionary’s principal definition of “nation” is “[a]n extensive aggregate of persons, so closely associated with each other by common descent, language, or history, as to form a distinct race or people, usually organized as a separate political state and occupying a definite territory,” thus reinforcing the connection between a “people” and a “nation.” 10 THE OXFORD ENGLISH DICTIONARY 231 (1989) (emphasis added).

138. (emphasis added).

139. (emphasis added).

140. U.S. CONST. art I, § 8, cl. 3.

141. Similarly, the Declaration notes that, upon the suspension of colonial legislatures, the “Legislative Powers, incapable of Annihilation, have returned to the People at large for their exercise,” and that the king has refused to pass “Laws for the Accommodation of large Districts of People.” These are both unmistakable references to “people” in their distinctively political capacity.
that the British will now be treated “as the rest of Mankind, Enemies in War, in Peace, Friends.”

The Declaration charges that British troops were protected “from Punishment for any Murders which they should commit on the Inhabitants of these States,” and that the King had “endeavoured to bring on the Inhabitants of our Frontiers, the merciless Indian Savages.” The use of “Inhabitants” in the latter clause distinguishes the Indians from the “People” who are exercising the right of revolution. Similarly, the condemnation of George III for murders committed on “Inhabitants” includes a variety of persons who may not be full members of the political community: children, Indian tribes friendly to the colonial cause, and aliens temporarily resident in the United States.

The special meaning of the term “people” is well-evidenced in the Constitution of 1787, which carefully distinguishes between “people” and “persons.” The two instances of the term “people” both denote a specific political group making a specific political choice. The preamble states, “We, the People of the United States . . . do ordain and establish this Constitution for the United States of America,” and Article One requires that “[House members shall be] chosen every second Year by the People of the several States.” The Constitution employs the term “Persons” to describe members of Congress (“the Names of Persons voting for and against the Bill shall be entered on the Journal”) and to describe candidates for the presidency (“[The electors] shall vote for two Persons . . . . And they shall make a List of all the Persons voted for”). Most notoriously, the term is used in calculating representation

142. (emphasis added).
143. (emphasis added).
144. The meaning of these terms is of more than antiquarian interest. In the 2000 presidential election, voters unsuccessfully challenged the right of Texas electors to vote for both George W. Bush and Richard Cheney, arguing that Cheney was not an “inhabitant” of another state, as required under the Twelfth Amendment to the Constitution. See Jones v. Bush, 122 F. Supp. 2d 713 (N.D. Tex. 2000). The court, incorrectly in my view, equated “inhabitant” with modern concepts of domicile, tying it to the narrow issue of public assessments. The problem with the court’s approach is that it collapses the constitutionally distinct terms “citizen” and “inhabitant.” The Eleventh Amendment prohibits federal jurisdiction over suits between one of the United States and “Citizens of another State.” The framers of the Twelfth Amendment rejected the term “Citizen” when formulating the restriction on voting in the electoral college, relying instead on the broader term “inhabitant.”
146. Id. art. I, § 2, cl. 1 (emphasis added).
147. Id. art. I, § 7, cl. 2 (emphasis added).
148. Id. art. II, § 1, cl. 3 (emphasis added).
in the House of Representatives: “by adding to the whole Number of free Persons . . . three fifths of all other Persons.”

So who are “the good People of these Colonies” in whose name the Declaration speaks? I believe that they are one American people, announcing their separation from the British people of whom they used to be a part. As Section II.B argues, the Declaration created an American nation. It thus makes eminent sense for the Declaration to speak in the voice of the constituent members of that nation. The American people that effect this change are a “People,” in the distinct sense of members of a political community. This sense is aptly captured in the Declaration’s opening lines: “WHEN in the Course of human Events, it becomes necessary for one People to dissolve the Political Bands which have connected them with another . . . .” The phrase “one people” is singular. It is not only that the Americans are united into “one”; they are united into one people. It is this one aggregate body—this one People—that now rises to assert independence from Britain. The singular sense of “one People” is reiterated when the Declaration asserts that George III is “unfit to be the Ruler of a free People.” Interestingly, the delegates did not sign the parchment copy of the Declaration as representatives of states; their signatures (unlike those of the Constitution) are undifferentiated and random. Indeed, as if to

149. Id. art. I, § 2, cl. 3 (emphasis added).

150. Prior to the revolution, few Americans would have doubted that they and the British were part of one people. For example, the New York Committee of Correspondence had resolved on July 19, 1774:

That we are one people, connected by the strongest ties of affection, duty, and interest, and that we lament as the greatest misfortune, every occurrence which has the least tendency to alienate or disturb that mutual harmony and confidence, which, if properly cultivated, could not fail rendering the British empire the admiration and envy of all the world.

Proceedings of the Committee of Correspondence, July 19, 1774, PA. GAZETTE, July 27, 1774 (emphasis added). The Declaration itself refers to “our British Brethren” and to the “Ties of our common Kindred.”

151. (emphasis added).

152. (emphasis added). In Jefferson’s draft, the phrase was “a people who mean to be free”; Congress’s tightening of the phrase increased its rhetorical power. See MAIER, supra note 3, at 147. Congress did not, however, change the singular to a plural—powerful evidence that the singular represents a conscious congressional choice.

153. This point is obscured by some printed editions of the Declaration that insist on grouping the delegates by state and inserting the name of their state as a heading. Although such alterations perhaps bring clarity to the document, and render its form closer to that of the Constitution, they nonetheless wreak editorial violence on the text. For the signatures as they actually appear on the parchment copy, see BECKER, supra note 114, at 192–93.
emphasize the unitary nature of the American people, the Declaration never mentions any state by name.

This reading best makes sense of the Declaration’s text. The Declaration is clearly not in the voice of the states or of the delegates to the Continental Congress. Nor can it comfortably sustain the proposition that it refers to thirteen independent peoples completely separating themselves both from each other and from the larger people of which they were once a part. The historical sources in Section II.C confirm that this reading is not simply a modern contrivance, but is amply supported by contemporary evidence.

B. What Kind of Independence Did They Declare?

The second part of the textual puzzle is to determine precisely what sort of legal change the Declaration effected. Plausible arguments might be made that the Declaration created one unitary nation-state, or that it created thirteen separate nations, completely independent of each other. The best reading of the text, however, is that the Declaration created a union of states, a confederated republic, which stood as one nation with respect to the rest of the world. It was the Declaration, not the Articles of Confederation or the Constitution, that initially united the colonies and gave birth to an American nation.

The debate, to the extent there is one, revolves primarily around this passage from the final paragraph of the Declaration:

We . . . solemnly Publish and Declare, That these United Colonies are, and of Right ought to be, FREE AND INDEPENDENT STATES; that they are absolved from all Allegiance to the British Crown, and that all political Connection between them and State of Great-Britain, is and ought to be totally dissolved; and that as FREE AND INDEPENDENT STATES, they have full Power to levy War, conclude Peace, contract Alliances, establish Commerce, and to do all other Acts and Things which INDEPENDENT STATES may of right do.

From this language it is easy to conclude that the Declaration created thirteen independent states. After all, it refers to “Free and Independent States,” not to one nation. And the passage strongly suggests that each state individually has the right to do all things which “Independent States” may of right do.

154. See, e.g., supra notes 92–93 and accompanying text.
This conclusion is much too quick, and the passage cannot support the weight that the thirteen nations argument would place on it. In its final paragraph, the Declaration declares “these United Colonies” to be “Free and Independent States.” Elsewhere, the Declaration simply refers to “these Colonies.” Yet here, the Declaration reiterates that the Colonies are united. Although the Declaration refers to the colonies as “Free and Independent,” in context this simply means that the colonies are free and independent of Great Britain; it does not at all follow that they are independent of each other. Interestingly, the Declaration never once applies the term “sovereign” to the states, and the passage most suggestive of state sovereignty is capable of an alternate reading. “They” (these Colonies), the Declaration claims, have “full Power to levy War,” and so on. But if the Declaration meant to create thirteen independent states, a much better formulation would have been that “each” has the “full Power to levy War,” and so on. In this context, “they” is ambiguous, and even restrictive. For example, one might say of members of the House of Representatives, “they have the power to impeach the President of the United States,” but no one would interpret that phrase as meaning that any individual member of the House has the power to impeach the President. They can only operate as a group; likewise, the Declaration most plausibly means that they (the colonies) can only levy war, and so on, as a group. It is surely of some significance that no state ever individually asserted any of the powers to which the thirteen nations argument would entitle them. No state on its own declared war; no state on its own contracted Alliances; and no state on its own established Commerce with foreign powers. Not one of the state constitutions drafted after the Declaration said one word about war, peace, or treaties. It was clear that these powers belonged to the Continental Congress, not to the states.

Twentieth-century orthography has blinded us to another powerful piece of textual evidence in this passage. In formal writing of the late eighteenth-century, nouns generally were capitalized; adjectives and pronouns were not. See Richard B. Morris, The Forging of the Union Reconsidered: A Historical Refutation of State Sovereignty over Seabeds, 74 COLUM. L. REV. 1056, 1072 (1974). See, e.g., THE PAPERS OF BENJAMIN FRANKLIN xii: JANUARY 1 THROUGH DECEMBER 31, 1768 (Leonard W. Labaree ed., 1959) (discussing Benjamin Franklin’s insistence on capitalizing all nouns). The Constitution displays such capitalization. For example, the Constitution repeatedly refers to a “supreme Court,” not to a “Supreme Court.” U.S. CONST. art. II, § 2, cl. 2; art. III, § 1;
Declaration of Independence

Declaration.\textsuperscript{158} The adjective/noun rule, for example, is evident in “separate and equal Station,” “unalienable Rights,” “light and transient Causes,” “absolute Despotism,” “former Systems,” “direct Object,” “public Records,” “manly Firmness,” “pretended Legislation,” “undistinguished Destruction,” and “native Justice.” The pronoun rule is evident in phrases such as “they should declare,” “they are endowed,” “alter or abolish it,” and “neglected to attend to them.” The adjective/noun rule is seemingly violated only in twelve places, five of which occur in the Declaration’s final passage. These five are “General Congress, Assembled,” “Supreme Judge of the World,” “United Colonies,” and, twice, “Free and Independent States.” The other seven are “Political Bands,” “Legislative Bodies,” “Representative Houses,” “Legislative Powers,” “Judiciary Powers,” “Standing Armies,” and “Armed Troops.” Yet these anomalies can be explained quite simply. These adjectives function not as separable adjectives, but as integral parts of the nouns themselves. In context, all of these noun phrases describe particular entities. A modern example might help illustrate the point. We do not capitalize “cuckoo clock,” but if nouns were required to be capitalized, the phrase would appear as “Cuckoo Clock.”

This evidence from elsewhere in the Declaration explains the significance of the capitalized phrases “United Colonies” and “Free and Independent States.” They must be read as complete noun phrases, not as nouns modified by adjectives. Thus the “United Colonies” are a distinct entity, not merely a group of colonies who happen to be united. The Declaration creates an entity, “Free and Independent States,” not States that happen to be free and independent. These “Free and Independent States” are an entity as surely as the “United States of America” of art. III, § 2, cl. 2. Although the capitalization rule has disappeared from modern English, it remains the rule in modern German.

158. In only three places does the Declaration fail to capitalize a noun: the word “causes” in the preamble, the word “consent” in the charges, and the word “act” in the conclusion to the charges. I can think of no logical reason for failing to capitalize these nouns; each appears capitalized elsewhere in the Declaration. This lapse is most likely an inadvertent oversight by the printer, who may have carelessly construed these words as verbs.

The argument here rests on the Dunlap broadside. The parchment, by contrast, leaves many nouns uncapsilized, \textit{see} BECKER, \textit{supra} note 114, at 191–92, a peculiarity that may reflect the personal quirks of its preparer. However, the parchment does capitalize “Free and Independent States,” and to the extent that this secret document is relevant at all, \textit{see supra} note 113 and accompanying text, it does not undercut my argument.
which the Constitution speaks. 159 Because we have lost the understanding of capitalization that was common knowledge in the eighteenth century, it is less obvious to us that the Declaration created an American nation. Yet many people pointed to precisely this final passage of the Declaration to support a nationalist reading. 160 It does not seem implausible that capitalization played some role in their arguments.

The case for a nationalist reading of the Declaration’s concluding passage is bolstered by other passages in the document. Most telling is the preamble: “WHEN in the Course of human Events, it becomes necessary for one People to dissolve the Political Bands which have connected them with another, and to assume among the Powers of the Earth, the separate and equal Station to which the Laws of Nature and of Nature’s God entitle them . . . .” This passage describes the American people as doing two distinct things. First, they are dissolving the political bands which have connected them with another. This is consistent with both the one-nation and the thirteen-nation argument. Second, they are assuming a “separate and equal Station” among the “Powers of the Earth.” This language, however, is almost impossible to reconcile with a thirteen-nation thesis. The people themselves are asserting a claim to an equal station (note the singular form) as a “Power[] of the Earth.” 161 This “one People” asserts equality with the nations of the world. 162 It is hard to imagine a more striking endorsement of the nationalist view. If the Declaration were creating thirteen independent nations, it ought to say “separate and equal Stations.” Furthermore, the Declaration charges George III with constraining “our fellow Citizens taken Captive on the high Seas to bear Arms against their Country.” 163 To what “Country” does this refer? Obviously, the country cannot be Great Britain. It must refer to America as a whole, a point nicely amplified in the phrase “fellow Citizens,” which connects Americans in all the states with the bonds of citizenship.

159. The Constitution specifically states that the United States is a singular entity when it extends federal judicial power to controversies “to which the United States shall be a party.” U.S. CONST. art. III, § 2, cl. 1 (emphasis added).
160. See Subsection II.C.2, infra.
161. Cf. Lucas, Justifying America, supra note 104, at 82 (“[T]he Laws of Nature and of Nature’s God’ refers to the doctrine of eighteenth-century international law that all nations are by nature ‘free, independent, and equal’ and entitled to the same rights and privileges.”)
163. (emphasis added).
Finally, it is worth noting that some of the language in the final paragraph of the Declaration comes directly from Richard Henry Lee’s motion in the Continental Congress of June 7, 1776. Lee’s motion stated, “That these United Colonies are, and of right ought to be, free and independent States, that they are absolved from allegiance to the British Crown, and that all political connection between them and the State of Great Britain is, and ought to be, totally dissolved.”\(^{164}\) In debate, Lee explained what he thought his motion would accomplish: “Let this happy day give birth to an American Republic!”\(^{165}\) Lee clearly saw his motion as creating a singular republic, not a multitude of independent states. This view is reiterated in the Declaration that Congress ultimately approved.

**C. The Historical Evidence**

The argument thus far has been grounded almost exclusively in the text of the Declaration. Interpreting the Declaration as the act of one American people creating an American nation best makes sense of the Declaration’s text and structure. But is such an interpretation grounded in historical reality? Although the evidence is not unambiguous, there is compelling historical support for this interpretation. This article does not undertake an exhaustive survey of the historical sources and does not aim to resolve conclusively any historical issues. Rather, I wish to assuage doubts that my reading of the Declaration is somehow deeply antithetical to its historical understanding.

Subsection One examines the understanding of the Declaration during the Revolution itself. Subsection Two addresses the debates surrounding the drafting and ratification of the Constitution. Subsection Three considers Independence Day celebrations in the early republic. Subsection Four analyzes the treatment of the Declaration in the earliest reported judicial decisions. Subsection Five focuses on the seminal work of Joseph Story, whose *Commentaries on the Constitution* address many of the issues investigated in this Article. Subsection Six evaluates the Articles of Confederation. Subsection Seven summarizes the main points of this Section.

\(^{164}\) 5 JOURNALS OF THE CONTINENTAL CONGRESS 425 (Worthington Ford ed., 1906). Note the change in the final version, which capitalized “free” and “independent.”

1. The Declaration in the Throes of the Revolution

Pauline Maier has carefully traced how the momentum for independence grew in towns and villages across America. The “authority” of the people of the United States that the delegates to the Second Continental Congress invoked was firmly grounded in the actions of these communities, in which the people themselves debated and resolved the pressing issue of independence. On May 27, 1776, the town of Malden, Massachusetts, instructed its delegate that “it is now the ardent wish of our soul that America may become a free and independent state” and called for the creation of “an American republic.” Topsfield, Massachusetts, expressed its hope that Congress would “declare America to be independent of the Kingdom of Great Britain.” The Pennsylvania Provincial Conference declared its willingness to “concur in a vote of the Congress declaring the United Colonies free and independent States, provided the forming of the Government, and the regulation of the internal police of this colony, be always reserved to the people of the said Colony.” This statement clearly implies that a Declaration of Independence would effect the creation of a new nation, with the states reserving power over their internal affairs only.

That the Declaration had done precisely this was apparent from the very beginning. On July 4, 1776, the Continental Congress appointed Benjamin Franklin, John Adams, and Thomas Jefferson to design a great seal for the United States America, a truly odd assignment if the Declaration had merely created thirteen independent nations. John Hancock, President of the Continental Congress, sent copies of the Declaration to the several states, observing, “The important Consequences to the American States from this Declaration of

166. See MAIER, supra note 3, at 47–96.
167. See Morris, supra note 156, at 1069–71 (discussing the legitimacy of the Continental Congress’s claim to powers delegated directly from the people).
169. Resolution of the Town of Topsfield, Massachusetts, June 21, 1776, reprinted in MAIER, supra note 3, at 233–234.
Declaration of Independence

Independence, considered as the Ground & Foundation of a future Government, will naturally suggest the Propriety of proclaiming it in such a Manner, that the People may be universally informed of it.”172 Ezra Stiles, who two years later would become President of Yale College, received his copy of the Declaration on July 13, 1776. Upon reading it, he declared, “Thus the CONGRESS have tied a Gordian knot, which the Parliament will find they can neither cut nor untie. The thirteen united Colonies now rise into an Independent Republic among the kingdoms, states and empires on earth.”173 Savannah, Georgia, celebrated the Declaration on August 10, 1776. The President and the Council of Georgia read the Declaration in the Council chamber.174 They then proceeded to a square in front of the Assembly House and read the Declaration “before a great Concouse of people.”175 They read the Declaration again in front of the Georgia Battalion.176 Finally, they proceeded to the Battery and read the Declaration for a fourth time.177 That evening the Council buried a representation of George III and roused the people with their understanding of what the now very familiar Declaration meant: “America is free and independent; that she is, and will be, with the blessing of the Almighty, great among the nations of the earth.”178

The state constitutions adopted after the Declaration did nothing to contradict this. None of these constitutions claimed any power over war and peace or any other incident of national sovereignty. Indeed, ten of these constitutions specifically prescribed methods for electing delegates to the Continental Congress, an implicit recognition of Congress’s legitimacy and authority.179

In a 1776 speech in the Continental Congress, Benjamin Rush, a signer of the Declaration, noted that “We are now a new Nation,” and

175. Id.
176. Id.
177. Id. at 323.
178. Id.
179. See Morris, supra note 156, at 1071.
proclaimed himself a “Citizen of America.” To Rush, it was clear that the states were “dependent on each other—not totally independent states.” Rush’s views were far from atypical. In 1777, before the Articles of Confederation were even proposed to the states, much less ratified, the Continental Congress, after a lengthy debate, found itself equally divided on the issue of whether states had an inherent right to meet with each other without the approbation of Congress. If the Declaration had simply created thirteen independent nations, that issue should have been easily resolved.

In 1778, a writer to the Pennsylvania Gazette requested that the newspaper republish the Declaration of Independence. The writer stated:

Too much cannot be said in Favour of this excellent Composition. It is to the true Whig what the Bible is to the true Christian—it contains his Right to the fair Inheritance of true Liberty. It cannot be too much admired, nor too often read by every American . . . . In order to excite the good People of the United States to a Remembrance of their Birth into a world of Freedom [you are requested to re-publish the Declaration].

For this writer, the Declaration was about the “birth” of the “People of the United States.” With the Declaration, the People of the United States had announced themselves on the world stage.

Another writer in the Pennsylvania Gazette discoursed on the relation between the states in 1780, at which point the Articles of Confederation had not yet gone into effect. Maryland had enacted a law that, read literally, seemed to prohibit the export of provisions from Maryland to


181. Id. at 247. Peter Onuf points out that after the Declaration, “The American states did not seem to be acting as independent sovereignties were expected to behave; they did not act like true states.” PETER ONUF, THE ORIGINS OF THE FEDERAL REPUBLIC 3 (1983). With regard to territorial disputes, “state sovereignty was identified not with the will to make and enforce claims but with claims that could be made by right, and that should be upheld by all the states, individually and collectively.” Id. at 7.

182. RAKOVE, supra note 102, at 165–66.

183. PA. GAZETTE, June 13, 1778. The phrase “Birth into a world of Freedom” foreshadows similar natal imagery in Abraham Lincoln’s Gettysburg Address: “Four score and seven years ago our fathers brought forth on this continent, a new nation, conceived in Liberty, and dedicated to the proposition that all men are created equal . . . . [We here highly resolve] that this nation, under God, shall have a new birth of freedom.” Abraham Lincoln, Address Delivered at the Dedication of the Cemetery at Gettysburg (Nov. 19, 1863), in ABRAHAM LINCOLN: GREAT SPEECHES 103–04 (John Grafton ed., 1991) (emphasis added).
The writer defended the Maryland statute, arguing that it could not possibly have such a meaning. Invoking the first line of the Declaration of Independence, he argued:

We are but one people, from one end of the continent to the other (I speak only of the United States) though under our respective forms of government: The union, the confederacy, necessarily imply mutual assistance and free intercourse and commerce. . . . An act of any State to the purpose would be void in itself, as being against the fundamental laws of the union. . . . Every man knows that no State can constitutionally enact a law contrary to the spirit of the Union with the other States. 185

This writer recognized that the “one people” of America were largely governed by their respective state governments, but he also recognized that there were “fundamental laws of the Union” that the states could not breach. Since the Articles were not yet in effect, he could have been referring only to the Declaration of Independence or the nationalism that it affirmed. If this proposition had been at all remarkable, it is doubtful the writer would have claimed that “every man” knew it to be true.

A year earlier the Congress had officially asserted that it possessed “the supreme sovereign power of war and peace.” 186 This power gave Congress the right “ultimately and finally to decide on all matters and questions touching the law of nations.” 187 Since the Articles of Confederation had not yet gone into effect, the source of this power could only be the Declaration of Independence.

In December of 1780, the Continental Congress affirmed the importance of the Declaration to the governance of America by appointing a committee “to collect and cause to be published 200 correct copies of the declaration of independence, the articles of confederation and perpetual union, the alliances between these United States and his Most Christian Majesty, with the constitutions or forms of government of the several states.” 188 Were the Declaration simply the “propaganda

184. See PA. GAZETTE, Feb. 9, 1780.
185. Id.
187. Id. at 284. The Congress and its ambassadors in Europe also issued passports and oaths of citizenship in the name of the United States. Morris, supra note 156, at 1087.
188. 18 JOURNALS OF THE CONTINENTAL CONGRESS 1217 (Gaillard Hunt ed., 1910). The final product was UNITED STATES CONTINENTAL CONGRESS, supra note 114. Subsequent collections of the federal and state constitutions included the Declaration as their first document. See, e.g., THE
statement” that later scholars have declared it to be, its inclusion in this collection of governing documents would have been most inappropriate.

2. The Debates Surrounding the Drafting and Ratification of the Constitution

The Constitutional Convention and the subsequent ratification struggle raised important questions about the relation between the states and the union. These debates helped bring the Declaration into proper focus, as both proponents and opponents of the new Constitution were called on to explain just what the Declaration stood for.

The meaning of the Declaration was discussed in an important exchange in the Constitutional Convention on June 19, 1787. In the course of a debate about the scope of state authority, Massachusetts delegate Rufus King observed that the states “were not ‘Sovereigns’ in the sense contended for by some.” They lacked the “peculiar features of sovereignty, they could not make war, nor peace, nor alliances, nor treaties.” A “Union of the States,” King noted, “is a Union of the men composing them, from whence a national character results to the whole.” If the states had “formed a confederacy in some respects—they formed a Nation in others.”

Maryland’s Luther Martin, who would become a virulent Anti-Federalist, responded that the Declaration had authorized no such thing. For Martin, the Declaration had “placed the 13 States in a state of Nature towards each other” and created thirteen independent sovereignties.

This notion was immediately denounced by James Wilson of Pennsylvania, one of America’s most brilliant lawyers and one of the
Convention’s most able men.\textsuperscript{195} Wilson, one of only six men to sign both the Declaration and the Constitution, replied that Martin fundamentally misunderstood what the Declaration meant. Wilson noted that the Declaration of Independence preceded the state constitutions.\textsuperscript{196} He could “not admit the doctrine that when the Colonies became independent of Great Britain, they became independent also of each other.”\textsuperscript{197} Wilson read the Declaration of Independence and observed that “the United Colonies were declared to be free and independent States” and that they were “independent, not individually but Unitedly.”\textsuperscript{198} Alexander Hamilton heartily agreed.\textsuperscript{199}

Both Hamilton and Wilson expressed similar views outside the Convention. In February of 1787, Hamilton explicitly tied the Declaration to nationhood in a speech in the New York Assembly. Hamilton read verbatim the sentence of the Declaration that declared the United Colonies “free and independent states.”\textsuperscript{200} To Hamilton, this meant that

the union and independence of these states are blended and incorporated in one and the same act; which taken together clearly, imports, that the United States had in their origin full power to do all acts and things which independent states may of right do; or in other words, full power of sovereignty.\textsuperscript{201}

In the Pennsylvania ratifying convention, Wilson articulated a strong nationalist view of the Declaration:

I consider the people of the United States, as forming one great community; and I consider the people of the different states as forming communities again on a lesser scale. . . .


\textsuperscript{196} Notes of Robert Yates, June 19, 1787, in DRAFTING THE CONSTITUTION, supra note 190, at 75 (remarks of James Wilson).

\textsuperscript{197} Notes of James Madison, June 19, 1787, in DRAFTING THE CONSTITUTION, supra note 190, at 1053 (remarks of James Wilson).

\textsuperscript{198} Id. Wilson had made a virtually identical argument in a 1785 publication. See JAMES WILSON, CONSIDERATIONS ON THE BANK OF NORTH AMERICA, reprinted in 2 THE WORKS OF JAMES WILSON 824, 829 (Robert Green McCloskey ed., 1967).

\textsuperscript{199} Notes of James Madison, June 19, 1787, in DRAFTING THE CONSTITUTION, supra note 190, at 1053, 1057 (remarks of Alexander Hamilton).


\textsuperscript{201} Id.
. . . I view the states as made for the People, as well as by them, and not the People as made for the states; the People, therefore, have a right, whilst enjoying the undeniable powers of society, to form either a general government, or state governments, in what manner they please; or to accommodate them to another; and by this means preserve them all; this, I say, is the inherent and unalienable right of the people, and as an illustration of it, I beg to read a few words from the Declaration of Independence, made by the representatives of the United States and recognized by the whole Union.202

James Madison can likewise be counted among those with a nationalist interpretation of the Declaration. In The Federalist No. 45, he asked,

Was, then, the American Revolution effected, was the American Confederacy formed, was the precious blood of thousands spilt . . . not that the people of America should enjoy peace, liberty, and safety, but that the government of the individual States . . . might enjoy a certain extent of power and be arrayed with certain dignities and attributes of sovereignty?203

For Madison, to ask the question was to answer it. Much later in life, Madison conferred with Thomas Jefferson on the appropriate subjects of study for the University of Virginia Law School. Madison’s sketch of a course outline began, “And on the distinctive principles of the Government of our own State, and that of the U. States, the best guides are to be found in—1. The Declaration of Independence, as the fundamental act of Union of these States.”204

Charles Cotesworth Pinckney, a signer of the Constitution, offered a similar reading of the Declaration in the South Carolina convention in 1788. Pinckney argued, “[T]he Declaration of Independence . . . sufficiently refutes the doctrine of the individual sovereignty and


204. Letter from James Madison to Thomas Jefferson (Feb. 8, 1825), in 9 THE WRITINGS OF JAMES MADISON 218, 221 (Gaillard Hunt ed., 1910) (emphasis added).
independence of the several States.”²⁰⁵ He noted, “The several States are not even mentioned by name in any part, as if it was intended to impress the maxim on America that our freedom and independence arose from our union,” and urged his colleagues to “consider all attempts to weaken this union, by maintaining that each State is separately and individually independent, as a species of political heresy.”²⁰⁶

In The Federalist No. 2, John Jay stated:

To all general purposes we have uniformly been one people—each individual citizen everywhere enjoying the same national rights, privileges and protection. As a nation we have made peace and war—as a nation we have vanquished our common enemies—as a nation we have formed alliances and made treaties, and entered into various compacts and conventions with foreign states.²⁰⁷

The text of the Constitution confirms the nationalist reading of the Declaration of Independence that was offered during the ratification debates. Article One of the Constitution requires Representatives to have been citizens of the United States for seven years, and Senators to have been citizens for nine years.²⁰⁸ Because the Constitution was drafted in 1787, these requirements logically require United States citizenship to have existed in 1778. There must therefore have been an entity called the United States of which an individual could be a citizen. If so, this status would have nothing to do with the Articles of Confederation, which were ratified in 1781, but could only be a result of the Declaration of Independence. Indeed, that the Declaration created American citizenship is clearly recognized in American law.²⁰⁹ Article Six of the Constitution states that “all Treaties made, or which shall be made, under the Authority of the United States, shall be the supreme Law of the Land.”²¹⁰ This means that treaties executed prior to the Constitution are valid and binding. There is nothing to suggest that treaties made by the United States prior to the Articles of Confederation would be excluded from this clause. The United States entered into many treaties prior to the

²⁰⁵ 1 J OSEPH STORY, COMMENTS ON THE CONSTITUTION OF THE UNITED STATES 150 (Boston, Little & Brown 1873) (1st ed. 1833) (quoting DEBATES IN SOUTH CAROLINA, 1788, at 43 (Charleston, A.E. Miller 1831)) [hereinafter 1 J OSEPH STORY].
²⁰⁶ Id.
²⁰⁸ U.S. CONST. art. I, § 2, cl. 2 & art. I, § 3, cl. 3.
²⁰⁹ See supra notes 64–68 and accompanying text.
²¹⁰ U.S. CONST. art. VI, cl. 2 (emphasis added).
ratification of the Articles in 1781, and since treaties were traditionally only entered into between sovereign nations, the clear implication is that the United States, as one entity, possessed all the powers of external sovereignty at least since the Declaration of Independence. Finally, it was worth noting how the opening and closing phrases of the Constitution evoke the Declaration of Independence. The phrase “We the People” is not significantly different from the “one people” that announces itself in the preamble to the Declaration. The Constitution closes by citing the Declaration as a temporal reference point: “Done in convention . . . in the year of our Lord one thousand seven hundred and Eighty seven and of the Independence of the United States of America the Twelfth.”

3. **The Celebration of Independence Day**

Fourth-of-July celebrations have seldom generated much interest among legal scholars, yet they are a fascinating window into how early Americans perceived the Declaration of Independence. Americans celebrated the anniversary of the Declaration with energy and enthusiasm beginning in 1777 and continuing until the present day. No one celebrated the days on which the colonial legislatures had authorized their congressional delegates to vote for independence; no one celebrated July 2, the date Congress actually voted for independence; no one celebrated the date of the ratification of the Articles of Confederation; no one celebrated the dates of the signing or ratification of the Constitution; no one celebrated the dates on which the new federal government under the Constitution began operating. But Americans consistently celebrated the anniversary of the Declaration, telling evidence of the importance of the Declaration in the development of American national identity.

The first anniversary of the Declaration in 1777 was celebrated with cannon salutes, decorated navy ships, military parades, ringing church

---

211. *Id*. art VII.

Of course, Article VII also provides that the ratification of only nine states would be sufficient for establishing the Constitution between those nine states. Although the Constitution does not specify the status of a non-ratifying state, it presumably would have been free to go its own way. *See The Federalist No. 43, at 286 (James Madison) (Isaac Kramnick ed., 1987) (noting the “merely hypothetical” prospect of non-ratifying states). This demonstrates not so much that the United States was not one nation, at least to the great attributes of sovereignty, but that it was not yet “one nation, indivisible.”
bells and extravagant fireworks across the eastern seaboard.\textsuperscript{212} One Pennsylvania newspaper declared, “Thus may the fourth of July, that glorious and ever memorable day, be celebrated through America, by the sons of freedom, from age to age till time shall be no more.”\textsuperscript{213} By the time of the drafting of the Constitution, the Fourth of July and the Declaration had become tied to a strong sense of American national identity. The \textit{Pennsylvania Herald} observed in 1787 that the “auspicious Fourth of July, which crowned the toils of America with freedom and sovereignty has been commemorated in every district of the continent, with the fullest demonstrations of joy and gratitude.”\textsuperscript{214} The paper expressed its hope that the Federal Convention would produce “a system of government adequate to the security and preservation of those rights, which were promulgated by the ever-memorable Declaration of Independency.”\textsuperscript{215} In 1788, the “largest, most lavish procession ever seen in the United States” was held in Philadelphia on the Fourth of July.\textsuperscript{216} The “Grand Federal Procession,” as it was called, was over a mile and half long and involved over five thousand people.\textsuperscript{217} Intended as a massive propaganda display on the part of Pennsylvania Federalists, the procession pointedly linked the proposed new Constitution with the celebration of the Declaration of Independence.\textsuperscript{218} The event ended with a speech by James Wilson before a crowd of approximately 17,000 people.\textsuperscript{219} As Francis Hopkinson subsequently wrote to Thomas Jefferson, “Nothing can equal the rejoicings in the cities towns & villages thro’out the States on the late fourth of July in Celebration of the Declaration of Independence & the Birth of the new Constitution.”\textsuperscript{220}

\begin{footnotesize}
\begin{enumerate}
\item \textit{Id.} at 24 (quoting PA. EVENING POST, July 5, 1777).
\item \textit{Id.}
\item \textit{Id.} supra note 212, at 71.
\item \textit{Id.} at 71–72.
\item \textit{Id.} at 78.
\item \textit{Id.} at 77. The speech was not a complete success. “Owing to some mistake, the cannon began firing just as he began to speak, so that no one could understand anything he said.” \textit{Id.} (quoting CHARLES BIDDLE, \textit{The Autobiography of Charles Biddle, Vice-President of the Supreme Executive Council of Pennsylvania} 226 (Philadelphia, E. Claxton 1853)).
\item \textit{Letter from Francis Hopkinson to Thomas Jefferson (July 17, 1788), in 18 The Documentary History of the Ratification of the Constitution} 270, 271 (John P. Kaminski & Gaspare J. Saladino eds., 1995). One such celebration was in New Haven, Connecticut, where the Declaration of Independence was read in a ceremony with “uncommon splendour.” \textit{Id.} at 235.
\end{enumerate}
\end{footnotesize}
When George Washington was attempting to discern the protocol suitable to his new office as President of the United States, Alexander Hamilton recommended limiting public entertainments, but insisted that an entertainment be held on the “day of the Declaration of Independence.” 221

By 1791, people were referring to the Fourth of July as a “natal day,”222 and, at least by 1815, it was common to speak of the Fourth as “the National Birth Day.”223 Thomas Jefferson referred to Fourth of July celebrations as “an anniversary assemblage of the nation on its birthday,”224 and to the Fourth of July as “our nation’s birthday.”225 As Jefferson explained, the Declaration of Independence “made us a nation.”226 In 1821, Secretary of State John Quincy Adams attended the celebration of the Fourth of July in Washington, D.C. He read from the original Declaration of Independence that was in the custody of the Department of State, and observed,

The interest, which in this paper has survived the occasion upon which it was issued; the interest which is of every age and every clime; the interest which quickens with the lapse of years, spreads as it grows old, and brightens as it recedes, is in the principles which it proclaims. It was the first solemn declaration by a nation of the only legitimate foundation of civil government. . . . It announced in practical form to the world the transcendent truth of the unalienable sovereignty of the people.227

This theme was echoed in the last letter Thomas Jefferson ever wrote. Referring to the upcoming fiftieth anniversary of the Declaration of

222. INDEPENDENT GAZETTEER, July 2, 1791, quoted in Travers, supra note 212, at 110.
223. BOSTON GAZETTE, July 15, 1815, quoted in Travers, supra note 212, at 4. In 1871, Congress specifically declared that July 4, 1776, was the “birthday of the nation.” Act of Mar. 3, 1871, ch. 105, 16 Stat. 470.
227. JOHN QUINCY ADAMS, AN ADDRESS DELIVERED AT THE REQUEST OF A COMMITTEE OF THE CITIZENS OF WASHINGTON, ON THE OCCASION OF READING THE DECLARATION OF INDEPENDENCE ON THE FOURTH OF JULY, 1821 (Washington, 1821), quoted in Detweiler, supra note 202, at 574 (emphasis added).
Independence, Jefferson stated, “May [the Declaration] be to the world what I believe it will be . . . the signal of arousing men to burst the chains under which monkish ignorance and superstition had persuaded them to bind themselves, and to assume the blessings and security of self-government.” 228 Jefferson urged, “let the annual return of this day forever refresh our recollections of these rights, and an undiminished devotion to them.” 229

In 1831, John Quincy Adams again spoke at a Fourth of July ceremony and affirmed that the “declaration was joint, that the united colonies were free and independent states, but not that any one of them was a free and independent State, separate from the rest.” 230 Rather, “The Declaration of Independence announced the severance of the thirteen united colonies from the rest of the British Empire, and the existence of their people, from that day forth, as an independent nation.” 231

In all of these Fourth of July celebrations, Americans affirmed their understanding of the Declaration as grounded in the inherent sovereignty of the American people. They did not understand the Declaration as having created thirteen independent nations; rather, the Declaration marked the Americans’ birth as a new people and as a new nation.

4. Early Court Cases

Further evidence of the role of the Declaration in establishing one American nation can be found in the earliest decisions of American courts. The cases, although not unambiguous, strongly suggest that American courts recognized the Declaration as the act of one people creating an American nation.

228. Letter from Thomas Jefferson to Roger Weightman, June 24, 1826, quoted in THE LIFE AND SELECTED WRITINGS OF THOMAS JEFFERSON, supra note 226, at 665–66. Both Thomas Jefferson and John Adams would die on July 4, 1826, the fiftieth anniversary of the Declaration of Independence, a coincidence which Americans imbued with mystical significance. See TRavers, supra note 212, at 220.


230. 1 Joseph Story, supra note 205, at 150 n.3. (quoting John Adams).

231. Id. Justice Scalia recently echoed this view in his dissent in Lee v. Weisman, 505 U.S. 577 (1992), where he describes the Declaration as “the document marking our birth as a separate people.” Id. at 633 (Scalia, J., dissenting) (emphasis added). Scalia’s description is similar to that of Justice Bradley, who, in the Slaughter-House Cases, referred to the Declaration as “the first political act of the American people in their independent sovereign capacity.” Slaughter-House Cases, 83 U.S. (16 Wall.) 36, 116 (1873) (Bradley, J., dissenting).
The first discussion of the Declaration in the Supreme Court appears in Chief Justice Jay’s opinion in *Chisholm v. Georgia.* 232 In that case, the Court unanimously held that Article III of the Constitution subjected states to suit in federal court by citizens of another state. 233 Jay’s opinion addressed the legal effect of the Declaration of Independence:

The revolution, or rather the Declaration of Independence, found the people already united for general purposes, and at the same time providing for their more domestic concerns by state conventions, and other temporary arrangements. From the crown of Great Britain, the sovereignty of their country passed to the people of it; and it was then not an uncommon opinion, that the unappropriated lands, which belonged to that crown, passed not to the people of the colony or states within whose limits they were situated, but to the whole people; on whatever principles this opinion rested, it did not give way to the other, and thirteen sovereignties were considered as emerged from the principles of the revolution, combined with local convenience and considerations; the people nevertheless continued to consider themselves, in a national point of view, as one people; and they continued, without interruption, to manage their national concerns accordingly . . . . 234

Jay’s opinion confirms the idea that the Declaration was a national act of one people, creating one nation, at least with respect to “national affairs.”

The 1795 case of *Penhallow v. Doane’s Administrators* 235 raised the issue of the status of the Declaration in even more pointed fashion. The case arose out of a New Hampshire admiralty dispute dating back to 1777. In 1779 and 1783, the Continental Congress had asserted its

---

232. 2 U.S. (2 Dall.) 419 (1793).
233. Id. at 424. This particular holding was altered by the ratification of the Eleventh Amendment. But it is unlikely that the Supreme Court misinterpreted Article III. Two of the five justices who heard the case, James Wilson and John Blair, had been delegates to the Constitutional Convention, where Wilson took a leading role. Chief Justice John Jay was one of the authors of *The Federalist Papers,* and Justice William Cushing had played a prominent part in the Massachusetts ratifying convention. Although this is not the place to enter into the thicket of Eleventh Amendment controversies, the best interpretation is that the Eleventh Amendment simply limits a particular class of diversity cases, not federal question suits brought against states. See Amar, supra note 101, at 1474–75.
235. 3 U.S. (3 Dall.) 54 (1795).
power, through its Committee of Appeals, to review the earlier New Hampshire state court decisions in the case. The 1795 proceedings forced the Supreme Court to determine the validity of this jurisdiction, and by extension, the powers of the Continental Congress prior to the Articles of Confederation. Both parties were represented by exceptionally able counsel, and oral arguments extended over eight days.

The plaintiffs in error contended that the Continental Congress lacked the power to review prize decisions of the state courts: “The colonies, totally independent of each other before the war, became distinct, independent states, when they threw off their allegiance to the British crown, and congress was no longer a convention of agents for colonies, but of ambassadors from sovereign states.” The Declaration, they contended, “made no difference as to the sovereignty of the several States”; Congress was merely an advisory body and possessed only the power to recommend.

The defendants in error vehemently challenged these assertions. As they saw it, “On the declaration of independence, a new body politic was created; congress was the organ of the declaration; but it was the act of the people, not of the state legislatures, which were likewise nothing more than organs of the people.” The Continental Congress had a “national sovereignty” and derived its authority “from the whole people of America, as one united body.”

237. See id. at 394.
238. Penhallow, 3 U.S. (3 Dall.) at 67–68. Justice Paterson’s notes of the oral arguments confirm that the appellants specifically linked this argument to the Declaration of Independence. See William Paterson’s Notes of Arguments in the Supreme Court (Feb. 6, 1795), in 6 THE DOCUMENTARY HISTORY OF THE SUPREME COURT OF THE UNITED STATES, supra note 236, at 428, 431.
239. William Paterson’s Notes of Arguments in the Supreme Court (Feb. 10, 1795), in 6 THE DOCUMENTARY HISTORY OF THE SUPREME COURT OF THE UNITED STATES, supra note 236, at 448–49.
240. Penhallow, 3 U.S. (3 Dall.) at 76.
241. Id. In the original notes, the argument is phrased “On the decl’n of independance, the sovereignty was vested in Congress,” William Paterson’s Notes of Arguments in the Supreme Court (Feb. 11, 1795), in 6 THE DOCUMENTARY HISTORY OF THE SUPREME COURT OF THE UNITED STATES, supra note 236, at 457–58, and “The King of Britain was formerly Sovereign of U.S. The people divested the King, of it, & placed it in Congress,” William Tilghman’s Notes of Arguments in the Supreme Court (Feb. 11, 1795), in 6 THE DOCUMENTARY HISTORY OF THE SUPREME COURT OF THE UNITED STATES, supra note 236, at 461, 463.
pointed out, led to the fairly ridiculous conclusion that the Declaration had actually lessened the power of the Continental Congress. ②42 If there was no national government until the Articles of Confederation were ratified in 1781, as the plaintiffs maintained, then “How did we get along?” the defendants asked sarcastically. “Were there 13 different wars?” ①243

The Supreme Court was thus squarely confronted with the legal status of the Declaration. The Court minced no words in upholding the power of the Continental Congress prior to the Articles of Confederation and the importance of the Declaration in forming one American nation. The opinion of Justice William Paterson (who had proposed the New Jersey plan at the Constitutional Convention and was certainly no enemy of state rights) stated,

[I]t became necessary for the people or colonies to coalesce and act in concert . . . they accordingly grew into union, and formed one great political body, of which congress was the directing principle and soul. As to war and peace, and their necessary incidents, congress, by the unanimous voice of the people, exercised exclusive jurisdiction, and stood, like Jove, amidst the deities of old, paramount, and supreme. The truth is, that the states, individually, were not known nor recognised as sovereign, by foreign nations, nor are they now. ①244

For Paterson, “These high acts of sovereignty were submitted to, acquiesced in, and approved of, by the people of America.” ①245 Justice John Blair agreed. He observed that the Continental Congress “in short, acted in all respects like a body completely armed with all the powers of war; and at all this, I find not the least symptom of discontent among all the confederated states, or the whole people of America.” ①246 Indeed, “congress were universally revered, and looked up to as our political fathers, and the savours of their country.” ①247 Justice Iredell recognized

①242. See William Paterson’s Notes of Arguments in the Supreme Court (Feb. 9, 1795), in 6 THE DOCUMENTARY HISTORY OF THE SUPREME COURT OF THE UNITED STATES, supra note 236, at 434, 439.

①243. Id. at 440.

①244. Penhallow, 3 U.S. (3 Dall.) at 81 (Paterson, J.).

①245. Id. at 80. According to Paterson, the only way New Hampshire could have avoided the power of the Continental Congress was to withdraw completely from the union. Id. at 82.

①246. Id. at 111 (Blair, J.).

①247. Id.
the Committee’s jurisdiction on slightly different grounds. Iredell felt that any authority exercised by the Continental Congress could only be conveyed by the people of each state acting individually, not “by all the people in the several provinces, or states, jointly.” Nonetheless, Iredell found it “unquestionable” that Congress, with the “acquiescence of the states” exercised the “high powers” of “external sovereignty,” and that the “articles of confederation amounted only to a solemn confirmation of it,” A state could only avoid this power by “withdrawing from the confederation” completely.

These views did not go completely unchallenged. In the 1796 case of *Ware v. Hylton*, Justice Chase suggested that the Declaration did not mean that the “united colonies jointly, in a collective capacity, were independent states . . . but that each of them was a sovereign and independent state.” Nonetheless, Chase conceded that the Continental Congress was effectively the organ of one nation, admitting that “the several states retained all internal sovereignty; and that congress properly possessed the great rights of external sovereignty.” In “deciding on the powers of congress . . . before the confederation,” Chase “[saw] but one safe rule, namely, that all the powers actually exercised by congress, before that period, were rightfully exercised, on the presumption not to be controverted, that they were so authorized by the people they represented.

In the early nineteenth century, lawyers repeatedly cited the nationalist implications of the Declaration. As William Rawle argued in 1805, “The independence of America was a national act. The avowed object was to throw off the power of a distant country; to destroy the political

248. Id. at 94 (Iredell, J.).
249. Id. at 91.
250. Id. at 96.
251. Id. at 95.
252. 3 U.S. (3 Dall.) 199 (1796).
254. In oral argument, John Marshall contended both that “Virginia . . . was an independent nation” in 1777, and that “America . . . must, from the 4th of July 1776, be considered as independent a nation as Great Britain.” *Ware*, 3 U.S. (3 Dall.) at 210–11 (oral argument of John Marshall).
255. On the other side, Mr. Lewis contended that the revolutionary war was “waged against all America, as one nation.” Id. at 219 (oral argument of Attorney Lewis) (emphasis added).
256. *Ware*, 3 U.S. (3 Dall.) at 232 (Chase, J.).
subjection; to elevate ourselves from a provincial to an equal state in the great community of nations.”256 Another lawyer argued that, “By the declaration of independence, the colonies became a separate nation from Great Britain.”257 In an 1817 case, an attorney pointed out that “by express decisions on the point, the principle was settled that our existence as an independent nation commenced with our declaration of independence in 1776.”258 Chief Justice John Marshall, sitting on circuit, agreed with that proposition.259

5. The Treatise of Joseph Story

The first great American constitutional law treatise is the 1833 treatise by Joseph Story, one of the towering figures in the history of American law. Story’s treatise synthesized many of the elements discussed in this section and firmly endorsed the nationalist understanding of the Declaration of the Independence.

Story argued that the Continental Congress derived its powers directly from the people, and exercised its authority “not as the delegated agents of the governments de facto of the colonies, but in virtue of original powers derived from the people.”260 The period surrounding the adoption of the Declaration was important, Story maintained, because of its relevance to “several very important considerations respecting the political rights and sovereignty of the several colonies.”261 For Story, the Declaration was manifestly the act of one American people:

[The Declaration of Independence] was not an act done by the State governments then organized; nor by persons chosen by them. It was emphatically the act of the whole people of the united

257. Lambert’s Lessee v. Paine, 7 U.S. (3 Cranch) 97, 125 (1805) (oral argument of Attorney Mason) (emphasis added); see also id. at 114 (oral argument of Francis Scott Key) (“By . . . the declaration of independence . . . a new sovereignty was created.”); Kilham v. Ward, 2 Mass. 236, 256 (1806) (oral argument of Nathan Dane) (arguing that a person became an American citizen on July 4, 1776, the date of “the birth of the American nation”) (emphasis omitted).
259. See id. However, Marshall found that this did not mean that the American judiciary was free to recognize as a foreign nation any entity that had issued a declaration of independence; rather, the courts could only recognize those foreign nations that the executive had already recognized. See id. at 441–42.
260. STORY, supra note 205, at 140.
261. Id. at 144.
colonies, by the instrumentality of their representatives, chosen for that, among other purposes . . . . It was an act of original, inherent sovereignty by the people themselves, resulting from their right to change the form of government, and to institute a new one, whenever necessary for their safety and happiness. . . . The people of the united colonies made the united colonies free and independent states, and absolved them from all allegiance to the British crown. The declaration of independence has accordingly always been treated, as an act of paramount and sovereign authority, complete and perfect per se, and ipso facto working an entire dissolution of all political connection with and allegiance to Great Britain. And this, not merely as a practical fact, but in a legal and constitutional view of the matter by courts of justice. 262

Story forcefully rejected the notion that the American people had created thirteen independent states. The Declaration placed the states “under the dominion of a superior controlling national government, whose powers were vested in and exercised by the general congress with the consent of the people of all the states.” 263 Accordingly, “[f]rom the moment of the declaration of independence . . . the united colonies must be considered as being a nation de facto, having a general government over it created, and acting by the general consent of the people of all the colonies.” 264

6. A Note on the Articles of Confederation

At first glance, the Articles of Confederation would seem to contradict the idea that an American nation emerged with the adoption of the Declaration of Independence. The second Article states, “Each state retains its sovereignty, freedom, and independence, and every power, jurisdiction, and right, which is not by this Confederation expressly delegated to the United States, in Congress assembled.” This section, and other provisions of the Articles (such as voting by states), suggest that the Declaration had created thirteen independent nations. 265 With the Articles, these nations confederated together for limited purposes, but each nonetheless retained the fundamentals of

262. *Id.* at 149–50.
263. *Id.* at 152.
264. *Id.* at 152–53.
265. ARTICLES OF CONFEDERATION art. II.
nationhood. Although there is much to be said for this argument, it does not necessarily undermine the idea that the Declaration created one nation, at least with respect to the rest of the world. Indeed, the most plausible reading of the Articles is that they are merely declarative of what the Declaration had already accomplished.

The first thing worth noting about the Articles is that we got along quite well without them. The Continental Congress did not bother to approve them until November 15, 1777, and they did not go into effect until March of 1781. Yet throughout this period, the Continental Congress acted as a de facto national government. It conducted almost the entire War of Independence against one of the most powerful nations on earth. It requisitioned supplies, supported, trained, and governed the Continental Army, organized a navy, negotiated treaties with foreign nations, sent American ambassadors throughout Europe, heard appeals from state courts in admiralty suits, and issued currency. In short, the Congress acted as if it were the representative of one American nation.


267. For an account of the ratification debates over the Articles of Confederation, see Eric M. Freedman, Why Constitutional Lawyers and Historians Should Take a Fresh Look at the Emergence of the Constitution from the Confederation Period: The Case of the Drafting of the Articles of Confederation, 60 TENN. L. REV. 783 (1994).

Jack Rakove observes:

Nothing in the general reception the Articles received suggests that Americans were deeply interested in discussing the nature of the union they were forming. No pamphlets were written about them, and when the Articles were printed in American newspapers they appeared only as another scrap of news, probably less important than reports of victory at Saratoga and almost certainly less controversial than a growing number of essays proposing remedies for inflation.

RAKOVE, supra note 102, at 185.

268. It is also worth noting that the Articles did not look like typical treaties entered into between sovereign nations. They were drafted by the Continental Congress and sent to the states for ratification. Thus, no state created any separate agreements with any of the other states. The process was centrally directed and states had the option of approval or disapproval, but they could not adjust the Articles’ terms.

269. See generally RAKOVE, supra note 102.

270. See, e.g., id. at 184–85 (“[T]he idea that the confederation was essentially only a league of sovereign states was ultimately a fiction[,] Congress was in fact a national government, burdened with legislative and administrative responsibilities unprecedented in the colonial past.”).

The only substantive difference that ratification of the Articles brought to Congress was that its directives were now styled “ordinances,” rather than “resolves” or “recommendations.” See Richard P. McCormick, Ambiguous Authority: The Ordinances of the Confederation Congress, 1781–1789, 41 AM. J. LEGAL HIST. 411, 411 (1997).
These powers are best seen as powers vested in the Congress by the American people at the time of the Declaration. By creating the new state constitutions, the people of each state vested control of state and local affairs in the state governments. In no state did the people explicitly confer any power over foreign affairs to their state government. It was understood that these powers rested with the Congress.\textsuperscript{271} Thus, the Articles of Confederation did not operate on a clean slate. Under the Articles each state retained “every power, jurisdiction, and right, which is not by this Confederation expressly delegated to the United States, in Congress assembled.” Strictly speaking, this does not mean that Congress had only those powers delegated to it by the states; it means that states retained the powers which they had not delegated to the Congress. \textit{And states could not retain that which was not theirs to begin with.}

Likewise, although each state “retains its sovereignty, freedom, and independence,” this clause reveals nothing about the scope of what was retained. In light of the extensive powers Congress exercised at the time of the ratification in 1781, the best reading of this clause is simply that the states retained their sovereignty over their internal affairs, their freedom to govern themselves as they saw fit, and their independence from Britain, and were not absorbed into one American super-state\textsuperscript{272}

This is not to say that a state might not gain the powers of external sovereignty for itself. If the people of the state agreed to withdraw affirmatively from the union, at least prior to the Constitution, such secession would probably have transferred those powers to the state. Just as the powers of external sovereignty fell to the United States when it withdrew from the British empire, so would those powers fall to a state that withdrew from the American Union. But this would be a result of the people’s ultimate power to distribute sovereignty among organs of

\textsuperscript{271} See, e.g., RAKOVE, supra note 102, at 150 (“Only in Connecticut—whose existing charter, widely regarded as a model for republican government, required no revision—was a serious question raised about the status of Congress when a Litchfield County convention proposed that members of Congress be elected by the people.”).

\textsuperscript{272} Cf. id. at 170–71 (suggesting that the clause was “ambiguous” and that the Congress viewed it as “less significant” than its sponsor believed it to be).

Supporters of a thirteen nations argument might also point to the Articles’ assertion that it formed a “firm League of Friendship.” ARTICLES OF CONFEDERATION art. III. Yet the Articles also describe themselves as “Articles of Confederation and Perpetual Union.” ARTICLES OF CONFEDERATION pmbl. Perpetual union is a quite different matter than a “firm League of Friendship.”
government; it would not be a function of a state’s simply reclaiming a power it had granted away under the Articles.273

Finally, it is worth noting that contemporaries were familiar with the argument that the Articles implied the existence of thirteen separate nations. This argument was nonetheless decisively rejected by many people who thought carefully about these issues, among them such distinguished figures as James Wilson, William Paterson, and Joseph Story. None of these people saw the Articles as contradicting the idea that the Declaration was the act of one people creating an American nation. Neither should we.

7. The Historical Evidence: A Summary

Sections A and B argued the Declaration is best interpreted as the act of one American people creating an American nation. The evidence outlined in this section confirms the merits of such a reading. Such an interpretation was endorsed by such leading figures as James Madison, James Wilson, Alexander Hamilton, John Jay, John Marshall, and John Quincy Adams. It was endorsed in the Constitutional Convention; it was endorsed in the courts of law; and it was endorsed by ordinary Americans on Independence Day. In all these ways, Americans reaffirmed their understanding of the origins of the American nation and of the underlying sovereignty of the American people.

III. THE DECLARATION OF INDEPENDENCE AND CONSTITUTIONAL FORMALISM

We hold these Truths to be self-evident, that all Men are created equal, that they are endowed by their Creator with certain unalienable Rights, that among these are Life, Liberty, and the Pursuit of Happiness—That to secure these Rights, Governments are instituted among Men, deriving their just Powers from the Consent of the Governed, that whenever any Form of Government becomes destructive of these Ends, it is the Right of the People to

273. See also RAKOVE, supra note 102, at 173–74 n. *:
A reconstruction of the precedents, events, and atmosphere of 1774–76 does not validate a states’-rightist interpretation of the origins of the union. In addition to congressional prerogatives over war and diplomacy, the procedures used to authorize the creation of new governments in 1775–76 clearly demonstrate that sovereign powers were vested in Congress from the start and that the emerging provincial regimes were regarded as subordinate bodies.
alter or abolish it, and to institute new Government, laying its Foundation on such Principles, and organizing its Powers in such Form, as to them shall seem most likely to effect their Safety and Happiness.

This eloquent passage is the most celebrated in the Declaration of Independence and among the most celebrated of all political writings in American history. It is a powerful statement of American principles and beliefs. Our understanding of this passage, however, has often been obscured by simplistic readings that strip it of much of its subtlety. At its worst, the passage simply becomes a convenient tool for courts wishing to strike down any legislation that strikes them as violating some sort of “natural right.” This specter of rampant Lochner-ism is a powerful force in legal academics’ traditional aversion to the Declaration. Standing in isolation, this passage has an ungrounded quality and an ethereal feel that makes lawyers doubtful of its analytical power. What constitutes a violation of these “unalienable rights,” and what sort of rights are they? Lacking specificity, the passage could be used to support almost anything.

These concerns can be substantially eliminated by reading the passage in the context of the entire Declaration. About one half of the Declaration is devoted to the specific charges laid against George III. Although these charges are almost entirely ignored today, for the Continental Congress they were the most important part of the Declaration. As Pauline Maier has put it, “Independence was justified only if the charges against the King were convincing and of sufficient gravity to warrant the dissolution of his authority over the American people.”


275. See supra notes 32–41 and accompanying text.

276. MAIER, supra note 3, at 105. Maier notes, “[The charges] were therefore essential to the Declaration’s central purpose, not subordinate to an assumed premise, as Carl Becker argued in one of the more tortured passages in his book on the Declaration.” Id. (citation omitted).
the American Congress, spent only four of its 129 pages discussing the preamble. As Lind noted, “Of the preamble I have taken little or no notice. The truth is, little or none does it deserve.” For Lind, the issue was whether the charges were accurate, a challenge subsequent American defenders of the Declaration would take up. This task was more daunting then it appeared. Pauline Maier has observed, “Today most Americans, including professional historians, would be hard put to identify exactly what prompted many of the accusations Jefferson hurled against the King, which is not surprising since even some well-informed persons of the eighteenth century were perplexed.” However, to understand the philosophy of the Declaration, the historical accuracy of the charges is of little consequence. What is important is that the delegates thought that the actions charged had occurred, and that these charges were sufficient to justify independence.

The charges against the King constitute the principal focus of this Part. It is important to note that the charges do not simply recount instances in which the King had deprived individuals of life, liberty, or the pursuit of happiness. They were addressed to the far more serious concern of a “Design to reduce [the Americans] under absolute Despotism.” It was this fear that justified revolution: “The History of the present King of Great-Britain is a History of repeated Injuries and Usurpations, all having in direct Object the Establishment of an absolute Tyranny over these States. To prove this, let Facts be submitted to a candid World.” The charges seek to establish that George III is a tyrant, which is not necessarily the same thing as saying that he has violated the rights of life, liberty, and pursuit of happiness. Rather, the presumption is that a despotism is so unlikely to safeguard these rights that the people are justified in overthrowing it. Thus the Declaration asserts, “A Prince, whose Character is thus marked by every act which may define a Tyrant, is unfit to be the Ruler of a free People.” The problem is not that George III is a prince, but that he has become a tyrant.

277. See WILLS, supra note 93, at 65.
278. John Lind, An Answer to the Declaration of the American Congress, quoted in id. at 66.
280. MAIER, supra note 3, at 106.
281. The Declaration refers to a “Design.” By 1776, Americans had become convinced that there was a conspiracy in Britain to deprive them of their liberties. See BAILYN, supra note 20, at 144–59.
This emphasis on despotism is linked to the Declaration’s deep concern with what it terms “form.” The Declaration legitimizes revolution “whenever any Form of Government becomes destructive of these Ends.” Notice the prominence of the term “form.” It is not strictly necessary here, as the phrase could easily have read “whenever any government becomes destructive of these ends.” In quick succession, the term appears twice more: The people are to organize new government “in such Form, as to them shall seem most likely to effect their Safety and Happiness.” Then, “Mankind are more disposed to suffer, while Evils are sufferable, than to right themselves by abolishing the Forms to which they are accustomed.” Later, the Declaration charges the King with “altering fundamentally the Forms of our Governments.” Form had a slightly more definite meaning in the eighteenth century than it does today; it implied a clear structure—something with definite shape. This concern with form, of course, is a hallmark of constitutionalism. By analyzing the appropriate forms of government, the Declaration speaks in constitutional language.

The charges against the King, therefore, are not so much rooted in particular violations of natural rights, as they are in what I term “constitutional formalism.” Although the Declaration suggests that “any form of government” might become destructive of the rights of the people, the Declaration is far from neutral as to form. It categorically rules out “Despotism,” “arbitrary Government,” and “absolute Rule,” for example, as acceptable forms of government. More importantly, the Declaration’s charges display a close attention to the proper and distinct spheres of legislative, executive, and judicial power and to the paramount importance of the rule of law. The rights that George III allegedly violated were not so much individual rights, but the collective rights of the American people to self-government.

282. (emphasis added).
283. (emphasis added).
284. (emphasis added).
285. (emphasis added).
287. The Declaration’s grounding in self-government is clearly expressed in a letter written by John Adams on June 23, 1776. Adams felt that with a declaration of independence, it would be easier for the states to “ascertain the criminality of toryism” and to prevent the presses from producing “seditious or traitorous speculations.” Letter from John Adams to John Winthrop, June 23, 1776, reprinted in SPIRIT OF SEVENTY-SIX, supra note 168, at 307. It would lessen “[s]landers upon public men,” and provide the civil governments with a “vigor hitherto unknown.” Id. at 307–
This Part explores these themes through a close reading of the charges against the King. Section A analyzes the Declaration’s commitment to the rule of law. Sections B, C, and D respectively address the legislative, executive, and judicial powers. Section E summarizes the main points of this Part.

A. The Rule of Law

One of the most prominent themes in the Declaration’s charges against the King is the importance of law in a well-governed society. Indeed, the term “laws” serves as a sort of a pervasive refrain throughout the body of the charges:

He has refused his Assent to Laws . . . .
He has forbidden his Governors to pass Laws . . . .
He has refused to pass other Laws . . . .
He has . . . obstruct[ed] the Laws for Naturalization of Foreigners . . . .
He has . . . refus[ed] his Assent to Laws for establishing Judiciary Powers . . . .
He has . . . subject[ed] us to a Jurisdiction . . . unacknowledged by our Laws . . . .
For abolishing the free System of English Laws in a neighbouring Province . . . .
For . . . abolishing our most valuable Laws. 288

Two important principles emerge from this collection of charges. First, laws are not necessary evils, but positive and necessary goods. Second, the law must be defended and obeyed, even by kings.

---

8. Whatever Adams perceived a declaration to mean, it had little to with the rights of minorities to utter expressions of loyalty to Great Britain. Cf. United States ex rel. Guar. Trust Co. of N.Y. v. Gehr, 116 F. 520, 522 (N.D.W.Va. 1902) (“[T]here is not a word in the Declaration of Independence about [freedom of speech].”).

288. (emphasis added).
1. The Importance of Law

The first charge that the Declaration levels against the King is a ringing affirmation of the importance of law: “He has refused his Assent to Laws; the most wholesome and necessary for the public Good.”\(^{289}\) What would it mean for a law to be “necessary?” Garry Wills has argued that the term “necessary” is firmly rooted in the mechanistic world-view of the Enlightenment: “It was the proudest boast of the world opened by Newton’s *Principia* that men could discern necessity at work, invariable, in the flow of apparent chance.”\(^{290}\) This sense of necessity is pervasive in the Declaration, which begins, “WHEN in the Course of human Events, it becomes necessary . . . .”\(^{291}\) Later it refers to the “Necessity which constrains [the colonies] to alter their former Systems of Government.”\(^{292}\) At the end, it simply resigns itself to the inevitable: “We must, therefore, acquiesce in the Necessity, which denounces our Separation.”\(^{293}\) Just as separation from England was required by the course of events leading up to the Declaration, so are certain laws required if a society is to function properly.\(^{294}\)

The term “wholesome” also plays an important role here. A “wholesome” law is one that performs some sort of improving function. The word derives from Middle English terms for “health” and it came to imply a promotion of well-being.\(^{295}\) As Shakespeare’s Marcellus noted of Christmas Eve: “And then, they say, no spirit can walk abroad; / The nights are wholesome; then no planets strike, / No fairy takes, nor witch hath power to charm.”\(^{296}\) A law that is “wholesome” for the “public good,” then, is a sort of medicine for the body politic—something that improves the people and their lives and banishes the evils that would otherwise beset them.

\(^{289}\) (emphasis added).

\(^{290}\) *WILLS, supra note 93, at 94.*

\(^{291}\) (emphasis added).

\(^{292}\) (emphasis added).

\(^{293}\) (emphasis added).

\(^{294}\) The Constitution reiterates the necessity of some laws. Article One, Section Eight of the Constitution authorizes Congress to make all laws “necessary and proper” for carrying out its enumerated powers. Likewise, Article Two, Section Three requires the President to recommend to Congress “such Measures as he shall judge necessary and expedient.”


\(^{296}\) *WILLIAM SHAKESPEARE, HAMLET* act 1, sc. 1 (emphasis added).
The image of law is of an active, positive force in the lives of the people. The claim is not that George III has permitted bad laws, but that he has prevented the passage of good ones, a theme taken up in subsequent charges: The King has “forbidden his Governors to pass Laws of immediate and pressing Importance.” He has “refused to pass other Laws for the Accommodation of Large Districts of People.” He has “endeavoured to prevent the Population of these States; for that Purpose obstructing the Laws for the Naturalization of Foreigners.” He has “obstructed the Administration of Justice, by refusing his Assent to Laws for establishing Judiciary Powers.” It is striking that the charges thus begin, not with a complaint that the King has violated natural rights, nor with a complaint that he has approved bad law, but that there was not enough good law. The lack of good law can evidently be as despotic as the overwhelming presence of bad law. The Declaration has plenty to say about bad law, but for the moment it is important to note how far the Declaration is from an endorsement of the minimalist state.

2. Obedience to Law

An equally compelling principle is that kings are not above the law. In this principle, we can see the deeply legalistic nature of the Declaration. In an intriguing passage, the Declaration asserts, “He has combined with others to subject us to a Jurisdiction foreign to our Constitution, and unacknowledged by our Laws; giving his Assent to their Acts of pretended Legislation.” The “others” in this charge are members of Parliament, a body the Declaration never mentions by name. Carl Becker has suggested that for the Continental Congress it was a “point of principle not on any account to pronounce the word Parliament,” the source of so many of the colonists’ disputes with Britain. What is most striking, however, is the contention that the King and Parliament had together violated “our Constitution” and “our Laws.” It is a bit of a mystery precisely what sort of “Constitution” is being referred to. In an

297. (emphasis added).
298. (emphasis added).
299. The Declaration iterates nine instances of bad laws (“pretended legislation,” as the Declaration puts it) passed by Parliament and approved by the King.
300. At a later point, the Declaration asserts, “We have warned [our British Brethren] from Time to Time of Attempts by their Legislature to extend an unwarrantable Jurisdiction over us” (emphasis added).
301. BECKER, supra note 114, at 19–20.
Declaration of Independence

earlier draft, the phrase read “our constitutions,” but Congress changed this to the singular form.\textsuperscript{302} In the plural, the phrase would almost undoubtedly have referred to the constitutions and charters of the individual colonies. But the singular rules this out. There must be a constitution common to all Americans. It is hard to imagine what this might be other than the British Constitution, which was the subject of much discussion in the revolutionary period.\textsuperscript{303} The British Constitution, of course, was not a written constitution, but an accumulation of unwritten practices that resulted in a balance of power between the king, the lords and the commons. Americans perceived this constitution as binding even the king himself.\textsuperscript{304} Accordingly, any \textit{ultra vires} action by the king would be null and void. The parliamentary legislation he approved was thus “pretended Legislation.” Likewise, the Declaration’s accusation that the King had committed “repeated Injuries and Usurpations” is itself a denunciation of illegal action. The term “injury” in its technical sense refers specifically to a violation of law; this is the essence of the Latin \textit{injuria}, that is, an act contrary to right.\textsuperscript{305} Similarly, the term “usurpation” is meaningless outside of a normative framework for evaluating the legality of exercises of power. The Declaration repeats these terms in its conclusion, asserting, “Our repeated Petitions have been answered only by repeated Injury,” and that the Americans had asked the British people to “disavow these Usurpations.”

The Declaration’s charge that the King and Parliament had asserted a jurisdiction “unacknowledged by our Laws,” closely echoes the charge of “foreign to our Constitution.” This passage suggests the two different senses in which the Declaration employs the word “law.” On the one hand, there are the statutes and common law decisions, ordinary law which the king is obligated to uphold and defend. In that sense, the Declaration denounces the King and Parliament for “abolishing our most valuable Laws” and for “abolishing the free System of English Laws in a neighbouring Province.” On the other hand, there is “law” in a larger sense, the constitutional structures that limit and channel the authority of the king. The king’s obligations to preserve and protect the more

\textsuperscript{302} See id. at 178.

\textsuperscript{303} See, e.g., BAILYN, supra note 20, at 67 (“The word ‘constitution’ and the concept behind it was of central importance to the colonists’ political thought; their entire understanding of the crisis in Anglo-American relations rested upon it.”).

\textsuperscript{304} See, e.g., THOMAS JEFFERSON, A SUMMARY VIEW OF THE RIGHTS OF BRITISH AMERICA (1774), quoted in ELLIS, supra note 274, at 35–36.

\textsuperscript{305} See 7 THE OXFORD ENGLISH DICTIONARY 981 (1989).
ordinary law follow directly from his duties under law in its larger sense. When the king and Parliament abolished ordinary law, their actions were thus illegal, not because they violated the laws that were abolished, but because they violated the “Constitution” and the “Laws.” Likewise, their attempts to legislate for the colonies were an extension of “an unwarrantable Jurisdiction.” In other words, the king and Parliament cannot legitimately act outside of the scope of the “Constitution” and the “Laws.”

But who is to judge whether the king and Parliament have violated the laws? The Declaration makes abundantly clear that this power rests in the people themselves: “when a long Train of Abuses and Usurpations . . . evinces a Design to reduce them under absolute Despotism, it is their Right, it is their Duty, to throw off such Government.” The final judge of the law is in a sense an immense national jury, composed of the people deliberating about whether to throw off their government. The ultimate interpretation of the “Constitution” and “Laws” is therefore not in the hands of a mandarin class, but in the hands of the people. This is not an invitation to lawless violence, but a serious charge to the people as a whole. Revolution is justified only when the people are convinced that their government has persistently and consistently violated the law.

B. The Legislative Power

The Declaration places great value in representative legislatures as protectors of the rights of the people. Many of the charges against the King address obstructions of the legislative process. These obstructions do not directly abridge any individual’s life, liberty, or the pursuit of happiness, but they do interfere dramatically with the people’s right to self-government.

The term “right” appears only twice in the charges, and in both instances it relates to the functioning of representative legislatures. The Declaration asserts that the King “has refused to pass other Laws for the Accommodation of large Districts of People, unless those People would relinquish the Right of Representation in the Legislature, a Right inestimable to them, and formidable to Tyrants only.” In Britain, many districts were not entitled to send representatives to Parliament; this was justified under the theory that they were “virtually” represented by

306. (emphasis added).
representatives from other districts. The Declaration forcefully rejects this notion of virtual representation. Representation is not simply a convenient way to organize a government—it is a fundamental right to which all people are entitled. Accordingly, the charge pointedly links a lack of representation with tyranny.

Why does representation matter? Because the legislature is the primary guarantor of the rights of the people. The second instance of the term “right” in the charges states, “He has dissolved Representative Houses repeatedly, for opposing with manly Firmness his Invasions on the Rights of the People.” Here the Declaration again emphasizes the representative function of legislative bodies by referring to them as “Representative Houses,” rather than as “legislative houses.” The legislature is to protect the people’s rights with “manly Firmness,” a phrase that suggests a certain stern republican virtue. This resonates nicely with an image in the second paragraph: The people are entitled to “provide new Guards for their future Security.” Representative houses filled the role of “guard” perfectly, which was why George III felt compelled to dissolve them. Representative bodies will protect the rights of the people precisely because they are representative of the people.

The importance of representation animates certain other charges the Declaration lays against the King. With Parliament, he is accused of “imposing Taxes on us without our Consent,” and with “suspending our own Legislatures.” In one of the most maligned charges in the Declaration, he is accused of “call[ing] together Legislative Bodies at Places unusual, uncomfortable, and distant from the Depository of their public Records, for the sole Purpose of fatiguing them into Compliance with his Measures.” This charge apparently referred to the moving of the Massachusetts legislature from Boston to Cambridge by the royal governor in 1768. British critics of the Declaration could scarcely contain their derision of this clause. Thomas Hutchinson doubted that Harvard College, four miles from Boston, was unusual, uncomfortable, or distant from the public records, and could not fathom how this “unimportant dispute between an American Governor and his Assembly” could possibly be a “ground to justify Rebellion.”

307. See generally BAILYN, supra note 20, at 161–75.
308. Cf. Gray v. Sanders, 372 U.S. 368, 381 (1963) (“The conception of political equality from the Declaration of Independence, to Lincoln’s Gettysburg Address, to the Fifteenth, Seventeenth, and Nineteenth Amendments can mean only one thing—one person, one vote.”) (citation omitted).
309. See MAIER, supra note 3, at 110.
310. Quoted in id. at 111.
the charge as “truly ridiculous,” and wondered whether it was “inserted by an enemy . . . to throw an air of ridicule on the declaration in general.” The Congress wisely chose not to dwell on the details of this incident, but instead cast it in more general terms. Against the backdrop of representation, the charge takes on more meaning. If a king could “fatigue” legislative bodies into compliance with his own will, then the entire purposes of representation would be destroyed. If the legislature was forced to meet at places distant from the public records, it would be impossible to legislate intelligently about public issues and the people’s right to adequate representation would thereby be diluted.

The Declaration contends that representative bodies are the most appropriate organs to exercise the legislative power. In an interesting charge, the Declaration reveals its understanding of this power:

He has refused for a long Time, after such Dissolutions, to cause others to be elected; whereby the Legislative Powers, incapable of Annihilation, have returned to the People at large for their exercise; the State remaining in the mean time exposed to all the Dangers of Invasion from without, and Convulsions within.

This charge makes clear that the legislative powers originate with the people, to whom those powers return on the dissolution of legislative assemblies. But this is obviously an extremely undesirable outcome. The people at large should not be exercising the legislative power. Only a representative legislature can wield these powers effectively and protect the state from invasions and convulsions. By dissolving legislative houses, the king reduces the colonies to a tumultuous ekklesia at best, and to complete anarchy at worst.

In sum, the Declaration views representative legislatures as the primary protector of the rights of the people. It is critically important that these assemblies be truly representative, that they are protected from interference and dissolution by the executive, and that their powers not return to the people at large. Once again, the deeply formalist nature of the Declaration is apparent. By placing such stress on the importance of representative legislatures, the Declaration suggests that they are not simply one possible form of government among many plausible alternatives. Rather, representation in the legislature is essential to self-government, the most important right the Declaration intends to protect.

311. Quoted in id. at 111.
C. The Executive Power

In one sense, the entire Declaration is an essay on the misuse of executive power, because each of the charges laid against the King describes an abuse of his authority. This Section attempts to bring more precision to the subject of the Declaration’s perspective on executive power. A good place to start is with the conclusion to the charges, where the Declaration asserts, “A Prince, whose Character is thus marked by every act which may define a Tyrant, is unfit to be the Ruler of a free People.” This sentence demonstrates that the relationship between monarchy and tyranny is not as direct as is often made out. Indeed, the sentence implies that a “Prince” who did not engage in tyrannous acts would be perfectly fit to be the “Ruler of a free People.” To modern ears, it seems strange that anyone might be the “Ruler” of a free people; after all, the people should be ruling themselves. But in this context, “Ruler” is simply a synonym for the executive authority. “Head of a free People” might have captured the meaning slightly better, but with less elegance; “Ruler” served just as well. A people might thus be under a “Prince” yet still be “free.” The key issue is whether that “Prince” conducts himself as a constitutional monarch governed by the rule of law, or acts as a lawless tyrant.

Notice how this becomes a question of personal character: “whose Character is thus marked by every act which may define a Tyrant.” It is not simply that George III has done tyrannous things, but that he has a tyrannous character. The distinction is subtle, but important. One who has done tyrannous things might be restrained or reformed; one with a tyrannous character must be overthrown, because he poses the greater threat to the people’s liberties. This is a reflection of the revolutionary generation’s deep concern about

312. Cf. Metro. Wash. Airports Auth. v. Citizens for the Abatement of Aircraft Noise, Inc., 501 U.S. 252, 273 (1991) (“The abuses by the monarch recounted in the Declaration of Independence provide dramatic evidence of the threat to liberty posed by a too powerful executive.”); Youngstown Sheet & Tube Co. v. Sawyer, 343 U.S. 579, 641 (1952) (Jackson, J., concurring) (“The example of such unlimited executive power that must have most impressed the forefathers was the prerogative exercised by George III, and the description of its evils in the Declaration of Independence leads me to doubt that they were creating their new Executive in his image.”).

313. Cf. Lucas, Justifying America, supra note 104, at 110 (“[Independence] was not likely to be won without support from European monarchies such as France and Spain, neither of which would be favorably impressed with a Declaration that inveighed against kingship.”).

314. (emphasis added).
public virtue, and about the essential character attributes of those who would govern.315

Of course, it will not always be easy for the people to determine whether or not their leader displays a tyrannous disposition. This was particularly true in the eighteenth-century, where few people ever met their leader face-to-face (only a handful of Americans had actually met George III). Accordingly, government must be structured to control the potential tyrannous tendencies of those in power. The Declaration thus provides an implicit framework for control of the executive. The main features of this framework are limited veto powers, positive duties of protection, civilian control of the military, and limits to the creation of executive offices.

1. The Veto

Some of the Declaration’s most bitter complaints concern the King’s use of his veto power to prevent the passage of laws favored by the colonial assemblies.316 At least five of the Declaration’s charges relate to this abuse of the veto power. The inescapable conclusion seems to be that the vesting of an absolute veto power in one individual creates a dangerous likelihood of tyranny. This is especially true when the veto power is wielded against representative legislatures, the very bodies the Declaration contemplates as the prime protectors of the people’s liberties. A representative legislature whose enactments can be readily discarded by an executive ceases to be a legislature at all, but becomes merely a forum for the expression of discontent. As Gordon Wood has explained, in 1776 “it seemed abominable that a single person should have a negative over the voice of the whole society.”317

2. Positive Duties of Protection

The principle that the king has duties to his people dates back at least to the medieval coronation oaths and is echoed in the oath that Article Two of the Constitution provides for the President of the United States: “I do solemnly swear (or affirm) that I will faithfully execute the Office

315. See, e.g., WOOD, supra note 20, at 65–70.
316. Technically speaking, the King’s disallowance of colonial acts, a power exercised through the Privy Council, was different from his veto power over acts of Parliament. See DUMBAULD, supra note 109, at 87. From the colonies’ point of view, such a distinction was of little significance.
317. WOOD, supra note 20, at 141.
of President of the United States, and will to the best of my Ability, preserve, protect and defend the Constitution of the United States.”

That George III owed positive legal duties to his people is made clear in the Declaration: He has “utterly neglected to attend” to the passage of important laws, and has “abdicated Government here, by declaring us out of his Protection and waging War against us.” This lack of action by the King constitutes a violation of his affirmative duties as an executive. Accordingly, he can be seen to have “abdicated” the office of king, in essence leaving that office vacant for the people to fill as they deem fit.

These charges suggest that some sort of covenant or contract exists between an executive and the people. The people entrust the executive with the executive power; in exchange, the executive must exercise those powers for the benefit of the people. As Thomas Jefferson put it in his 1774 pamphlet, *A Summary View of the Rights of British America*, the king is “the chief officer of the people, appointed by the laws, and circumscribed with definite powers, to assist in working the great machine of government erected for their use, and consequently subject to their superintendence.” It is not merely that the people have the power to select an executive when that office is indisputably vacant. They have the far more important power of determining for themselves the conditions under which executive misfeasance or nonfeasance will constitute an abdication of that office. Thus, in the Declaration it is the “good People of these Colonies” who determine that George III has “abdicated Government here.”

3. **Civilian Control of the Military**

Since the threat of an overpowering military was never far from the minds of the revolutionary generation, the Declaration also has much to say about the relationship between the military and the executive authority. The Declaration condemns the King for “affect[ing] to render the Military independent of and superior to the Civil Power.” In similar fashion, the Declaration asserts that the King “has kept among us, in Times of Peace, Standing Armies, without the consent of our Legislatures,” and that the King and Parliament have “quarter[ed] large Bodies of Armed Troops among us.” Indeed, these charges suggest that

---

318. U.S. CONST. art II, § 1, cl. 8.
319. See supra note 304, at 35-36.
it was not merely a despotism that George III was supposedly trying to impose on the American states; it was the worst sort of despotism—a military despotism. The fundamental problem was making the military both “independent” of and “superior” to the civilian authorities. Military independence freed the military of civilian control; military superiority in turn subjected the civilian authority to military control. To the revolutionary generation, a standing army represented one of the greatest threats to republican virtue and liberty. The Declaration takes pains to point out that the Hessian troops hired by George III were not just mercenaries, but an Army: “He is, at this time, transporting large Armies of foreign Mercenaries to compleat the Works of Death, Desolation, and Tyranny, already begun with circumstances of Cruelty and Perfidy, scarcely paralleled in the most barbarous Ages, and totally unworthy the Head of a civilized Nation.” It is this image that the Declaration contemplates as the inevitable result of an uncontrolled military. Military dominance is a trait of tyrannous and “barbarous” nations; by contrast, “civilized” nations can rigorously control the use of their military power.

What is the fundamental problem with military control? After all, in some circumstances, military control might be preferable to ineffective civilian government. The fundamental problem, in the Declaration’s view, is that military rule is fundamentally inconsistent with representative government. Military government is wrong, not so much because it is likely to infringe on the people’s individual liberties, which it almost assuredly will, but because it infringes on the people’s “Right of Representation in the Legislature, a Right inestimable to them, and formidable to Tyrants only.” Although the people might be represented in the local militia, they cannot be represented in armies, which are rigidly hierarchical and governed by their own distinct rules

321. See id. at 53.
322. Consider, for example, an address given by General John Cadwalader, a leading figure in the American Revolution, to the Pennsylvania Council of Safety in 1777:

We wish to see the civil authority regulate and direct all our public measures and should greatly lament the necessity, which may compel the military power to take the direction in their hands, in order to save this country from absolute ruin—but you may depend that the military will exert its authority whenever the weakness, languor, or timidity of your councils shall render it their duty so to do, and all the world will justify them in it.

General John Cadwalader, Address to the Pennsylvania Council of Safety (Jan. 15, 1777) (on file with the Cadwalader Collection, Historical Society of Pennsylvania).
323. (emphasis added).
and practices. It is the legislatures, not armies, who speak for the people; consequently the slightest intrusions on representative legislatures should be fervently resisted.

4. *The Proliferation of Offices*

Another abuse of executive authority is the creation of numerous and unnecessary offices. The Declaration expresses this in colorful terms: “He has erected a Multitude of new Offices, and sent hither Swarms of Officers to harrass our People, and eat out their Substance.” There is more than a touch of the hyperbolic here, but the charge is nonetheless interesting. New offices will of course be filled by new officers, and the Declaration’s use of the term “Swarms” readily equates these officers with a biblical plague of locusts.324 Part of the problem, no doubt, is that these officers would not be Americans, but faceless bureaucrats “sent hither” from Britain. Again, the issue is a lack of accountability in government and a disconnect between the people and their rulers.

5. *The Executive Power: A Summary*

In all of these ways, then, the Declaration suggests how the power of the executive can be controlled. The Declaration does not necessarily reject Alexander Hamilton’s contention in *The Federalist* that “[e]nergy in the executive is a leading character in the definition of good government,”325 but it does emphasize the importance of channeling that energy within narrow bounds. The Declaration is much more disturbed by executive encroachment on the rights of the representative colonial assemblies than it is about potential legislative encroachments on the prerogatives of the Crown. As a means of preventing tyranny, this makes perfect sense. In the minds of the drafters of the Declaration, executive tyranny was a far more serious threat than legislative tyranny. To be sure, the Declaration minces no words in denouncing the actions of Parliament. The problem, though, was not that Parliament had encroached on the prerogatives of the King; indeed, the Declaration heartily denounces Parliament for having “combined,” that is, “conspired,” with the King to form an absolute tyranny over the American colonies. The problem was that Parliament was simply not the

324. See MAIER, supra note 3, at 110.
relevant legislature. The legislatures that mattered were the colonial assemblies, and it was their rights that the King and Parliament were infringing.

Once again, the Declaration is speaking not so much about individual rights, but about structure. Properly constructed and faithfully executed, the executive power can play an important role in preserving the fundamental rights of self-government. Poorly structured and imperfectly controlled, the executive power can quickly degenerate into tyranny.

D. The Judiciary Power

The Declaration’s discussion of the judiciary power is also grounded in the preservation of self-government. The first charge to deal with the judiciary asserts that the king had “obstructed the Administration of Justice, by refusing his Assent to Laws for establishing Judiciary Powers.” The Declaration here specifically links the judiciary powers to the administration of justice. But it is only a particular kind of judiciary that is likely to administer anything resembling justice. The Declaration thus carefully lays out its understanding of the proper role of judges and of juries.

I. Judges

The Declaration addresses one charge to the role of judges. It contends that the King “has made Judges dependent on his Will alone, for the Tenure of their Offices, and the Amount and Payment of their Salaries.” American judges have often cited this charge in support of judicial independence,326 and, indeed, the charge does suggest the importance of an independent judiciary. Note the striking contrast, for example, between this charge and the military charge. The King is denounced for making the military “independent,” but he is equally denounced for making judges “dependent.” One might read the Declaration of Independence, then, as a declaration of dependence for the American military, but as a declaration of independence for the American judiciary.327

326. See supra note 82 and accompanying text.

327. Indeed, one of the earliest and most forceful assertions of judicial independence in the revolutionary era was directed not against the civilian government, but against the Continental
There is much to be said for such a reading, but it is not quite accurate. The issue was not so much that judges were dependent, but that they were dependent on the wrong entity. Judges were supposed to be accountable to the colonial assemblies, not to the king. A judge appointed and paid by the king simply became one more petty executive official, one more locust in the great “Swarm of Officers,” to use the phrase that immediately follows this charge. Such a person would have no sense of the local community and no duties to the people at large. Far better that judges be appointed by representative colonial assemblies or by the people. Only if judges are accountable in this fashion, can self-government be preserved. This does not necessarily mean that life tenure or salary protection for judges would be inconsistent with self-government, but it does mean that representative legislatures must have a role in their selection and that there be some mechanism for removal in cases of criminality or mental incapacity.

2. Juries

Although the Declaration discusses judges in only one charge, it addresses juries in no fewer than three charges. This imbalance suggests the vital role that the drafters of the Declaration envisioned for the jury. Quite simply, the jury was the democratic branch of the judicial department. Through juries, the people themselves would determine when to impose the coercive power of the state on an individual.

The Declaration asserts that the King and Parliament had “depriv[ed] us, in many Cases, of the Benefits of Trial by Jury.” The use of the term

Army. In April 1778, Pennsylvania Chief Justice Thomas McKean, a signer of the Declaration of Independence, ordered Robert Hooper, a colonel in the Continental Army, to appear before him to answer for a violent assault Hooper had committed on Jonathan Dickinson Sergeant, the Pennsylvania Attorney General, at the Reading Court of Quarter Sessions. Letter from Thomas McKean to Nathaniel Greene (June 9, 1778), quoted in John M. Coleman, Thomas McKean: Forgotten Leader of the Revolution 225 (1975). Hooper’s commanding officer, General Nathaniel Greene, wrote to McKean, claiming that he could not “without great Necessity consent” to Hooper’s absence. McKean replied:

I shall not ask your consent nor that of any other person in or out of the army, whether my Precept shall be obeyed or not in Pennsylvania. . . . I should be very sorry to find, that the execution of criminal laws should impede the operation of the army in any instance, but should be more so to find the latter impede the former.”

Id.

328. See, e.g., MAIER, supra note 3, at 111.
329. The best exposition of this idea is in AMAR, supra note 320, at 81–118.
“Benefits” nicely captures the importance of the jury. Jury trial was not simply a “right” of the accused; it had “benefits” for the entire community. Tocqueville, one of the most astute observers of the American jury, would later observe, “The institution of the jury . . . places the real direction of society in the hands of the governed . . . and not in that of the government.” It was this element of popular participation that drove the revolutionary generation’s concern for the preservation of the jury. Popular participation on juries lies beneath two other charges that the Declaration lays against the King and Parliament. Immediately after the jury charge, the Declaration denounces “transporting us beyond Seas to be tried for pretended Offences.” This charge echoes the Declaration’s early dismissal of Parliamentary acts as “pretended Legislation.” At issue was the application of a statute from the reign of Henry VIII allowing the trial in England of treasons committed outside the realm. The fundamental problem with this, of course, was that it destroyed the privilege of trial by a local jury. Not only would the individual be deprived of jurors drawn from his community, but the local community would also be deprived of its right to decide whether to apply the treason statute in a given case. In similar fashion, the Declaration denounces the protection of British troops by “a mock Trial, from Punishment for any Murders which they should commit on the Inhabitants of these States.” A “mock trial” is one in which local civilian juries do not have a meaningful voice.

E. The Constitutional Formalism of the Declaration

This Part has argued that the Declaration displays a pervasive concern with the structure of government and with the rule of law. In this manner, the Declaration is deeply formalistic and speaks in constitutional language. This is powerful evidence that the “rights”...
about which the Declaration’s famous second paragraph is most concerned are not so much the natural rights of individuals against the state or against the majority, but the rights of the American people to self-government. The most important phrase in the second paragraph is not so much “Life, Liberty, and the Pursuit of Happiness,” but rather “the Consent of the Governed.” Charge after charge reiterates the ways in which the King and Parliament have interfered with American self-government. To be sure, the Declaration describes some direct assaults on life, liberty, and the pursuit of happiness. The charge, “He has plundered our Seas, ravaged our Coasts, burnt our Towns, and destroyed the Lives of our People,” is a good example of this. But even this refers to an assault on the American people as a whole; it describes violent acts that directly interfere with the rights of self-government.

This deep concern for the preservation of self-government is reflected in the Declaration’s concern with the structure and the form of government. Far from being neutral as to form, the Declaration specifies the forms that are most likely to preserve self-government and, by extension, the inalienable rights of life, liberty, and the pursuit of happiness. The Declaration imagines a government structured along these lines as most conducive to securing the people’s “Safety and Happiness”: the government must be founded in the rule of law and must value the preservation of law as a valuable end in its own right. It must recognize that no official is above the law and must provide mechanisms for dealing with official lawlessness. It must have representative legislatures responsive to the people, but it must not allow the legislative powers to return to the people at large. It must limit the executive’s power of the veto, specify the executive’s legal duties, and limit the proliferation of executive offices. It must ensure that the military power remains subordinate to the civilian authorities. It must protect judges from interference by the executive power, and it must preserve inviolate the role of juries in the courts. This basic outline of government is not inconsistent with the state constitutions that were drafted after the Declaration nor is it inconsistent with the Constitution framed by the Convention of 1787. Although they differ in some details, all are broadly concerned with protecting the people’s right to self-government and with protecting the people from self-interested government.

In his 1998 book, *The Bill of Rights*, Akhil Amar argues that the Bill of Rights was not primarily about vesting “individuals and
minorities with substantive rights against popular majorities”; rather, it was about “structural ideas” intended “not to impede popular majorities but to empower them.” This Part has offered a similar analysis of the Declaration of Independence. For too long, the Declaration has been seen as a rather vacuous endorsement of indefinable and ghostly natural rights. When we strip away the interpretive overlay of the *Lochner* era, we can see the Declaration that its drafters saw: a Declaration with very specific notions about the structure of government, and a Declaration that, above all else, is committed to the right of the American people to govern themselves. On this, Jefferson’s last letter serves well as the last word: “May [the Declaration of Independence] be to the world, what I believe it will be . . . the signal of arousing men to burst the chains under which monkish ignorance and superstition had persuaded them to bind themselves, and to assume the blessings and security of self-government.”

IV. CONCLUSION

On July 3, 1776, John Adams wrote to his wife Abigail:

The second day of July, 1776, will be the most memorable epocha in the history of America. I am apt to believe that it will be celebrated by succeeding generations as the great anniversary festival. It ought to be commemorated as the day of deliverance, by solemn acts of devotion to God Almighty. It ought to be solemnized with pomp and parade, with shows, games, sports, guns, bells, bonfires and illuminations, from one end of this continent to the other, from this time forward, forevermore.

You will think me transported with enthusiasm, but I am not. I am well aware of the toil, and blood, and treasure, that it will cost us to maintain this declaration, and support and defend these States. Yet, through all the gloom, I can see the rays of ravishing light and glory. I can see that the end is more than worth all the means, and

335. *Id.* at xii.
336. Letter from Thomas Jefferson to Roger Weightman (June 24, 1826), in JEFFERSON’S LETTERS, supra note 224, at 666 (emphasis added).
that posterity will triumph in that day’s transaction, even although we should rue it, which I trust in God we shall not.\footnote{337}

Adams’s predictions were remarkably prescient. He only got one thing wrong—the date. Americans do not (and never did) celebrate the day that Congress voted to approve Richard Henry Lee’s motion on independence. We celebrate the day that Congress approved the Declaration of Independence. Why? Because even more important than independence itself was how that independence was declared and justified. We celebrate the Declaration because it marks the birth of an American nation, because it dedicates America to the principle of self-government, and because it confirms the inherent sovereignty of the American people. The vote on July 2 did not do this, but the vote on July 4 did.\footnote{338}

For too long, these important features of the Declaration have been obscured in the legal community by a misguided focus on individual natural rights. As a result, it has been easy to consign the Declaration to irrelevance and oblivion. This Article has argued that the deepest principles of the Declaration are not about an individual’s natural rights against the state, but about the right of the American people to self-government and about the formal structures that allow self-government to flourish. When we understand the Declaration in this fashion, we can better appreciate the Declaration’s place in American law.

\footnote{337. Letter from John Adams to Abigail Adams, July 3, 1776, \textit{reprinted in Spirit of Seventy-Six}, supra note 168, at 321. Late in life, Adams, envious of Thomas Jefferson’s fame as the principal draftsman of the Declaration and fearful that his own contributions were being eclipsed, dismissed the Declaration as a “theatrical side show,” and complained that “Jefferson ran away with the stage effect . . . and all the glory of it.” \textit{Quoted in Ellis, supra note 274, at 292.}}

\footnote{338. Even as a purely legal matter, the Declaration, and not the July 2 vote, created American independence. The voicing of the Declaration suggests that it is doing something in the present moment. Thus, it states, “we declare,” not we “have declared” or we “have resolved.” “Declaring,” of course, had a distinctly legal connotation. A declaratory judgment, familiar from equity, declared the rights of individual parties in the same way that the Declaration declared the rights of the American people. \textit{Cf.} Peter Charles Hoffer, \textit{The Law’s Conscience: Equitable Constitutionalism in America} 72–77 (1990) (tying the Declaration to forms of equitable pleading). A declaration of war altered the legal status between two countries in the same way that the Declaration altered the relationship between the colonies and Great Britain. (The Constitution grants Congress the power to “\textit{declare} War.” U.S. \textit{Const.} art I, § 8 (emphasis added).) The resolution of July 2 might easily have been the subject of a motion to reconsider; once independence was “declared,” such a motion could not have succeeded. Finally, all legal decisions, from the eighteenth century forward, have treated the Declaration, not the July 2 vote, as the legal act of independence.}
First, reading the Declaration as the act of one American people creating an American nation makes sense of a number of judicial decisions that cannot be justified if the Declaration created thirteen independent nations. These include cases dating United States citizenship to the Declaration of Independence, cases justifying United States control over the shoreline, and cases concerning the foreign affairs powers of the federal government.

Second, the nationalist understanding of the Declaration is relevant to one of the most divisive issues in modern constitutional law, the so-called “sovereign immunity” of the states. A paramount principle of the Declaration is that the people, not governments, are sovereign. Any “sovereignty” the state governments possess is theirs because the sovereign American people gave it to them. Quite simply, the states have never been completely sovereign and independent nations. Invoking their “sovereignty” to avoid suit in federal court for violations of federal law fundamentally misunderstands the nature of the federal union. A Declaration that denounces “mock trials,” celebrates the role of juries and the rule of law, attacks judicial subservience to the executive, and condemns a “swarm of officers” for “harrass[ing] our people” offers little support for a doctrine that keeps citizens out of court when government lawlessness infringes on their rights.

Third, recognizing the grounding of the Declaration in popular sovereignty rescues it from the Lochner-esque reasoning with which it is

---

339. See cases cited supra note 64.
340. See, e.g., United States v. Maine, 420 U.S. 515, 520 (1975) (“[D]ominion over the marginal sea was first accomplished by the National Government rather than by the Colonies or by the States.”); United States v. California, 332 U.S. 19, 31 (1947) (“[W]e cannot say that the thirteen original colonies separately acquired ownership to the three-mile belt or the soil under it.”).
342. The classic statement of popular sovereignty as a rejection of governmental sovereignty is Amar, supra note 101.
343. Some judges have recognized that the Declaration may be relevant to sovereign immunity. See, e.g., Seminole Tribe v. Florida, 517 U.S. 44, 95 (1996) (Stevens, J., dissenting) (“[T]he recitation in the Declaration of Independence of the wrongs committed by George III made [the proposition that the king could do no wrong] unacceptable on this side of the Atlantic.”); Nevada v. Hall, 440 U.S. 410, 415 (1979) (“[T]he fiction that the king could do no wrong] was rejected by the colonists when they declared their independence from the Crown.”); City and County of Denver v. Madison, 351 P.2d 826, 831 (Colo. 1960) (Frantz, J., dissenting) (arguing that the Declaration repudiates governmental immunity from tort suits), Bulman v. Hulstrand Constr. Co., 521 N.W.2d 632, 641 (N.D. 1994) (Sandstrom, J., concurring) (noting inconsistency between the Declaration and sovereign immunity); Fowler v. City of Cleveland, 126 N.E. 72, 78 (Ohio 1919) (Wanamaker, J., concurring) (“The Declaration of Independence makes no exception in favor of governmental sovereignty.”).
often associated. Anyone who argues that a particular law violates the Declaration of Independence because it infringes on some perceived individual natural right misunderstands the Declaration. There is not one line in the Declaration about the empowerment of minorities against popular majorities. The core principle of the Declaration is the people’s right of self-government; it is this principle that pervades all of the charges against the King. The mistake of the *Lochner* era was to read into the Declaration’s second paragraph a vision of economic laissez-faire that is unsupported by anything else in the Declaration’s text. Courts attentive to the entire text of the Declaration would have recognized that individual rights are not the only rights at stake when a law is challenged. Equally at stake is the people’s right to self-government through representative legislatures, a right far more important to the Declaration than the rights, for example, of butchers and bakers to carry on their livelihood in whatever fashion they saw fit.

Fourth, a reconsideration of the Declaration helps define with more precision when it is appropriate for courts to cite the Declaration. This is particularly relevant in those states that appear to have adopted the “principles of the Declaration of Independence” as part of their positive law. As should now be clear, it will rarely be sufficient to cite the Declaration in opposition to some law that supposedly violates individual rights, because there will almost always be an opposing interest of self-government. It is in cases involving the structural principles of government that the Declaration can play a more valuable role. The most suggestive cases along these lines are those that employ the Declaration as a “negative constitution.” These cases have intuitively understood that the Declaration can speak to structure, but none of them quite realized the degree to which structural concerns animate the entire Declaration. This Article has attempted to move the analysis forward by demonstrating how the charges form a coherent structural vision—a vision that may be persuasive in resolving structural

344. By this, I do not mean to suggest that the protection of minorities is not a critical thread in the fabric of American law. It is. But this is much more a result of the Reconstruction Amendments than of the political ideas that animate the Declaration of Independence. See generally AMAR, supra note 320.

345. See supra notes 73–74 and accompanying text.

346. It is possible that some laws might be so oppressive that they create a *prima facie* case that the process of representation has utterly broken down. In such a case, violation of individual rights may be indicative of the violation of the collective right of the American people to self-government.

347. See supra notes 76–90 and accompanying text.
questions in the interpretation of the state constitutions and the Constitution of the United States.

Fifth, focusing on the Declaration of Independence brings to light the fascinating pre-constitutional history of the United States. In the legal community, it often seems as if nothing of any significance happened before 1787. This Article has demonstrated the fallacy of that belief. By carefully studying the period between 1776 and 1787, we can gain a much better understanding of the origins of our country and of the principles upon which it was founded.

* * *

The parchment copy of the Declaration of Independence now rests in a massive bulletproof container in the National Archives Building in Washington, D.C., where it is viewed by thousands of visitors every year. Immediately below the Declaration, one can view two pages of the parchment copy of the Constitution and the parchment copy of the twelve amendments to the Constitution proposed by the First Congress, ten of which became our Bill of Rights. The writing on the Constitution and the Bill of Rights is still bright and crisp, but the Declaration has worn badly. Years of exposure to the sun and a poorly executed attempt to create a facsimile in 1823 have rendered it almost entirely illegible. The placement of the documents in the Archives accentuates the elusiveness of the Declaration. Whereas the Constitution and Bill of Rights are conveniently placed at eye level, the Declaration is higher and further back; the visitor to the Archives must strain mightily to lean over the Constitution to make out any words in the Declaration.

This physical arrangement of our nation’s founding documents is perfectly symbolic. The familiar words of the Constitution and the Bill of Rights have been parsed and re-parsed by lawyers in thousands of cases over the past two centuries. But the Declaration is always just beyond the horizon, in the nether twilight where familiarity blends into shadows. It is the vague ethereal presence that somehow looms so deeply behind the Constitution that even legally-trained Presidents of the United States can easily confuse the two documents.

This Article has attempted to bring the Declaration into sharper focus, bringing clarity to those lines that are so elusive in the National Archives, and, perhaps, bridging the gap between the Declaration and

348. See MAIER, supra note 3, at xi.
349. See supra note 23 and accompanying text.
the documents that lie beneath it. Both Abraham Lincoln and Roger Brooks Taney would have understood why the Declaration holds such a pride of place in the Archives. These men recognized, in their own way, the important role the Declaration plays in American law. For Lincoln, the Declaration was the founding document of the American nation, a document committed to “government of the people, by the people, for the people.” For Taney, the Declaration was a source of deep structural principles of government. This Article has argued that on these points, both men were right. Right not only because of their understanding of the Declaration, but because they bothered to consider it at all. For both men correctly understood the Declaration not as a curiosity of interest only to antiquarians, but as a living source of principles of government—principles that 225 years later continue to define and shape this great experiment in self-government that we call the United States of America.

350. Lincoln, supra note 183, at 104.
Appendix: The Dunlap Broadside, the First Printing of the Declaration of Independence

IN CONGRESS, JULY 4, 1776.

A DECLARATION
BY THE REPRESENTATIVES OF THE UNITED STATES OF AMERICA,
IN GENERAL CONGRESS ASSEMBLED.

WHEN in the Course of human Events, it becomes necessary for one People to dissolve the Political Bands which have connected them with another, and to assume among the Powers of the Earth, the separate and equal Station to which the Laws of Nature and of Nature’s God entitle them, a decent Respect to the Opinions of Mankind requires that they should declare the causes which impel them to the Separation.

WE hold these Truths to be self-evident, that all Men are created equal, that they are endowed by their Creator with certain unalienable Rights, that among these are Life, Liberty, and the Pursuit of Happiness—That to secure these Rights, Governments are instituted among Men, deriving their just Powers from the Consent of the Governed, that whenever any Form of Government becomes destructive of these Ends, it is the Right of the People to alter or abolish it, and to institute new Government, laying its Foundation on such Principles, and organizing its Powers in such Form, as to them shall seem most likely to effect their Safety and Happiness. Prudence, indeed, will dictate that Governments long established should not be changed for light and transient Causes, and all Experience hath shewn, that Mankind are more disposed to suffer, while Evils are sufferable, than to right themselves by abolishing the Forms to which they are accustomed. But when a long Train of Abuses and Usurpations, pursuing invariably the same Object, evinces a Design to reduce them under absolute Despotism, it is their Right, it is their Duty, to throw off such Government, and to provide new Guards for their future Security. Such has been the patient Sufferance of these Colonies; and such is now the Necessity which constrains them to alter their former Systems of Government. The History of the present
King of Great-Britain is a History of repeated Injuries and Usurpations, all having in direct Object the Establishment of an absolute Tyranny over these States. To prove this, let Facts be submitted to a candid World.

HE has refused his Assent to Laws, the most wholesome and necessary for the public Good.

HE has forbidden his Governors to pass Laws of immediate and pressing Importance, unless suspended in their Operation till his Assent should be obtained; and when so suspended, he has utterly neglected to attend to them.

HE has refused to pass other Laws for the Accommodation of large Districts of People, unless those People would relinquish the Right of Representation in the Legislature, a Right inestimable to them, and formidable to Tyrants only.

HE has called together Legislative Bodies at Places unusual, uncomfortable, and distant from the Depository of their public Records, for the sole Purpose of fatiguing them into Compliance with his Measures.

HE has dissolved Representative Houses repeatedly, for opposing with manly firmness his Invasions on the Rights of the People.

HE has refused for a long Time, after such Dissolutions, to cause others to be elected; whereby the Legislative Powers, incapable of Annihilation, have returned to the People at large for their exercise; the State remaining in the mean time exposed to all the Dangers of Invasion from without, and Convulsions within.

HE has endeavoured to prevent the Population of these States; for that purpose obstructing the Laws for Naturalization of Foreigners; refusing to pass others to encourage their Migrations hither, and raising the Conditions of new Appropriations of Lands.

HE has obstructed the Administration of Justice, by refusing his Assent to Laws for establishing Judiciary Powers.

HE has made Judges dependent on his Will alone, for the Tenure of their Offices, and the Amount and Payment of their Salaries.

HE has erected a Multitude of new Offices, and sent hither Swarms of Officers to harrass our People, and eat out their Substance.

HE has kept among us, in Times of Peace, Standing Armies, without the consent of our Legislatues.

HE has affected to render the Military independent of and superior to the Civil Power.
HE has combined with others to subject us to a Jurisdiction foreign to our Constitution, and unacknowledged by our Laws; giving his Assent to their Acts of pretended Legislation.

FOR quartering large Bodies of Armed Troops among us:

FOR protecting them, by a mock Trial, from Punishment for any Murders which they should commit on the Inhabitants of these States:

FOR cutting off our Trade with all Parts of the World:

FOR imposing Taxes on us without our Consent:

FOR depriving us, in many Cases, of the Benefits of Trial by Jury:

FOR transporting us beyond Seas to be tried for pretended Offences:

FOR abolishing the free System of English Laws in a neighbouring Province, establishing therein an arbitrary Government, and enlarging its Boundaries, so as to render it at once an Example and fit Instrument for introducing the same absolute Rule into these Colonies:

FOR taking away our Charters, abolishing our most valuable Laws and altering fundamentally the Forms of our Governments:

FOR suspending our own Legislatures, and declaring themselves invested with Power to legislate for us in all Cases whatsoever.

HE has abdicated Government here, by declaring us out of his Protection and waging War against us.

HE has plundered our Seas, ravaged our Coasts, burnt our Towns, and destroyed the Lives of our People.

HE is, at this Time, transporting large Armies of foreign Mercenaries to compleat the Works of Death, Desolation, and Tyranny, already begun with circumstances of Cruelty and Perfidy, scarcely paralleled in the most barbarous Ages, and totally unworthy the Head of a civilized Nation.

HE has constrained our fellow Citizens taken Captive on the high Seas to bear Arms against their Country, to become the Executioners of their Friends and Brethren, or to fall themselves by their Hands.

HE has excited domestic Insurrections amongst us, and has endeavoured to bring on the Inhabitants of our Frontiers, the merciless Indian Savages, whose known Rule of Warfare, is an undistinguished Destruction, of all Ages, Sexes and Conditions.

IN every state of these Oppressions we have Petitioned for Redress in the most humble Terms: Our repeated Petitions have been answered only by repeated Injury. A Prince, whose Character is thus marked by every act which may define a Tyrant, is unfit to be the Ruler of a free People.
NOR have we been wanting in Attentions to our British Brethren. We have warned them from Time to Time of Attempts by their Legislature to extend an unwarrantable Jurisdiction over us. We have reminded them of the Circumstances of our Emigration and Settlement here. We have appealed to their native Justice and Magnanimity, and we have conjured them by the Ties of our common Kindred to disavow these Usurpations, which, would inevitably interrupt our Connections and Correspondence. They too have been deaf to the Voice of Justice and of Consanguinity. We must, therefore, acquiesce in the Necessity, which denounces our Separation, and hold them, as we hold the rest of Mankind, Enemies in War, in Peace, Friends.

WE, therefore, the Representatives of the UNITED STATES OF AMERICA, in GENERAL CONGRESS, Assembled, appealing to the Supreme Judge of the World for the Rectitude of our Intentions, do, in the Name, and by the Authority of the good People of these Colonies, solemnly Publish and Declare, That these United Colonies are, and of right ought to be, FREE AND INDEPENDENT STATES; that they are absolved from all Allegiance to the British Crown, and that all political Connection between them and the State of Great-Britain, is and ought to be totally dissolved; and that as FREE AND INDEPENDENT STATES, they have full Power to levy War, conclude Peace, contract Alliances, establish Commerce, and to do all other Acts and Things which INDEPENDENT STATES may of right do. And for the support of this Declaration, with a firm Reliance on the Protection of divine Providence, we mutually pledge to each other our Lives, our Fortunes, and our sacred Honor.

Signed by ORDER and in BEHALF of the CONGRESS,

JOHN HANCOCK, PRESIDENT.

ATTEST. CHARLES THOMSON, SECRETARY.

PHILADELPHIA: PRINTED BY JOHN DUNLAP.
E-PROXIES FOR SALE? CORPORATE VOTE-BUYING IN THE INTERNET AGE

Douglas R. Cole*

Abstract: Advances in electronic communications technology promise to invigorate shareholder voting as a viable tool for corporate governance, for example by decreasing the cost, and thereby increasing the frequency and effectiveness, of proxy fights. Increased use of shareholder voting, though, forces renewed focus on issues related to the shareholder voting process. One such issue is vote-buying. Traditionally, courts have treated vote-buying in the corporate context as per se illegal. More recently, however, courts have relaxed their attitude toward such transactions, a move generally applauded by commentators. This article argues that the newfound judicial acceptance of vote-buying is problematic, at least for publicly-held corporations. The article examines the reasons offered in support of vote-buying in such corporations, and suggests that the same benefits could be obtained, without the threat of harm presented by vote-buying, through the use of turnout payments to encourage shareholder participation in corporate voting contests. With regard to closely-held corporations, however, the article argues that vote-buying serves a useful preference aggregation function and generally should be permitted.

“Vote-buying” evokes images of illicit deals by campaign workers attempting to fix political races on behalf of their candidates.¹ Vote-buying, however, is not limited to the political arena; it has implications for corporate governance as well. Recently, questions regarding the appropriate treatment of vote-buying, particularly vote-buying in the context of political elections, have sparked renewed interest among commentators.² While focusing their efforts largely on civic voting, these

¹. For instance, in the late 1800s, William Tweed and his Tammany Hall cronies made extensive use of vote-buying (among other things) to control New York politics. See ALEXANDER B. CALLOW, JR., THE TWEED RING 208–09 (1966) (“On the eve of election, the boss and his ward leaders collected an army of the party faithful, the unemployed, the underworld, the flotsam and jetsam of the slums—the bums, frowsy, bleary-eyed, and ragged—and entertained them royally at the saloons, where votes were openly bought and sold.”). More recently, vote-buying allegations arose in the 2000 presidential election. See Richard S. Dunham, Sleight of Hand at the Polls: Florida Isn’t an Exception, BUSINESS WEEK, Nov. 27, 2000, at 50, available at 2000 WL 24486679 (reporting that in Wisconsin a campaign worker “was caught on tape trying to influence voters by offering homeless people cigarettes if they would cast absentee ballots”); cf. Richard L. Hasen, Vote Buying, 88 CAL. L. REV. 1323, 1327–28 (2000) (giving brief historical overview of vote-buying in political contests and tracing it as far back as Greece and Rome).

². See generally Hasen, supra note 1 (analyzing the potential bases supporting a ban on vote-buying in civic elections); Saul Levmore, Voting with Intensity, 53 STAN. L. REV. 111 (2000).
scholars express support for the proposition that vote-buying in the corporate governance realm largely is, and should be, permissible. In doing so, they draw on recent case law and earlier commentary suggesting that a judicial tolerance toward corporate vote-buying is consistent with shareholder wealth-maximization principles.

This Article argues that a permissive attitude toward vote-buying arrangements in the corporate governance realm is inappropriate, and that both courts and commentators have failed to recognize or appreciate the harm to shareholders such arrangements could entail. This failure takes on special significance in light of the potential for an increase in such transactions provided by the reduced communication costs resulting from increased Internet connectivity. This Article focuses on corporate vote-buying in the Internet age, identifying both the possibilities and the risks that such technology provides for enhancing shareholder participation in corporate governance.

Whether vote-buying matters turns, at least in part, on whether shareholder voting matters. The role and importance of shareholder voting in corporate governance has long been the subject of debate among courts and commentators. Theoretically, the power of the vote (discussing the use of voting as a technique for aggregating preferences and considering the role vote-buying might play in such preference aggregation).

3. The support these commentators offer for vote-buying is by no means unqualified. Saul Levmore, for instance, states that “[a]lthough there is not much law on the subject, I will go along with the current wisdom that vote-buying in corporate law is now more acceptable than it once was and that we are soon likely to see more explicit legislative and judicial approval of trading in shareholder voting rights.” Levmore, supra note 2, at 138.


offered the ultimate corporate owner (i.e., the shareholder) a means to retain control over corporate management. Recognition of the collective action problems that plague the voting process, however, as well as the availability of much stronger tools to discipline corporate managers, led many commentators to conclude that the power to vote was almost meaningless as a mechanism for corporate governance.\textsuperscript{6} Indeed, by the mid-1980s, discussion of shareholder voting as a viable means of removing entrenched corporate management was supplanted almost

\textsuperscript{6} The recognition of the collective action problems presented by voting among a widely-dispersed share ownership, with the concomitant agency cost problems, is widely attributed to Adolf Berle and Gardiner Means in \textit{Adolf A. Berle, Jr. & Gardiner C. Means, The Modern Corporation and Private Property} (1932). For a discussion of the impact of these collective action problems on the viability of shareholder voting as a control mechanism, see sources cited supra note 5. \textit{See also} Mark J. Roe, \textit{A Political Theory of American Corporate Finance}, 91 Colum. L. Rev. 10, 12–13 (1991) (discussing Berle-Means view of importance of shareholder voting); Zohar Goshen, \textit{Controlling Strategic Voting: Property Rule or Liability Rule?}, 70 S. Cal. L. Rev. 741, 749–63 (1997); Sanford J. Grossman & Oliver D. Hart, \textit{Takeover Bids, the Free-Rider Problem and the Theory of the Corporation}, 11 Bell J. Econ. 42 (1980) (discussing collective action problems shareholders face in responding to a takeover bid). These collective action problems are discussed in greater detail \textit{infra} Section I.B.
entirely by reliance on the possibilities presented by corporate acquisition through means of a tender offer. 7

Since the late 1980s, however, corporate managers seeking to insulate themselves,8 often aided by state legislatures leery of seeing local corporations acquired and the corporation’s assets moved beyond the reach of the legislature’s taxing power,9 have erected sophisticated takeover defenses designed to prevent acquisition. Thwarted by these defenses, potential acquirers once again turned to the power of the shareholder vote, attempting to wrest control of the corporation through proxy contests or consent solicitations,10 but their initial attempts proved both costly and largely unsuccessful.11

7. See, e.g., Peter J. Henning, Corporate Law After the Eighties: Reflections on the Relationship Between Management, Shareholders and Stakeholders, 36 ST. LOUIS U. L.J. 519, 533 (1992) (stating that tender offers replaced proxy fights as principal means of hostile acquisition); Lyman Johnson & David Millon, Misreading the Williams Act, 87 MICH. L. REV. 1862, 1864 (1989) (“Over the past two decades, the hostile takeover has replaced the proxy fight as the more potent vehicle for wresting corporate control from incumbent management.”); Guhan Subramanian, A New Takeover Defense Mechanism: Using an Equal Treatment Agreement as an Alternative to the Poison Pill, 23 DEL. J. CORP. L. 375, 383–97 (1998) (discussing history of shift from proxy fights to tender offers as means of obtaining change in corporate control). Of course, the new owner (i.e., the successful tender offer bidder) would accomplish the actual change in management through use of his or her shareholder powers. The concentration of all of the shares in a single owner, however, would remove the collective action problems that had protected the entrenched management prior to the transaction.

8. Not surprisingly, corporate managers do not express their goals so directly. Rather, they typically speak of protecting the shareholders’ “long-term interests.” According to these managers, the market is “undervaluing” the company, and the shareholders need only wait for the market to return to rationality for them to realize share gains far in excess of those offered by the corporate “raiders.” See, e.g., Bank of Scotland Launches Surprise $34 Billion Hostile Bid for NatWest, DOW JONES BUSINESS NEWS, Sept. 24, 1999 (reporting that the NatWest Board requested shareholders to reject a takeover bid at a 20% premium to market prices because it “undervalues” the company); Allied Colloids Snubs Hercules’ $1.8-Billion Bid, CHEMICAL WEEK, Dec. 3, 1997, at 11 (stating that in response to bid at 23% premium to market price, Allied Colloids’ board “advised shareholders to reject the bid, saying that it undervalues the company”).

9. In the words of one commentator:

What motivates states to enact antitakeover legislation? Wary of raiders’ tendencies to liquidate companies, close plants, and lay off workers, state legislators seek to protect home-based businesses. More specifically, the impetus likely derives from two sources: the enacting state’s desire to protect nonshareholder constituencies, including managers who are unable or unwilling to persuade shareholders of the value of internal defensive measures, and financial protectionism, where states desire to retain and maximize tax-generating resources.


10. See J. Harold Mulherin & Annette B. Poulsen, Proxy Reform as a Single Norm? Evidence Related to Cross-Sectional Variation in Corporate Governance, 17 J. CORP. L. 125, 131–32 (1991) (discussing evidence regarding shift from hostile takeovers back to proxy solicitations as means to acquire corporate control). Proxy solicitations and consent solicitations are two different but related
The advent of new technology, in particular the communications opportunities afforded by personal computers and the Internet, holds renewed promise for the role of shareholder voting in corporate governance. Electronic mail and Internet-based communications provide dissident shareholders (or potential acquirers) a means of instant, widespread, low cost disclosure of the information they wish to share with fellow shareholders. This technology similarly offers the possibility to drastically reduce the price associated with soliciting votes, whether proxies or consents, from these shareholders. While these technological advances may herald the emergence of shareholder voting as a viable means of acquiring and/or exercising corporate control, they also focus renewed attention on the scope of permissible inducements one may means of achieving action through shareholder voting. Proxy solicitation refers to the process of obtaining a shareholder’s right to vote on issues presented at the annual meeting. See, e.g., DEL. CODE ANN. tit. 8, § 212 (1991 & Supp. 2000) (describing proxies). Consent solicitation, on the other hand, involves obtaining written shareholder consent to some proposed action not in conjunction with the annual meeting. See, e.g., id. tit. 8, § 228 (describing consents). For the regulations regarding proxy and consent solicitations in public companies, see SEC Rules 14a-1 to 14a-15; 17 C.F.R. §§ 240.14a-1 to 240.14a-15 (2000). Each solicitation method has distinct advantages and disadvantages. The machinery of the proxy solicitation process, as well as the timing of the annual meeting, are largely in the control of corporate management. This often makes meaningful participation by dissidents difficult. Consent solicitations, because they need not occur in conjunction with an annual meeting, are more easily initiated and controlled by a dissident shareholder. The difficulty here, however, is that action by written consent requires consent by a majority of all outstanding shareholders, while action by proxy requires only a plurality of the votes actually cast at the annual meeting (assuming a quorum is present). For a more complete discussion of the distinction between proxy solicitation and consent solicitations and the advantages and disadvantages of each, see generally Eric S. Robinson, Defensive Tactics in Consent Solicitations, 51 BUS. LAW. 677 (1996).


12. In 1998, for instance, ADP Investor Communications Services, a proxy solicitation firm, charged its corporate clients $0.03 for processing proxies returned by Internet, less than one-tenth the $0.34 it charged for processing proxies returned by mail. Howard M. Friedman, Securities Regulation in Cyberspace 11-16 (2d. ed. 1998); see also Catherine S. Powell, Electronic Proxy Voting in Wisconsin Would Benefit Corporation, Shareholders, 72 WIS. LAW. 28 n.6 (Feb. 1999) (reporting similar figures). If investors agree to receive their proxy materials electronically, as well as to vote electronically, the savings are even greater. Friedman, supra, at 11-16.
offer a shareholder in an attempt to secure her vote. In particular, should it be permissible to buy shareholder votes?

This Article examines corporate voting in the Internet age and discusses the appropriateness of vote-buying in light of the expanded possibilities for active shareholder involvement in corporate governance. Section I discusses the role shareholder voting plays within a corporation. After providing a brief historical perspective, Section I discusses the various collective action problems that undermine shareholders’ attempts to use voting as a means of corporate management. It then discusses the potential impact communications technology may have on those collective action problems.

Section II describes “vote-buying” and the role it has played in corporate voting. After briefly reviewing the early judicial treatment of vote-buying as per se illegal, it discusses the shift to a rule of reason regime, first announced in the seminal case of Schreiber v. Carney.13 This modern approach requires a case-by-case analysis of vote-buying transactions and attempts to separate “good” vote-buying from “bad.”

Section III addresses whether the newfound judicial receptiveness toward vote-buying (since Schreiber, courts have not struck down a single instance of alleged vote-buying)14 actually advances the goal of shareholder wealth maximization. It considers whether vote-buying in a publicly held corporation, in light of the reduced transaction costs promised by electronic communication, might present a viable mechanism for invigorating shareholder participation. It concludes, however, that despite suggestions by commentators to the contrary, a permissive attitude toward traditional vote-buying cannot be justified on this basis. Indeed, closer analysis of previous cases suggests that even apparently beneficial instances of traditional corporate vote-buying could instead be seen as examples of coercive wealth transfers between various classes of corporate stakeholders. This Section then distinguishes a separate, but related, concept—the idea of paying shareholders to participate in voting, independent of the voting option they select. It suggests that this type of vote-buying—a “turnout incentive”—may hold promise for empowering shareholder voting while avoiding the difficulties traditional vote-buying presents.

Finally, Section IV examines whether a different judicial treatment of vote-buying in the close corporation context may be appropriate. In

---

13. 447 A.2d 17 (Del. Ch. 1982).
14. For discussion of the post-Schreiber cases, see infra notes 150–87 and accompanying text.
particular, it suggests that with regard to close corporations, the shareholders’ principal concerns often arise from difficulties in coordination rather than collective action problems. This Section argues that shareholders in close corporations are less likely to share uni-peaked preferences on corporate voting issues than the shareholders in large public corporations, and suggests that close corporations could use vote-buying as a means of aggregating preferences. Section IV concludes that while other techniques could potentially achieve the same results, vote-buying might represent an efficient mechanism for accomplishing such preference aggregation, justifying a different treatment with regard to close corporations.

This Article concludes that while the Internet and electronic communication may reinvigorate shareholder voting as a useful tool for corporate governance, courts and legislatures should be reluctant to accept vote-buying as part of the new voting regime, at least with respect to publicly-held corporations.

I. THE ROLE OF SHAREHOLDER VOTING IN CORPORATE GOVERNANCE AND THE IMPACT OF THE INTERNET

Perhaps the fundamental problem plaguing the corporate structure is the divergence between ownership and control that occurs as a result of shareholders (i.e., owners) entrusting control over their capital contributions to a separate group, corporate managers, whose interests may diverge from those of shareholders. Shareholders confront the

15. “Uni-peaked preferences” refers to those situations in which the voters (here shareholders) share identical preferences with regard to the voting alternatives presented. DANIEL A. FARBER & PHILLIP P. FRICKEY, LAW AND PUBLIC CHOICE 48–49 (1991); see also Maxwell L. Stearns, Standing Back from the Forest: Justiciability and Social Choice, 83 CAL. L. REV. 1309, 1378–81 (1995) (providing an example of voting outcomes in the face of uni-peaked preferences). That is, the voters would all rank the various potential outcomes in the same order. By contrast, “multi-peaked preferences” refers to situations in which the members of the group have divergent preferences. Stearns, supra, at 1331–33. The magnitude of a given peak is a function both of the number of people that share the preference, and the intensity with which they hold it.

16. Berle & Means are credited with first recognizing the importance of the divergence problem. See BERLE & MEANS, supra note 6. The notion that the divergence between ownership and control is a central problem of corporate law is a staple of the law and economics literature. See, e.g., Bernard Black & Reinier Kraakman, A Self-Enforcing Model of Corporate Law, 109 HARV. L. REV. 1911, 1913 (1996) (explaining that corporate law “should have the same principal goal in developed and emerging economies—succinctly stated, to provide governance rules that maximize the value of corporate enterprises to investors”); Matheson, supra note 9, at 709 (stating that “The great challenge of corporate law in the modern era, then, is to minimize agency costs by constraining abuse of managerial discretion”); Henry N. Butler & Fred S. McChesney, Why They Give at the
choice between investing time and effort in monitoring the corporate managers, or suffering the deleterious effects (for example, excessive salaries or shirking) that may result from such divergence. Much of corporate law can be understood as an attempt to adopt rules that minimize these agency costs. In this Section, the Article discusses the agency costs shareholders face and examines the role voting plays in minimizing those costs. In particular, it discusses the collective action problems that affect voters in their attempt to coordinate voting efforts. It then discusses the promise that modern electronic communications, in particular the Internet, hold for enhancing shareholder coordination.

Office: Shareholder Welfare and Corporate Philanthropy in the Contractual Theory of the Corporation, 84 CORNELL L. REV. 1195, 1197 (1999) (discussing how both the agency cost model of corporate law and the contractual theory of the firm are grounded in concerns about divergence between ownership and control); ECONOMIC STRUCTURE, supra note 4, at 9–11; Frank H. Easterbrook & Daniel R. Fischel, The Proper Role of a Target’s Management in Responding to a Tender Offer, 94 HARV. L. REV. 1161, 1169–74 (1981). This view, while widely accepted, is controversial. Some commentators have suggested that the importance of this divergence as an explanatory force of corporate law has been overstates. For example, in Blair & Stout, supra note 5, the authors contend that this divergence theory of corporate law reflects a “shareholder primacy” view (i.e., that the directors should focus solely on the wealth effects of their decisions on shareholders) that fails to explain corporate law either normatively or positively. They suggest that corporate management is better understood as a “mediative hierarchy” that balances competing desires advanced by various corporate constituencies. Id. at 271–87; see also, e.g., William W. Bratton, The Economic Structure of the Post-Contractual Corporation, 87 NW. U. L. REV. 180, 208–15 (1992), and sources cited therein, discussing the corporate form in terms of its “mediative effects.” Other commentators have suggested that this “shareholder primacy” view should be replaced by a paradigm in which the scope of the directors’ focus extends to all corporate “stakeholders” or even to society as a whole. See generally Ronald M. Green, Shareholders as Stakeholders: Changing Metaphors of Corporate Governance, 50 WASH. & LEE L. REV. 1409 (1993) (suggesting beneficial aspects of “multi-constituency” approach to corporate law, whereby director duties are directed toward those affected by corporate acts); Morey W. McDaniel, Stockholders and Stakeholders, 21 STETSON L. REV. 121 (1991) (arguing that directors should be empowered to pursue dual goal of maximizing stockholder wealth and minimizing stakeholder loss); David Millon, Redefining Corporate Law, 24 IND. L. REV. 223 (1991).

17. See, e.g., Matheson, supra note 9, at 709; ECONOMIC STRUCTURE, supra note 4, at 34–35. As noted above, some commentators challenge the importance of these agency cost issues. See supra note 16. Examples of tools designed to minimize agency costs are the duties of care and loyalty. The prospect of ex post director liability for mismanagement or theft provides appropriate ex ante incentives for directors to avoid shareholder harms. As with voting, however, the effectiveness of judicial enforcement of these duties as a means to discipline corporate managers is open to question. See Lawrence A. Hamermesh, Corporate Democracy and Stockholder-Adopted By-Laws: Taking Back the Street?, 73 TUL. L. REV. 409, 465–66 (1998).
Corporate Vote-Buying

A. The Divergence of Ownership and Control

In most publicly held corporations, shareholders are not involved in the day-to-day operations of the company; these operations are largely left to management. Management, however, may have interests that are not perfectly aligned with those of the shareholders. The classic example is the issue of management salaries. Managers, as a rule, would prefer high salaries and little work; shareholders, of course, have diametrically opposed preferences.

This divergence in preferences between the corporate principal (i.e., the shareholders) and the agents raises the prospect of monitoring costs. That is, shareholders are faced with the choice of either suffering the costs the divergences impose or undertaking monitoring to ensure that the management is faithfully carrying out its task of maximizing shareholder wealth. Such monitoring is, of course, not free, and shareholders seek to minimize the net costs associated with it. This can involve, for instance, attempts to create self-enforcing mechanisms to align management’s financial incentives with those of the shareholders.

18. Members of “management” can include shareholders. Indeed, in many, if not most, corporations, the corporate officers and directors receive at least part of their compensation in stock or stock options. For instance, according to the KORN/FERRY INTERNATIONAL 22ND ANNUAL BOARD OF DIRECTORS STUDY 6 (1995), 62% of outside directors receive stock or stock options as part of their compensation. See also Michael S. Knoll, An Accretion Corporate Income Tax, 49 STAN. L. REV. 1, 21 (1996) (noting that corporate directors and officers are typically compensated in stock or options); Laura Lin, The Effectiveness of Outside Directors as a Corporate Governance Mechanism: Theories and Evidence, 90 N. W. U. L. REV. 898, 919 n.113 (1996) (citing Korn/Ferry study). In managing the day-to-day affairs of the corporation, however, the persons are acting as “managers,” not in their capacity as “shareholders.”

19. See George W. Dent, Jr., Toward Unifying Ownership and Control in the Public Corporation, 1989 WIS. L. REV. 881, 889–91. This is not to suggest that shareholders as a group prefer that the corporation pay low salaries. Low salaries might prevent the firm from attracting talented managers. Rather, the shareholders would prefer that for a given level of productivity, the corporation pay as little as is necessary. Managers, on the other hand, would prefer greater corporate largesse.

20. This again assumes a “shareholder primacy” view of corporate law. Commentators arguing against this view have suggested that management’s “goal” is actually much broader, and includes balancing the interests of competing corporate “stakeholders” or even society as a whole. See supra note 15. The proponents of these broader views of corporate responsibility, however, largely have failed to explain why shareholders would choose to invest their capital in firms that undertake such balancing acts (assuming the existence of other corporations that do not elect to treat shareholder interests as one of many “balancing” factors).

21. More correctly, shareholders attempt to minimize the sum of the monitoring costs plus the costs resulting from the agents’ dilatory conduct. That is, the rational shareholder will incur an additional dollar of monitoring cost only if it results in a savings of at least one dollar in losses arising from agent misconduct.
Examples of this include performance-based compensation and stock options.22

Shareholder voting is another weapon in the shareholders’ arsenal for use in this ongoing quest to minimize the costs resulting from the divergence between management and shareholder interests. Through their collective votes, shareholders can elect the board of directors who, in turn, are responsible for managing the affairs of the corporation.23 If the board is lax in performing its oversight role or, alternatively, has been “captured” by management, shareholders can theoretically elect a new board that presumably will exercise greater vigilance. That is, through their collective will, shareholders can effect changes in corporate control either by replacing directors and (indirectly) management24 in the event of poor shareholder returns, or even by dissolving the corporation or merging it with another.25

Courts and legislatures have viewed this right to control the corporation through voting as one of the most fundamental aspects of share ownership. Courts, for example, have characterized this right as the shareholder’s “supreme right and main protection,”26 or “a right so essential to the enjoyment of property in stock that, it has been suggested, it is a part of the property itself.”27 State legislatures and the


23. See, e.g., DEL. CODE ANN. tit. 8, § 141(a) (1991 & Supp. 2000) ("The business and affairs of every corporation organized under this chapter shall be managed by or under the direction of a board of directors 

24. Shareholders do not have the right to hire or fire corporate officers and cannot force directors to do so. See DEL. CODE ANN. tit. 8, § 141(a) (affairs of corporation managed by the board); Alabama By-Products Corp. v. Cede & Co., 657 A.2d 254, 265 (Del. 1995) ("It is a fundamental principle of Delaware General Corporation Law that directors, rather than shareholders, manage the business and affairs of the corporation."); see also Peter V. Letsou, Shareholder Voice and the Market for Corporate Control, 70 WASH. U. L.Q. 755, 759–63 (1992) (discussing limitations on ability of shareholder to direct day-to-day activities of corporation). The full extent of shareholder participation in day-to-day corporate management is the right to elect the board. Of course, the right to elect (or remove) directors provides indirect control over director choices. See, e.g., DEL. CODE ANN. tit. 8, § 141(k).

25. See DEL. CODE ANN. tit. 8, §§ 251 (merger), 271 (dissolution). The shareholders cannot accomplish either a merger or dissolution directly. A merger, for instance, must begin with a resolution by the board, id. tit. 8, § 251(b), which in turn is put to the shareholders for vote. The shareholders can, however, elect directors supportive of a merger, and thus accomplish indirectly what they cannot accomplish directly.


27. In re Diamond State Brewery, 2 A.2d 254, 256–57 (Del. Ch. 1938); see also, e.g., Hoschett v. TSI Int’l Software, Ltd., 683 A.2d 43, 44 (Del. Ch. 1996) (referring to the “critical importance of
Corporate Vote-Buying

drafters of the Model Business Corporation Act have taken a similar view toward the importance of shareholder voting rights, mandating that for every corporation, one or more classes of shares, taken together, must have unlimited voting power in the company.28

Of course, shareholder voting is not the only means, or even necessarily a primary one, of actually performing management oversight. For instance, reputational constraints often bind management’s discretion.29 Even the CEO of a corporation must be concerned about her next job. Laying waste to the corporate assets of the company she is managing is hardly a strong recommendation for her next position.

The market for corporate control provides an even more important monitoring function.30 The threat of hostile takeover with its resultant shareholder voting both to the theory and to the reality of corporate governance”); Carmody v. Toll Bros., Inc., 723 A.2d 1180, 1193 (Del. Ch. 1998) (shareholder vote is “ideological underpinning upon which the legitimacy of directorial power rests”) (quoting Blasius Indus. v. Atlas Corp., 564 A.2d 651, 659 (Del. Ch. 1988)).

28. For example, New York’s corporate code specifies that:

the certificate of incorporation may deny, limit or otherwise define the voting rights and may limit or otherwise define the dividend or liquidation rights of shares of any class, but no such denial, limitation or definition of voting rights shall be effective unless at the time one or more classes of outstanding shares or bonds, singly or in the aggregate, are entitled to full voting rights.

N.Y. BUS. CORP. LAW § 501 (McKinney 1986) (emphasis added). The Model Business Corporation Act (MBCA) provides that the articles of incorporation must authorize “one or more classes of shares that taken together have unlimited voting rights.” MODEL BUS. CORP. ACT ANN. § 6.01(b) (1998). Twenty-four states have adopted this statute. See id. at xxvii (listing states that have adopted the act). For examples of state statutes, see FLA. STAT. ANN. § 607.0601 (West 2000); N.C. GEN. STAT. § 55-6-01(c) (1999); WASH. REV. CODE § 23b.06.010(2) (2000). While not adopting the MBCA language, Delaware’s corporate statute provides a similar result, at least implicitly. See DEL. CODE ANN. tit. 8, § 251(b).

29. See Stephen J. Choi, Company Registration: Toward a Status-Based Antifraud Regime, 64 U. CHI. L. REV. 567, 583–84 (1997) (discussing reputational constraints on managers with respect to fraudulent disclosures); Smith, supra note 22, at 1079 (discussing limitations reputational concerns impose on managerial behavior). For a discussion of some limitations of reputational constraints as a means of minimizing agency costs, see Mitu Gulati, When Corporate Managers Fear a Good Thing Is Coming to an End: The Case of Interim Nondisclosure, 46 UCLA L. REV. 675, 694–702 (1999) (discussing the breakdown of reputational constraints in corporate end-game scenarios).

30. According to Matheson, supra note 9, at 710:

In addition, market forces, like the market for corporate control, may also constrain managerial abuses. At one extreme, this monitoring model views shareholders as owners of the corporation and posits that stock ownership is like ownership of any other property. Unhappy shareholders can simply sell their shares to others. At the least, such conduct should evidence their displeasure with management. If sold to a bidder in a tender offer, such a sale might result in the ouster of management. Throughout the 1970s and much of the 1980s, this market in corporate control acted as an important mechanism for monitoring corporate behavior.
removal of entrenched management can provide strong incentives to fully utilize corporate assets. 31 Notwithstanding these alternative methods for disciplining corporate management, however, an important question to consider in discussing any aspect of corporate voting, including vote-buying, is the extent to which it aids or injures the shareholders’ attempt to use their voting power as a means of monitoring management performance.

B. Collective Action Problems and Their Impact on Shareholder Voting as a Means To Minimize Agency Costs

While voting in theory empowers shareholders to monitor corporate managers, in practice, collective action problems often prevent shareholders from using the power of the vote to do so. These problems arise from three related sources. First, if there are a significant number of corporate shareholders, each may be rationally apathetic with respect to the outcome of a particular vote. Second, any group activity that results in a uniform benefit to the group independent of individual input of any group member is subject to what is commonly called the free-rider problem. Finally, even absent free-riding, large groups face significant transaction costs in attempting to organize group behavior. Each of these issues contributes to the problems shareholders face in attempting to use their vote to maximize corporate returns.

1. Rational Apathy

As a general matter, that people choose to vote at all, whether in corporate or civic elections, is somewhat of a mystery. 32 A rational voter would vote only if the benefits of voting (i.e., the increased likelihood of the voter’s preferred outcome winning the election) outweigh the costs associated with casting the vote. The only vote that matters, however, is the “swing vote,” and even if there is only a relatively small number of

31. How successful these incentives are at disciplining corporate management is a topic of strong debate. Easterbrook and Fischel, for example, have argued that the market forces are quite strong. See ECONOMIC STRUCTURE, supra note 4, at 96. Others have argued that the strength of these incentives has been vastly overstated. See, e.g., Dent, supra note 19, at nn.33–37 (arguing that the takeover market does not really act to constrain managerial discretion).

32. Commentators often refer to the “voter’s paradox” or the “paradox of voting.” For a detailed discussion of the paradox of voting, and some potential explanations for why people vote in spite of the seeming irrationality of doing so, see Richard L. Hasen, VOTING WITHOUT LAW, 144 U. PA. L. REV. 2135, 2138–47 (1996).
voters, the possibility that a given voter will cast this decisive ballot is small. Although this is especially true with respect to civic elections given the “one-person one-vote” rule, the same theory applies to large publicly-held corporations with widely-dispersed share ownership.33

Moreover, even if the shareholder were convinced that his or her vote would be the deciding vote, the shareholder may lack incentives to participate in the voting process. In order to vote correctly on a given proposition, a shareholder must first determine the probable returns associated with each possible voting option. For example, assume the shareholders are voting on a single issue with two outcomes—A or B. Correctly choosing between them requires the shareholder to determine the wealth effects of A versus B. Unfortunately, developing this information is, or at least may be, very expensive for the shareholder.34

For instance, assume that the company has put the question of whether it should market an additional $100 million in bonds in order to build a new production facility to shareholder vote.35 Each shareholder can either vote yes or no. The “proper” vote, of course, will depend on the interest rate of the bonds, the probable rate of return on the production facility, the opportunity cost associated with dedicating management time and energy into the development of new production facilities, and myriad other factors.

An in-depth assessment of these factors would prove very expensive. Any given shareholder choosing whether to make this expenditure must acknowledge that he will get but a very small proportion of the return of any increase in the net worth of the company. That is, assume that the investment would be advisable and would increase the net worth of the company by roughly $10 million. If there are 10 million shares, that results in only a $1 per share increase. Shareholders holding few shares could ill afford to invest significant amounts to determine whether they should vote yes or no. In such situations, it may well be rational for the shareholder to appear apathetic by choosing not to vote.36

33. There is one important difference between civic and corporate voting in this regard. Civic voting typically does not have a “quorum” requirement, while corporate voting does. See, e.g., DEL. CODE ANN. tit. 8, § 216 (1991 & Supp. 2000). Thus, if there are insufficient shareholder votes present either in person or by proxy, shareholder action cannot occur. In general, this should increase shareholder incentives to vote.

34. See Clark, supra note 4, at 779–82; Easterbrook & Fischel, supra note 5, at 402–03.

35. An example similar to the one posited here is offered in ECONOMIC STRUCTURE, supra note 4, at 66–67.

36. See Clark, supra note 4, at 779–82; Peter J. Henning, Corporate Law After the Eighties: Reflections on the Relationship Between Management, Shareholders and Stakeholders, 36 ST. LOUIS
2. **The Free-Rider Problem**

Even shareholders owning large blocks of shares who would otherwise have the proper incentives to invest in information may choose not to do so because of the free-rider problem. Assume, for instance, that in the example above two shareholders each hold three percent blocks in the company. Either of them could afford to invest roughly $90,000 in researching the issue and still increase their wealth. Each of them, however, would prefer that the other make the research expenditure. That is, each would prefer to free-ride on the information produced by the other. Because the shareholders benefit collectively (on a per share basis) from the “proper” outcome on the vote, each individual shareholder has the incentive to let another shareholder undertake the informational costs associated with voting.

The free-rider problem is further exacerbated in the voting context because information is a public good, and voting is an information-intensive activity. One of the prime motivators of information development in other markets—the ability to profit from private

---

37. Clark, supra note 4, at 783–84.

38. Of course, this is not entirely correct. The return on this investment to one of the three percent shareholders would indeed be $90,000, so either one could afford to spend $89,999 and still increase their wealth. However, before performing the analysis, the shareholder does not know what the profit will be; that is what he was attempting to determine. Thus, before the analysis, the shareholder would not have a good idea what investment in information would be prudent. For instance, assume that after investigation, the shareholder determines that the result of the vote would have no wealth effects on the corporation. That is, the net value of the corporation would remain exactly the same independent of which option was chosen. In that case, any expenditure by a shareholder would be wealth decreasing for that shareholder.

Corporate Vote-Buying

information—is not present in the voting arena. In other areas, people can often offset their information costs by an associated profit from developing information that is not available to others in the market. Patent, copyright, and trade secret law, for example, allow people who develop valuable information to benefit privately from it. With voting, on the other hand, private information does very little good. Imagine, for instance, that a minority shareholder invests the necessary amount to determine the “proper” voting choice. In order for that information to benefit the shareholder at all, the shareholder must convince the remaining shareholders, or at least a majority, to vote in the same manner as he does. That is, the information that he developed does him absolutely no good in the voting contest as private information; it is only by the dissemination of the information that the shareholder can hope to benefit at all. This results in a systematic underinvestment in voting information.

40. Note that this comment is strictly confined to voting. Shareholders may be able to profit from private information about a company through other means, for example by purchasing shares (if the information indicates that the market is undervaluing the company) or selling shares (if the information shows that the market is overvaluing the company). Concerns that corporate insiders might profit privately from information derived as a result of their positions in the corporation led to the inclusion of a ban on insider trading in the Securities and Exchange Act of 1934 (1934 Act). See 15 U.S.C. § 78p(b) (1994).

41. See, e.g., Rebecca S. Eisenberg, Patents and the Progress of Science: Exclusive Rights and Experimental Use, 56 U. Chi. L. Rev. 1017, 1024–27 (1989) (explaining how the patent system serves to overcome free-rider problems associated with information goods such as inventions). Securing private benefits is not the justification for the patent or copyright statutes. Courts and commentators generally agree that the principal purpose underlying these laws is securing public benefit through increased investments in invention and the arts. See, e.g., Carol M. Silberberg, Preserving Educational Fair Use in the Twenty-First Century, 74 S. Cal. L. Rev. 617, 621 (2001) (“The immediate effect of our copyright law is to secure a fair return for an author’s creative labor. But the ultimate aim is, by this incentive, to stimulate artistic creativity for the general public good.” (citing Twentieth Century Music Corp. v. Aiken, 422 U.S. 151, 156 (1975))). The means of achieving this end, however, is through allocation of exclusive rights, thus resulting in private benefits.

42. This is typically accomplished through a proxy or consent solicitation. See infra Section I.C.

43. Of course, this is not entirely true. Rather than try to influence the vote, the shareholder can attempt to profit from the privately-developed information by buying or selling shares of the stock. Assuming the shareholder has information that the other investors do not have indicating that the value of the company will increase, the shareholder could conceivably enter into transactions with the other shareholders to purchase their shares at advantageous prices. This immediately suggests an interesting question. Why would a shareholder convinced that the value of the company would increase with a correct voting outcome ever attempt to buy votes? It seems it would be much more likely that the shareholder would elect to buy shares. By buying votes, the informed shareholder must spread the gains derived from use of the privately-developed information with the other shareholders. By buying their shares, the shareholder could keep the entirety of the gains. See infra notes 211–17 and accompanying text. It may be that constraints imposed either by liquidity
3. Voting and Coordination Problems

Coordination problems among shareholders further compound this problem. Imagine, for instance, that in the example offered above a “correct” decision on the voting issue would cost $10,000 to develop. Imagine further that there are 1000 shareholders who each invest $1000 in information.45 If the shareholders each act independently, this would result in a total research expenditure of $1 million, and yet still result in each of the shareholders holding inadequate information to cast his or her ballot appropriately. Clearly, the shareholders would benefit from some type of coordinated research.

C. Proxy Solicitation and Collective Action Problems

One way in which shareholders could attempt to coordinate their voting activities would be through use of some form of proxy solicitation process. A simplified description of such a process would be one in which a dissident group puts forward a proposal inconsistent with a proposal offered by the board.46 This could be, for example, a different slate of proposed directors, a bylaw amendment, or a response to a merger bid from another firm.47 The board and the dissident group then compete for proxies (or consents) from the remaining shareholders.

Providing for a proxy solicitation process addresses, at least in theory, shareholder collective action problems in three ways. First, to the extent that one or more of the shareholders develops information, it gives those shareholders a method for distributing it to other shareholders. Even though shareholders may be rationally apathetic given the prospect of developing their own voting information, to the extent that they can receive it free of charge from others, they would presumably be willing

limitations or a shareholder rights plan prevent acquisition of additional shares. In such cases, sharing the information with the other shareholder voters is a second-best solution. See infra notes 249–62 and accompanying text. At least the shareholder will participate in the resulting gain to the extent of his or her holdings.

45. See ECONOMIC STRUCTURE, supra note 4, at 66.
47. In general, a proxy solicitation could occur in connection with any issue on which a shareholder vote is required. Common examples of such acts include those listed in the text. See, e.g., DEL. CODE. ANN. tit. 8 §§ 141(d) (shareholder vote elects board); 242 (shareholder vote required to amend charter), 251 (shareholder vote required to approve merger) (1991 & Supp. 2000).
to act on it.\textsuperscript{48} Second, an individual shareholder (at least one who owns a significant number of shares) may be more willing to invest in voting information if she knows that there is a method in place for distributing the information, thereby making it more likely that the information would actually influence the outcome of the vote. Finally, a proxy solicitation process ameliorates, at least to some degree, the shareholder coordination problem. It encourages shareholders to share information in a way that helps prevent multiple shareholders from making overlapping investments in voting information.

While the possibilities of such contests are at least conceptually beneficial to shareholders, in practice they are of questionable effectiveness. The Securities and Exchange Commission (SEC) has implemented rules for proxy contests with respect to securities registered under the Securities and Exchange Act of 1934 (1934 Act).\textsuperscript{49} Section 14(a) of the 1934 Act requires that any communications made with a shareholder for the purpose of securing the shareholder’s proxy must comply with this set of rules.\textsuperscript{50} These rules specify, for example, the information that must be sent,\textsuperscript{51} how that information must be presented,\textsuperscript{52} and the filing requirements the soliciting party must meet before sending the solicitation.\textsuperscript{53} While these rules may help in terms of increasing shareholder comprehension of, and willingness to rely on, the information provided, they also impose substantial costs on dissidents’ attempts to communicate with their fellow shareholders.\textsuperscript{54}

\textsuperscript{48} This is not a complete answer to the rational apathy problem. Note that even if the information is provided, the voting decision is not “free.” Shareholders must still invest sufficient time and energy to review and analyze the material presented in the proxy solicitation, and may even find it necessary to invest additional time and energy in verifying some or all of the information presented. Moreover, in a proxy solicitation contest, the shareholder would presumably need to read two or more competing proxy solicitations and attempt to determine which of the asserted positions best serves his or her interest. At the very least, however, use of the proxy solicitation process should lower the amount of investment in voting information, thereby presenting at least a partial response to the rational apathy situation.


\textsuperscript{50} 15 U.S.C. § 78n(a).

\textsuperscript{51} 17 C.F.R. §§ 240.14a-3 to 240.14a-4.

\textsuperscript{52} \textit{Id} § 240.14a-5.

\textsuperscript{53} \textit{Id} § 240.14a-6.

While the SEC continues to tinker with the proxy solicitation process in an attempt to invigorate it, it is fair to say that, to date, it has not been a panacea for the collective action problems shareholders face. To begin with, the contests tend to be relatively expensive to the challenger, on the order of several million dollars for a typical contest. Thus, in determining whether or not to develop voting information, the shareholder must not only consider the cost of developing the information, but also the cost of distributing it through the proxy process. At the same time, dissidents engaging in proxy solicitation contests have had only very limited success. One empirical study reviewed 76 proxy solicitation contests and determined that the challengers achieved complete success in only approximately 30% of cases, hardly the kind of success rate one would expect to induce major investment in such activities.

proxy rules inhibited the ability of shareholders to monitor management and check managerial excesses and inefficiency by placing a costly barrier in the way of communications among shareholders.”) (citing other sources); see also supra note 11 (collecting sources regarding average cost of proxy solicitation contest). It should be noted that the rules do also provide some benefits to dissidents. For instance, SEC regulations provide that if a dissident’s solicitation meets SEC requirements, the corporation must either provide a list of the voting shareholders to whom the information should be sent, or, at the soliciting party’s expense, mail the information to the shareholders itself. 17 C.F.R. § 240.14a-7.

55. See, e.g., SEC Release Nos. 33-7760, 34-42055, IC-24107, 64 FR 61408, available at 1999 WL 1014713 (Nov. 10, 1999) (making revisions to proxy solicitation rules for stated purpose of permitting “increased communications with security holders and the markets”); see also Bernard S. Black, Next Steps in Proxy Reform, 18 J. CORP. L. 26 (1993) (discussing then-pending SEC amendments to the proxy rules that subsequently have been enacted).

56. See Palmiter, supra note 11, at 896 n.71 (estimating expense of “about $5 million” for a typical proxy contest); SARGENT, supra note 11, at Intro. 2 (1993) (stating that a “dissident shareholder can conduct a proxy contest for $1 to $15 million”); Cowan, supra note 11, at 1 ($1.7 million exclusive of legal and investment banking fees); see also Bialkin et al., supra note 11, at 56; Mark A. Stach, An Overview of Legal and Tactical Considerations in Proxy Contests: The Primary Means of Effecting Fundamental Corporate Change in the 1990s, 13 GEO. MASON L. REV. 745, 776 (1991) (“Proxy contests can be very expensive. For example, during a proxy fight for control of Lockheed, the incumbents spent approximately $8 million and the insurgents spent approximately $6 million.”).

57. See André, supra note 4, at 578 n.184 (citing studies indicating that insurgents are successful only 20–40% of the time).

58. See Thomas & Martin, supra note 11, at 329–32 (conducting a review of 76 proxy contests conducted between 1986 and 1991). Thomas and Martin do note, however, that dissident groups are at least partially successful in slightly over 50% of all cases. Id. For a complete analysis of dissident success rates in proxy contests reported on an annual basis for the years 1957–77, and 1981–85, see RONALD E. SCHRAGER, CORPORATE CONFLICTS: PROXY FIGHTS IN THE 1980s 8–11 (1986).
D. The Impact of the Internet on Corporate Proxy Contests

As the foregoing discussion suggests, both shareholder voting generally and proxy solicitation in particular are relatively expensive. Participants in these activities face three principal types of costs: information costs (the costs of detecting managerial shortcomings and developing a competing position), disclosure costs (the regulatory compliance costs and the communication costs to transmit the information to other shareholders), and collection and tabulation costs (the costs to process the proxies returned as a result of the solicitation efforts). As such costs fall, however, proxy solicitation contests should become a more viable mechanism for encouraging shareholder participation.59

The development and implementation of Internet communications and electronic mail offer great promise to reduce all three types of costs.60 Unfortunately, until recently, attempts to adapt these technologies to the proxy contest realm faced both significant technological and legal hurdles. In 1990, less than 1% of American households had routine access to electronic messaging or the Internet.61 From a regulatory standpoint, perhaps the most significant impediment to the use of electronic communications in proxy contests was the statutory requirement under the corporate code of most states that a proxy be “signed” by the shareholder granting it, with no explicit authorization of electronic signature or transmission.62

---

59. See Andy Dworkin, Shareholder Knockout, THE OREGONIAN, Sept. 24, 2000, at E1: Several experts also said the Internet is turning fights in shareholders’ favor. People angry about a company’s performance can meet like-minded investors through stock chat rooms in Yahoo and other websites. They can post proxy statements and other campaign materials electronically, cutting down printing cost. “I think there’s going to be more and more dissident activity around the annual meeting as costs get lower and lower.” [John] Wilcox [Vice-Chairman of the proxy solicitation firm Georgeson Shareholder Communications, Inc.] said.

60. Friedman, for example, estimates that electronic distribution of proxy solicitation materials would result in savings of approximately $2.50 to $5.50 per shareholder in printing and mailing costs. FRIEDMAN, supra note 12, at 11-16. On the tabulation side, ADP Investor Communication Services, a firm that provides proxy distribution, collection, and tabulation services, charges $0.03 to process an Internet proxy, less than one-tenth the $0.34 it charges to process those proxies returned by mail. Id.

61. Thomas P. Vartanian, The Emerging Law of Cyberbanking: Dealing Effectively with the New World of Electronic Banking and Bank Card Innovations, in DOING BUSINESS ON THE INTERNET: THE LAW OF ELECTRONIC COMMERCE, 452 PLI/Pat 141, 197 (1996) (“In 1990, when Internet access first became more readily available to the public outside of government, research or academic organizations, there were approximately 1 million users.”).

Today, however, both the legal and the technological barriers have been significantly reduced. Over 40% of the shareholding public now has an e-mail address and daily Internet accessibility, up from approximately 26% only two years ago. Moreover, new subscribers are obtaining Internet access every day, and estimates have predicted that by the middle of 2001, over half of all individuals in the United States will have Internet access. Since 1985, thirty states have adopted statutes permitting electronic proxies. The Delaware code, for example, provides that:

Without limiting the manner in which a stockholder may authorize another person or persons to act for such stockholder as proxy pursuant to subsection (b) of this section, the following shall constitute a valid means by which a stockholder may grant such authority: . . . A stockholder may authorize another person or persons to act for such stockholder as proxy by transmitting or authorizing the transmission of a telegram, cablegram, or other means of electronic transmission to the person who will be the holder of the proxy . . . provided that any such telegram, cablegram or other means of electronic transmission must either set forth or be submitted with information from which it can be determined that the telegram, cablegram or other electronic transmission was authorized by the stockholder.

With these legal and technological barriers largely removed, the benefits of the Internet and electronic mail for proxy solicitation (and

---

64. See Digital Divide, supra note 63, at xv (Executive Summary).
65. See Jacqueline Dosick, State Law Amendments Make It Easier To Implement Electronic Proxy Voting, 3 No. 7 Wallstreetlawyer.com: Sec. Elec. Age 19 n.1 (Dec. 1999), available at WESTLAW, 3 No. 7 GLWSLAW 19 (last visited July 31, 2001) (listing 30 states plus Puerto Rico that, as of the date of publication, had statutes permitting electronic proxies).
shareholder voting generally) should be imminently realizable. Information costs should fall as corporate monitoring becomes both more easily, and more timely, accomplished. Corporations will have the opportunity, and potentially the obligation, to make more frequent disclosures of material information. Especially in light of the information technology used by most large corporations to track operational results, corporate web pages could be required to reflect up to date results, forecasts, and other firm-specific information. Such regulations would permit shareholders to detect corporate mismanagement more quickly and easily, and thus sooner realize the need to mobilize the shareholder vote.

Perhaps the greatest promise of the Internet is the potential for reducing disclosure costs. Electronic mail offers world-wide, instantaneous, nearly cost-free communications capabilities. While it would be difficult for a given dissident shareholder to identify the appropriate e-mail addresses of her fellow shareholders, proxy solicitation firms can be expected to fill the void by maintaining extensive databases containing that information. The required information could then easily, quickly, and cheaply be distributed directly to the shareholders for their consideration. Moreover, the use of

67. See, e.g., Oesterle, supra note 39, at 218–25 (discussing possibility of requiring continuous disclosure of corporate operating results). One of the first commentators to consider the disclosure possibilities offered by advances in electronic communications technology was Donald C. Langevoort, Information Technology and the Structure of Securities Regulation, 98 HARV. L. REV. 747 (1985). He more recently updated his proposal in Donald C. Langevoort, Toward More Effective Risk Disclosure for Technology-Enhanced Investing, 75 WASH. U. L.Q. 753, 770–76 (1997) (discussing version of continuous disclosure system). Of course, the extent to which corporations will be, as opposed to “could be,” required to provide continuous disclosure is largely dependent on SEC and state law disclosure requirements. Technological feasibility, however, is undoubtedly an important first step toward implementation of such requirements.

68. Proxy solicitation firms specialize in the distribution and collection of proxy solicitation forms. They are “middlemen” in the proxy solicitation process. Perhaps the largest such firm is ADP Investor Communication Services. See Cary I. Klafter, Using Web Sites for Investor Relations and Stockholder Meeting Materials, SEA GLASS-CLE 435, 437 (1998) (same); D. Craig Nordlund, Electronic Dissemination of Disclosure Documents, 1093 PLI/Corp. 39, 42 (1999) (describing ADP as the “key” proxy solicitation firm); Oesterle & Palmiter, supra note 5, at 510 (stating that over 70% of proxy solicitations are conducted by ADP). Both corporations and dissident shareholder groups typically utilize such firms as part of their proxy solicitation efforts. Because of the solicitation firms’ specialization in communications, one would imagine that they would be among the first to take advantage of the possibilities offered by the Internet.

69. Taking complete benefit of the communications possibilities presented by the Internet and electronic mail would require amendments to the SEC regulations. See FRIEDMAN, supra note 12, at 12-1 to 12-7 (discussing necessary changes). For instance, the regulations currently require that proxy statements be mailed to security holders, 17 C.F.R. § 240.14a-7 (2000). In contrast to e-mail,
web-pages may also allow for more complete disclosures of information relating to the voting options. For example, SEC regulations currently impose a 500-word limitation on descriptions of shareholder proposals included within proxy solicitation statements.\(^70\) Shareholders making proposals could include within this description the address of a website containing more detailed information regarding their proposal.\(^71\)

These possibilities have not gone unnoticed. Indeed, according to one commentator:

Dissidents and shareholder activists were among the first to recognize the potential of the Internet. Bennett LeBow and the Brooke Group took their proxy fight against RJR Nabisco into cyberspace, posting all their proxy materials and fight letters on the homepage of their proxy solicitation firm, Georgeson & Company Inc., and publishing the Web site address in all the printed materials and advertisements.\(^72\)

As technological and legal hurdles continue to disappear, one must assume that such tactics will increasingly become a part of everyday corporate life.

II. THE JUDICIAL TREATMENT OF VOTE-BUYING

The previous Section considered the corporate voting process, the problems endemic to it, and the potential for the Internet to address at least some of those concerns. If technological advances can serve to reinvigorate shareholder voting as a tool for corporate governance, however, it is important to revisit issues that may affect the voting process. One such issue is the appropriate treatment of vote-buying. This Section discusses vote-buying and the role it could play within the voting

\(^70\) 17 C.F.R. § 240.14a-8(d) (2000) ("The proposal, including any accompanying supporting statement, may not exceed 500 words.").


\(^72\) Wilcox, supra note 71, at 11.
Corporate Vote-Buying

process. It begins by defining vote-buying and distinguishing it from voting trusts and vote-pooling agreements—related, but distinct, methods of coordinating voting power. It then briefly describes the judicial treatment of vote-buying, including the historical antipathy to such arrangements and the modern shift toward more relaxed judicial scrutiny.

A. Vote-Buying Defined

Delaware courts have used the term “vote-buying” to refer broadly to any “voting agreement supported by consideration personal to the stockholder, whereby the stockholder divorces his discretionary voting power and votes as directed by the offeror.”73 In one case, a corporate agreement to extend a loan to a large shareholder in return for the shareholder’s support of a merger constituted vote-buying.74 Similarly, another court held that a guaranteed job with the corporation in exchange for voting in a particular manner constituted vote-buying.75 In short, a vote-buying agreement is one in which ownership (i.e., the right to claim a residual interest in the assets of the corporation) is separated from the shareholder’s only means of control (i.e., the right to vote on how those assets should be used) in exchange for consideration.

Of course, conceptually, vote-buying could take many forms. For instance, votes could be bought on a one-time basis. The shareholder would deliver a proxy giving the purchaser the right to vote on one issue, such as a given election of the board of directors.76 Alternatively, vote-buying could entail the permanent transfer of the right to vote the share.77

74. Schreiber, 447 A.2d at 23 (“It is clear that the loan constituted vote-buying as that term has been defined by the courts.”).
76. Indeed, SEC Rules provide that with regard to securities registered under the 1934 Act, proxies are limited in duration to the next annual meeting following the proxy solicitation. See 17 C.F.R. § 240.14a-4(d)(2) (2000).
77. There are limitations on the ability to accomplish long-term vote sales through use of proxies. Delaware law limits proxies to three years duration unless the proxy specifically states a longer duration. Del. Code Ann. tit. 8, § 212 (1991 & Supp. 2000). More importantly, if the share is registered under the 1934 Act, proxies are limited in duration to the next annual meeting following the proxy solicitation. See 17 C.F.R. § 240.14a-4(d)(2). This limitation on proxies does not prevent the use of other techniques to effect a
This would in effect create a derivative instrument, independent from the underlying share and encompassing only the right to vote the share of stock. Presumably a market could arise in which these voting rights were traded just like any other sort of instrument. In fact, some have suggested that there may be benefits to permitting just such a market. Both forms of vote-buying present interesting, albeit somewhat different, questions. Surprisingly, the law treats them the same.

---

78. Dual-class recapitalizations are a form of a permanent “sale” of votes. See generally André, supra note 4, at 620–23. In a dual-class recapitalization, a single class of stock is divided into two classes, one with superior dividend rights, and one with superior voting rights. For example, the “super-voting” class may have twenty votes per share, while the “super-dividend” class has only one vote per share, but the right to receive two times the per-share dividend awarded to shares in the super-voting class. In the dual-class recapitalization, each shareholder is given the right to trade each of his existing shares for either a “super-voting” share or a “super-dividend” share. This presents each shareholder with a classic example of the prisoner’s dilemma. Each shareholder would prefer to keep the voting power with non-management-aligned shareholders in order to permit effective monitoring. However, because no single shareholder can be certain that his election of super-voting stock will maintain outside control, the strictly dominant strategy for each shareholder is to elect the “super-dividend” shares. Management shareholders, because they face greatly reduced coordination problems, can elect the “super-voting” shares. Thus, through a dual-class recapitalization, the management shareholders can effectively deprive the outside (i.e., non-management) shareholders of their voting power permanently. Id. at 620. Because of the potential for abuse, in 1987, the SEC promulgated Rule 19c-4 preventing dual-class recapitalizations. 17 C.F.R. § 240.19c-4 (1990). In Business Roundtable v. S.E.C., 905 F.2d 406, 417 (D.C. Cir. 1990), the D.C. Circuit struck down the rule as an unconstitutional exercise of the SEC’s regulatory authority. As a practical matter, however, Rule 19c-4 is still effective with regard to issues traded on national exchanges. The New York Stock Exchange has made compliance with the text of the SEC rule a listing requirement for all companies traded on that exchange. See NYSE Listed Company Manual § 313, available at www.nyse.com/listed/listed.html (last visited July 31, 2001). The American Exchange and NASDAQ have similar rules in effect. See American Stock Exchange Listing Standard § 122, available at www.complianceintl.com/amex (last visited July 31, 2001); Nasdaq Rule 4351, available at www.nasdr.com (last visited July 31, 2001).

79. See Douglas H. Blair et al., Unbundling Voting Rights and Profit Claims of Common Shares, 97 J. Pol. Econ. 420, 421–22 (1989) (noting that Wall Street professionals were urging the creation of such a market and theorizing that such a market would make the market for corporate control more efficient).

80. Id.

B. Other Voting Arrangements

Vote-buying agreements are not the only means whereby the shareholder can separate ownership of the underlying share from the power to vote that share. While vote-buying itself has always been subject to governance by the common law,82 two other methods of divorcing ownership and control—voting trusts and voting agreements—are explicitly permitted by Delaware statute.83

In a voting trust, a group of shareholders transfers legal ownership of their stock to a trustee for the purpose of vesting in him the right to vote but retains beneficial ownership of the shares.84 Throughout the duration of the trust, the trustee has the right to vote the stock, but he maintains a fiduciary duty to the shareholders.85 Furthermore, the trust agreement may specify the manner in which the trustee must vote the shares.86

Voting agreements, by contrast, do not involve the transfer of shares to a trustee. The participants merely sign an agreement specifying that each of them will vote their shares as provided in the agreement.87 The agreement can either specify the actual vote (i.e., the shareholders can agree to vote for a specific candidate for the board), or it can simply specify a procedure for determining how to vote.88 Because of concerns about the separation of ownership and control, both voting trusts and

82. See discussion infra Sections I.C, I.D.
84. Id. § 218(a). Until recently, such trusts were limited to a term of ten years, but 1994 amendments to the Delaware corporate code abolished this limitation.
85. See 5 Fletcher Cyclopedia of Corporations § 2091.10 at 438 (1996) (“Voting trustees should be held to adhere to the usual fiduciary principles of a trust.”); James D. Cox et al., Corporations § 13.33, at 13.84 (1995 & Supp. 2001) (noting that trustees’ discretion to vote the shares is “subject to their fiduciary duties”). For cases discussing the existence of this duty, see Regnery v. Meyers, 679 N.E.2d 74, 79 (Ill. App. 1997) (holding that trustee has fiduciary duty to members of voting trust); Siegel v. Ribak, 249 N.Y. S.2d 903, 906 (N.Y. 1964) (same). At least one court has held that the parties to the trust can contractually narrow the scope of the fiduciary duty. Warchime v. Warchime, 761 A.2d 1138, 1141 (Pa. 2000).
87. Id. § 218(c).
88. Id. § 218(a).
voting agreements were originally limited to ten-year terms. A 1994 amendment to the Delaware code, however, abolished this limitation.

The main difference between voting trusts or voting agreements on the one hand, and vote-buying on the other, is the nature of the consideration employed. With regard to the former, the traditional form of consideration is the reciprocal pledge of the other parties involved to vote their shares in the prescribed manner. If cash or any other form of consideration personal to the shareholder were offered to induce a party to enter an otherwise legal voting agreement, the transaction would constitute vote-buying and should be analyzed as such.

Delaware courts have long considered voting trusts and voting agreements to be in derogation of the common law prohibition on interfering with voting rights, and thus have construed the statutory provisions permitting them narrowly. Historically, strict compliance with all of the statutory requirements for establishing a voting trust was a necessary prerequisite to judicial enforcement.

---


91. See, e.g., Ringling Bros.-Barnum & Bailey Combined Shows v. Ringling, 53 A.2d 441, 443 (Del. 1947) (describing voting agreement where consideration on the part of each party was promise to vote shares in specified manner); Pitman v. Lightfoot, 937 S.W.2d 496, 501–02 (Tex. App. 1996) (discussing various shareholders who contributed stock to voting trust to create control group). The principal purpose of voting trusts and agreements is to allow shareholders to exercise control over the corporate entity by pooling their voting power and voting it as a block. Ben Fixman v. Diversified Indus., Inc., 1 Del. J. Corp. L. 171, 178–79 (Del. Ch. 1975).

92. See Schreiber v. Carney, 447 A.2d 17, 23 (Del. Ch. 1982) (“Vote-buying, despite its negative connotation, is simply a voting agreement supported by consideration personal to the stockholder . . . .”). Including voting agreements supported by personal consideration within the category of “vote-buying” responds to one of Hasen’s criticisms of attempts to restrict vote-buying. In his article, Hasen argues that it makes little sense to regulate vote-buying because parties could easily evade limitations on it through use of voting trusts or voting agreements, and thus “a ban on explicit vote-buying will simply increase the transaction costs of engaging in vote-buying by requiring a more cumbersome method to reach the same result.” Hasen, supra note 1, at 1352–53. If, however, courts limit per se acceptance of voting trusts and voting agreements to those situations where no personal consideration is present, Hasen’s criticism seems misplaced. Parties could not engage in vote-buying simply by structuring the transaction as a voting agreement or voting trust.

93. See Abercrombie v. Davies, 130 A.2d 338, 344 (Del. 1957); see also Oceanic Exploration Co. v. Grynberg, 428 A.2d 1, 6–7 (Del. 1981) (discussing historical treatment of voting trusts).

94. See Abercrombie, 130 A.2d at 344.
Corporate Vote-Buying

performance of the voting agreements. Furthermore, the recent trend has been to relax the requirement of strict compliance. Thus, it seems clear that, at least in some instances, the unbundling of ownership and voting control is allowed. As discussed below, however, the nature of the consideration present in such arrangements—that is, a reciprocal voting pledge—prevents use of these techniques as a means to coercively transfer shareholder wealth. Accordingly, whatever we may think about these arrangements, they offer little guidance to the issue at hand.

C. The Historical Prohibition on Vote-Buying

While courts have grudgingly accepted the legislative mandate to enforce certain types of voting arrangements, the common law approach to vote-buying was a regime of per se illegality. Courts offered two principal reasons for this approach. First, vote-buying was seen as a corrupting influence on the fiduciary duty each shareholder in a corporation was thought to owe his or her fellow shareholders. Second, courts rejected vote-buying in the corporate context as analogous to vote-buying in the democratic process. Just as vote-buying would corrupt the democratic process, robbing the government of its legitimacy, so would it affect corporate activities and decisions. This Section examines these two arguments against vote-buying and concludes that while the articulated reasoning is superficially appealing, upon closer examination, neither basis provides much support for the historical judicial antipathy toward vote-buying in the corporate context.

95. See F. O’NEIL, O’NEIL’S CLOSE CORPORATIONS § 5.37 (1992) (collecting cases).
96. See Oceanic Exploration Co., 428 A.2d at 7; see also COX ET AL., supra note 85, § 13.33, at 13.85 (“The modern attitude is to uphold arrangements that do not comply with all the requirements of the state voting-trust statute as long as the arrangement does not otherwise violate public policy.”).
97. See infra note 210.
99. See infra Section II.C.1; see also André, supra note 4, at 543 n.27 (collecting cases).
100. See infra Section II.C.2; see also André, supra note 4, at 542 n.26 (collecting cases).
1. Vote-Buying and Fiduciary Duty

The fiduciary duty concept arose from the notion that each corporate shareholder owed every other shareholder the duty to vote using her own independent judgment. The stated basis for this obligation was the belief that “[t]he security of the small stockholders is found in the natural disposition of each stockholder to promote the best interests of all, in order to promote his individual interests.” Of course, this reasoning is more accurately stated in reverse: the “natural disposition” of each shareholder, assuming the shareholder acts rationally, is to promote his or her own individual interest. In the absence of vote-buying, the shareholder presumably could promote his or her individual interest only by maximizing the value of his or her shares, thus achieving the collective good of shareholder wealth maximization. In other words, an outright prohibition on vote-buying can be understood as an attempt to insure that shareholders’ interests remain closely aligned. It would be difficult, if not impossible, for one shareholder to benefit in his role as a shareholder without every other shareholder capturing the same benefit.

This judicial hostility toward vote-buying was merely an early example of the courts’ recognition of the problems presented by a divergence between ownership and control. A purchaser of corporate votes, especially if it was a person without an ownership interest in the company, could use his or her control to injure the corporate shareholders. Indeed, the only reason such a person would purchase votes would presumably be if he or she could obtain a return on the investment in vote-buying through exploitation of the corporate control. It was through exposing the other shareholders to this risk that a vote-selling shareholder violated his or her fiduciary duties.

   It is the policy of our law that ownership of stock shall control the property and the management of the corporation, and this cannot be accomplished, and this good policy is defeated, if stockholders are permitted to surrender all their discretion and will in the important matter of voting, and suffer themselves to be mere passive instruments in the hands of some agent, who has no interest in the stock, equitable or legal, and no interest in the general prosperity of the corporation.
   See also Cone, 21 A. at 849 (noting that the transfer of voting rights “is the more dangerous because the person intrusted with the power has no such inducement to promote the interests of the corporation as the stock-owner has.”); André, supra note 4, at 542 n.25.
Modern case law, however, rejects the notion that shareholders, at least minority shareholders, owe their fellow shareholder any fiduciary duty.103 Minority shareholders now have well-recognized rights to compete with the corporation in other business endeavors,104 to withhold corporate opportunities,105 and even to engage in business transactions that will result in financial detriment to the corporation’s other shareholders.106 In light of the courts’ rejection of any fiduciary duty in such contexts, it would be difficult to argue that there is some “fiduciary duty” that would prevent a shareholder from selling his vote.107


105. See Advanced Communications Design, Inc. v. Follett, 615 N.W.2d 285, 293–94 (Minn. 2000) (holding that even in close corporation, minority shareholder owed no fiduciary duty that would prevent him from soliciting corporate customers).

106. See generally Waters v. Double L, Inc., 769 P.2d 582, 583–84 (Idaho 1999) (stating that shareholder can vote on transaction in manner favorable to him, even if adverse to other shareholders); see also Kahn, 638 A.2d at 1113–14 (stating that fiduciary duty limited to controlling shareholder); Ivanhoe Partners, 535 A.2d at –1344 (same); In re Shoe-Town, Inc. Stockholders Litig., No. C.A. 9483, 1990 WL 13475 (Del.Ch. Feb. 12, 1990) (same).

107. While the rejection of “fiduciary duty” applies only to minority shareholders, it is doubtful that a majority shareholder would sell his votes. The majority shareholder has control, and thus is not subject to the collective action problems minority shareholders face in determining whether or not to sell their vote. See infra notes 205–13 and accompanying text. Interestingly, an argument could be made that the person purchasing the votes has become a “controlling shareholder,” and thus owes a fiduciary duty to the remaining shareholders. See Zirn v. VLI Corp., 621 A.2d 773, 778 (Del. 1993) (fiduciary duty as majority shareholder).
2. Analogies to Vote-Buying in the Democratic Process

The idea that vote-buying was illegal per se also had roots in an analogy to vote-buying in the political arena. The general purpose of the vote in both the commercial and the political forums is roughly analogous—to pick representative agents to manage an institution for the benefit of the voters, or in some cases to decide a specific issue. Courts and legislatures had uniformly rejected vote-buying in the political context as a perversion of the democratic process. The argument that, by analogy, vote-buying should be illegal in the corporate context had many supporters, both among courts and commentators.

More recently, commentators have begun to question whether this analogy is accurate. In a recent article, Richard Hasen discussed three potential rationales for the universal ban on vote-buying in political contests: equality, efficiency, and inalienability. The equality rationale recognizes that the poor are more likely to sell votes, and the wealthy are more likely to buy them, leading to a greater concentration of political power in the wealthy merely because of their wealth. The efficiency

108. In the case of corporations, shareholders elect the board of directors; in government, voters elect the officials.
109. Shareholders will sometimes vote on specific issues such as mergers or recapitalizations. In the democratic process the referendum performs a similar role.
110. All fifty states and the federal government have enacted statutes making it a crime to buy votes in government elections. See Hasen, supra note 1, at 1324 n.1 (collecting statutes). For examples of statutes banning vote-buying in political contests see ARIZ. REV. STAT. § 16-1014 (1998); MASS. GEN. LAWS ANN. ch. 56, § 32 (West 1990); N.M. STAT. ANN. § 1-20-11 (Michie 1995); OHIO REV. CODE ANN. § 3599.01 (Anderson 2000); WASH. REV. CODE ANN. § 29.85.060 (West 2000); WIS. STAT. ANN. § 12.11 (West 1996). More recently, commentators have begun to inquire whether there may be some benefits to permitting at least some forms of vote-buying in political contests. Saul Levmore, for instance, suggests vote-buying in such contests might be useful as a preference aggregation tool. Levmore, supra note 2, at 142–58. Richard Hasen, on the other hand, draws a distinction between what he refers to as “core vote-buying” (i.e., direct payment for votes) and “non-core vote-buying” (i.e., incentives such as campaign promises), and, while he generally condemns the use of the first in political contests, he is more ambivalent about the latter. Hasen, supra note 1, at 1370–71.
111. See, e.g., André, supra note 4, at 542 n.26 (collecting a list of cases); see also Peter N. Flocos, Toward a Liability Rule Approach to the “One Share, One Vote” Controversy: An Epitaph for the SEC’s Rule 19c-4?, 138 U. PA. L. REV. 1761, 1785–87 & nn.109–24 (1990) (discussing early academic support for the analogy).
114. Id. at 1329–31.
rationale suggests that vote-buyers might engage in rent-seeking behavior\textsuperscript{115} that diminishes net social wealth.\textsuperscript{116} The inalienability rationale turns on “a moral judgment that votes should not be salable,” which grows either out of the notion that voting is a group right belonging to the community as a whole and, accordingly, is not alienable by individual voters, or out of a broader anti-commodification norm that suggests that selling votes would do violence to our conception of what voting means.\textsuperscript{117} Hasen argues that none of these three rationales requires a ban on vote-buying in corporate contests. Equality is not an issue because of the broad, shared understanding that corporate law is not concerned with equality, but rather wealth maximization.\textsuperscript{118} Inalienability fails as a justification because there is simply no anti-commodification principle applicable to corporate law.\textsuperscript{119} Inefficiency, he suggests, fails to support a ban because courts can provide sufficient policing to weed out inefficient (i.e., non-wealth-maximizing) transfers.\textsuperscript{120}

While the efficiency point is potentially overstated,\textsuperscript{121} Hasen certainly seems correct that—as other commentators\textsuperscript{122} and modern case law\textsuperscript{123} have also recognized—there are significant differences between civic and corporate voting. Voting in political contests is a public, not a private right.\textsuperscript{124} It attaches, not as a result of contract law, but rather as a reflection of status. Furthermore, while a corporation’s existence, and the

\begin{itemize}
  \item \textsuperscript{115} “Rent-seeking” has been defined as “the attempt to obtain economic rents (i.e., rates of return on the use of an economic asset in excess of the market rate) through governmental intervention in the market. An example of rent-seeking is a firm’s attempt to secure government-granted monopolies.” Jonathan R. Macey, \textit{Transaction Costs and the Normative Elements of the Public Choice Model: An Application to Constitutional Theory}, 74 VA. L. REV. 471, 472 n.4 (1988).
  \item \textsuperscript{116} Hasen, \textit{supra} note 1, at 1331–35. In other words, the purchasers of the votes may use their voting power to secure politicians who would act favorably to the vote-purchasers at the expense of other constituents. For instance, a government contractor could use purchased votes to elect officials who would execute above-market government contracts with the contractor, or a special interest group could purchase votes to elect officials who would promote laws beneficial to that group.
  \item \textsuperscript{117} Id. at 1335–38.
  \item \textsuperscript{118} Hasen recognizes that his rejection of the equality argument turns on corporate law incorporating a strong shareholder wealth maximization norm. \textit{Id.} at 1353. If that norm is rejected, as some commentators suggest it should be, see \textit{supra} note 16, the equality argument may have greater applicability to vote-buying.
  \item \textsuperscript{119} Id.
  \item \textsuperscript{120} Id. at 1351–52.
  \item \textsuperscript{121} \textit{See infra} notes 221–74 and accompanying text.
  \item \textsuperscript{122} \textit{See Clark, supra} note 4, at 804–05; Fischel, \textit{supra} note 112, at 141; Flocos, \textit{supra} note 111, at 1784–90.
  \item \textsuperscript{124} \textit{See Clark, supra} note 4, at 804; Flocos \textit{supra} note 111, at 1785–93 & nn.109–49.
\end{itemize}
role of the shareholder’s vote therein, is predicated in large part on concepts of wealth maximization, government often concerns itself with allocative and redistributive issues.

Perhaps the single most important difference between the corporate and political voting regimes is the availability of a viable “exit” option. That is, share ownership is an entirely voluntary undertaking. If a shareholder disagrees with the course of action a corporation selects, the shareholder can terminate his or her ownership interest by selling the shares.125 Citizens, by contrast, do not have the same range of options open to them. Renouncing citizenship and acquiring citizenship in another country certainly entails more substantial costs than selling corporate shares on the market. The absence of a viable exit option makes strong protection of the “voice” option126 far more important in the political arena.127 Thus, whatever we conclude regarding vote-buying in the political sphere, it does little to inform our choice regarding vote-buying in the corporate world.

D. The Modern Treatment of Vote-Buying

The modern treatment of vote-buying in corporate law reflects an outright rejection of the earlier notions of per se illegality. The shift in the judicial attitudes toward vote-buying was announced in 1982 in *Schreiber v. Carney*.128 In *Schreiber*, a shareholder brought a derivative action on behalf of Texas International Airlines, challenging a loan made by Texas International to Jet Capital, a 35% shareholder in Texas International.129 The loan grew out of a restructuring involving a merger between Texas International and Texas Air Corporation.130 Under the corporate charter, the merger required a majority vote from each of a number of different classes of stock.131 Because Jet Capital owned a majority of the shares in one of the voting classes, it could maintain a

---

125. Note that the availability of “exit” is not a complete answer to shareholder concerns about manipulation of the voting process. To the extent that damage has been inflicted on a corporation as a result of the manipulation, the price a share will command is likely less than it would have been absent the wrongful conduct.

126. The voice option refers to the right to vote or, in other words, to make one’s voice heard.


128. 447 A.2d 17 (Del. Ch. 1982).

129. *Id.* at 18.

130. *Id.*

131. *Id.* at 19.
blocking position on the merger, even though the other shareholders overwhelmingly supported the merger.132

Jet Capital agreed that the merger was in the best interests of Texas International, but refused to support it, allegedly because of adverse income tax effects resulting from its ownership of a number of warrants.133 Although it could escape these negative tax consequences by exercising the warrants, it stated that it lacked sufficient cash to do so. Thus, it felt constrained to vote against the merger.134 In order to overcome this impasse, the board of Texas International agreed that Texas International should provide a loan to Jet Capital to fund its early exercise of the warrants.135 The loan was put to shareholder vote and overwhelmingly approved.136

At least one shareholder, Leonard Schreiber, however, did not approve of the transaction. He filed suit asserting that the transaction constituted vote-buying and, under well-established law, was illegal per se.137 The court agreed with Schreiber that the loan constituted vote-buying.138 According to the court, however, vote-buying was not per se illegal, but rather must be examined on a case by case basis.139 Here, the vote-buying passed judicial scrutiny.140

In order to reach this result, the court was forced to distinguish precedents that at least superficially appeared to hold vote-buying illegal per se.141 In doing so, the court reexamined the old cases, and found two propositions. First, vote-buying arrangements were illegal when they

132. Id.
133. Id.
134. Id.
135. Id. at 20.
136. Id.
137. Id.
138. Id. at 23. It is not entirely clear why this transaction involved vote-buying. After all, the corporation loaned the money to the shareholder, it didn’t merely give it away. It appears, however, that the terms of the loan were more favorable than the shareholder could have gotten elsewhere. According to the court, “borrowing money at the prevailing interest rates . . . was deemed too expensive by the management of Jet Capital.” Id. at 19. Presumably the 5% rate offered by Texas Air for the loan, id., was lower than market. A sub-market interest rate loan would constitute consideration.
139. Id. at 25.
140. Id.
acted “to defraud” or “disenfranchise” stockholders.” The Schreiber court characterized Macht v. Merchants Mortgage & Credit Co., for instance, one of the principal cases prohibiting vote-buying, as involving “a series of criminally manipulative transactions instigated by a designing director in order to wrest control of the corporation from the other stockholders for fraudulent motives.” According to the Schreiber court, this proposition retained its vitality, but simply was not implicated by the facts presented. The second proposition was that vote-buying was illegal per se because it violated public policy. The court, citing evolving ideas of “public policy” and changed attitudes toward the function of shareholders in corporate governance, explicitly rejected this as a basis for invalidating vote-buying agreements.

The court was quick to point out, however, that its rejection of per se invalidity for vote-buying agreements should not be taken as a sign that courts would openly embrace these agreements. The court noted that vote-buying agreements had a potential for abuse. Thus, although vote-buying agreements would no longer be void, they would still be voidable and subject to a test of intrinsic fairness. The agreement in question here clearly passed this test, according to the court, as the disinterested shareholders had overwhelmingly approved it following full disclosure.

To summarize, the state of the law after Schreiber required Delaware courts to undertake a three-part test. First, determine if the agreement in question constitutes vote-buying. That is, does it involve the transfer of voting power in exchange for consideration personal to the shareholder? If it does, then determine whether it acts to defraud or disenfranchise the other shareholders. Finally, even if there is no fraud or disenfranchisement, treat the agreement as a voidable transaction subject to the test of intrinsic fairness.

---

142. Schreiber, 447 A.2d at 23.
143. 194 A. 19 (Del. Ch. 1937). The Schreiber court described Macht as the leading case for the proposition that vote-buying was illegal per se. Schreiber, 747 A.2d at 23.
144. Schreiber, 447 A.2d at 23 (emphasis added).
145. Id. at 26.
146. Id. at 25.
147. Id. at 26.
148. Id.
149. Id.
150. Id.
Delaware courts have addressed allegations of vote-buying a number of times since the decision in Schreiber. In each of these cases, the court has whole-heartedly embraced the Schreiber court’s rejection of per se illegality. In fact, as described below, some of these decisions go even farther than Schreiber. Yet, in none of these cases was the voting agreement struck down as illegal.

On two of these occasions, the court determined that the agreement in question did not constitute vote-buying. In doing so, the courts narrowed the expansive definition of vote-buying adopted by Schreiber. In Henley Group, Inc. v. Sante Fe Southern Pacific Corp., for instance, the court held that Schreiber required as an “essential element of a vote buying agreement . . . that ‘. . . the stockholder divorces his discretionary voting power and votes as directed by the offeror.’” According to the Henley court, the fact that the party in the present case was not legally obligated to vote in a certain way precluded any finding that vote-buying had occurred. Thus, under Henley, a threshold requirement for vote-buying is a legally binding obligation to vote in a certain manner.

In the other case, Rainbow Navigation, Inc. v. Yonge, the court instead focused on the issue of consideration. In that case, Van Reesema, one of three shareholders in a closely held corporation, switched loyalties. As a result, the directors of the company were removed from office and the corporate bylaws were amended. The displaced directors brought suit, alleging that the vote that removed them was illegal because Van Reesema’s vote had been bought. According to the directors, Van Reesema switched only after being offered a “consulting agreement” and the release of certain legal claims against him. According to the judge, however, this arrangement did not constitute vote-buying for two reasons. First, Van Reesema’s expectation of a consulting agreement was not legally enforceable at the time of the vote. Secondly, the

---

152. Id. at *7 (quoting Schreiber, 447 A.2d at 23).
153. Id.
155. Id. at *1.
156. Id.
157. Id. at *2.
158. Id. at *5.
159. Id. at *7.
consideration was not “a principal motivating force” in Van Reesema’s decision to vote the way that he did.\footnote{160}

Taken together, these two cases narrow the broad definition of vote-buying \emph{Schreiber} offered. To constitute vote-buying, three things now must occur. First, the “seller” must have a legally enforceable right to the consideration offered. Second, that consideration must have been a “principal motivating force” in the shareholder’s vote. Finally, the seller must have a legal obligation to vote in a certain manner. Absent these three elements, a vote-buying charge will not stand.

Two other cases have addressed the second prong of the \emph{Schreiber} test, whether the alleged vote-buying acted to “defraud or disenfranchise” the other shareholders. First, in \emph{Kass v. Eastern Air Lines, Inc.},\footnote{161} \emph{Schreiber} was extended to a case involving bondholder consents. Eastern was attempting a merger with Texas Air. As part of the deal, Eastern desired to pay a $1.75 dividend to its shareholders.\footnote{162} Payment of such a dividend, however, would violate certain covenants contained in various outstanding convertible debentures.\footnote{163} These covenants could be altered only with the consent of two-thirds of the holders of these instruments.\footnote{164} In an attempt to secure these consents, the board of Eastern offered to pay each debenture holder who consented either $35 in cash or $125 in free Eastern ticket vouchers for each $1000 in face amount of debentures for which consent was given.\footnote{165}

Not surprisingly, debenture holders who were opposed to relaxing the covenants challenged this as illegal vote-buying.\footnote{166} In analyzing this claim, the court, much like the \emph{Schreiber} court, quickly rejected the argument that analogized corporate vote-buying to vote-buying in political contests.\footnote{167} According to the court, the political system involved important non-commercial considerations, and thus a trade of votes for money would be corrupting.\footnote{168} Commercial transactions, on the other hand, involved solely commercial values, and thus, “commerce in votes”

\begin{thebibliography}{9}
\bibitem{160} \textit{Id.} at *10.
\bibitem{161} 1986 WL 13008 (Del. Ch. Nov. 14, 1986).
\bibitem{162} \textit{Id.} at *1.
\bibitem{163} \textit{Id.}
\bibitem{164} \textit{Id.}
\bibitem{165} \textit{Id.}
\bibitem{166} \textit{Id.}
\bibitem{167} \textit{Id.} at *2--*3.
\bibitem{168} \textit{Id.}
\end{thebibliography}
was much more acceptable.\textsuperscript{169} The court went so far as to say that in the commercial setting, “to conclude that the offering of money in exchange for consent (vote) is necessarily corrupt or a per se violation of public policy strikes [the Court] as quaint.”\textsuperscript{170}

The court thus rejected per se illegality and chose to adopt the \textit{Schreiber} analysis. Because the consideration was openly offered to each of the bondholders on the same terms, the court concluded that it was not fraudulent and could not act to disenfranchise any of the voters.\textsuperscript{171} Interestingly, it never addressed the third prong of the \textit{Schreiber} test. It did not treat the offer as a voidable transaction subject to approval by the disinterested voters.\textsuperscript{172}

Most recently, a Delaware court addressed this issue in \textit{In re IXC Communications, Inc. Shareholders Litigation}.\textsuperscript{173} There, IXC, a telecommunications company, placed itself “in play” by hiring an investment bank to develop merger options.\textsuperscript{174} Ultimately, Cincinnati Bell, Inc. (CBI) forwarded a proposal to merge the two companies by having the IXC shareholders exchange their shares for shares in CBI.\textsuperscript{175} During the course of negotiations that led to the merger proposal, CBI met with General Electric Pension Trust (GEPT), IXC’s largest shareholder with nearly a 40\% stake.\textsuperscript{176} CBI and GEPT reached a “side-deal” regarding the proposed merger.\textsuperscript{177} Pursuant to that deal, CBI purchased half of GEPT’s IXC holdings for $50 per share.\textsuperscript{178} In exchange, GEPT agreed to vote its remaining shares in favor of the merger.\textsuperscript{179}

After CBI presented its merger proposal to the board and the board approved, dissident shareholders attempted to enjoin the shareholder vote

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{169} \textit{Id.} at *3.
\item \textsuperscript{170} \textit{Id.}
\item \textsuperscript{171} \textit{Id.} at *4. This conclusion overlooks the coercive aspects of group action in light of a prisoner’s dilemma situation. \textit{See supra} notes 98–107 and accompanying text.
\item \textsuperscript{172} Of course, when the offer is open to all, it would be difficult to find disinterested voters. However, \textit{Schreiber} then requires the court to perform a judicial inquiry into the intrinsic fairness of the transaction, an inquiry the court in \textit{Kass} failed to undertake.
\item \textsuperscript{173} No. C.A. 17324, C.A. 17334, 1999 WL 1009174 (Del. Ch. Oct. 27, 1999).
\item \textsuperscript{174} \textit{Id.} at *1.
\item \textsuperscript{175} \textit{Id.} at *2.
\item \textsuperscript{176} \textit{Id.}
\item \textsuperscript{177} \textit{Id.} at *3.
\item \textsuperscript{178} \textit{Id.}
\item \textsuperscript{179} \textit{Id.}
\end{itemize}
\end{footnotesize}
necessary to complete the merger.\textsuperscript{180} The shareholders argued, inter alia, that the CBI-GEPT deal constituted impermissible vote-buying.\textsuperscript{181} In addressing this charge, the court concluded that the arrangement could well be considered vote-buying.\textsuperscript{182} Under \textit{Schreiber}, however, the court continued, such arrangements were void only if they acted to defraud or disenfranchise the remaining shareholders.\textsuperscript{183} Here, according to the court, the plaintiffs admitted there was no fraud, so the only question was disenfranchisement.\textsuperscript{184} With regard to that issue, the court concluded that, because GEPT held only a 40\% interest (prior to its sale of 50\% of its shares to CBI), the agreement could not act to disenfranchise the remaining independent shareholders.\textsuperscript{185} Finally, in response to the plaintiffs’ argument that the precommitment of the 40\% interest in support of the merger would “almost lock up” the vote, leaving the remaining shareholders with “scant power” to oppose the deal, the court said merely: “‘Almost locked up’ does not mean ‘locked up,’ and ‘scant power’ may mean less power, but it decidedly does not mean ‘no power.’”\textsuperscript{186}

Under \textit{IXC}, then, it appears that so long as less than 50\% of the vote is in the hands of the vote-purchaser prior to the election, the transaction is not subject to challenge on disenfranchisement grounds. Moreover, as in \textit{Kass}, the court in \textit{IXC} failed to address the third prong of \textit{Schreiber}. That is, it did not subject the vote-buying to approval by the disinterested shareholders or a test of “intrinsic fairness.” Whether this indicates that intrinsic fairness is no longer part of the vote-buying analysis is unclear.\textsuperscript{187} What is clear, however, is that under modern case law, vote-buying agreements receive far more generous treatment from the courts than they have in the past.

\begin{itemize}
\item 180. \textit{Id.} at *1.
\item 181. \textit{Id.} at *7–*9.
\item 182. \textit{Id.} at *3.
\item 183. \textit{Id.} at *8.
\item 184. \textit{Id.}
\item 185. \textit{Id.} at *9.
\item 186. \textit{Id.}
\item 187. \textit{See generally} Royce de R. Barondes, \textit{An Economic Analysis of the Potential for Coercion in Consent Solicitations for Bonds}, 63 \textit{Fordham L. Rev.} 749 (1994) (suggesting based on game theoretic model of bond solicitations that disclosure, rather than regulation, is sufficient to protect bondholders from coercion, thus obviating need for fairness inquiry).
\end{itemize}
III. VOTE-BUYING AND SHAREHOLDER WEALTH EFFECTS

As discussed above, the historical prohibition on vote-buying was premised on two separate arguments: a moral legitimacy argument that condemned vote-buying as a matter of principle,\(^\text{188}\) and a shareholder wealth effects argument that decried vote-buying based on its presumed harm to other shareholders’ economic interests.\(^\text{189}\) The modern treatment of vote-buying rejects as inappropriate any consideration of the former, and focuses exclusively on the latter. With regard to the wealth-effects argument, it concludes that while shareholder wealth effects are the appropriate standard against which to judge vote-buying, those considerations do not support the earlier per se prohibition on vote-buying.\(^\text{190}\) Instead of rejecting vote-buying transactions wholesale, the Schreiber court’s three-prong approach attempts to segregate vote-buying transactions on a case-by-case basis. The “defraud or disenfranchise” prong, operating in conjunction with the requirement of “intrinsic fairness” (to the extent that prong remains), appears to direct courts to undertake this separation based on a determination of the shareholder wealth effects resulting from the particular vote-buying transaction at issue.\(^\text{191}\)

The shift to a paradigm in which the regulation of vote-buying centers exclusively on a shareholder wealth maximization norm is not, in and of itself, troublesome. Shareholder wealth effects are a core concern throughout corporate law,\(^\text{192}\) and there is no immediately apparent reason why those effects should not also be a core concern with regard to vote-buying. Nevertheless, accepting shareholder wealth maximization principles as an appropriate basis for regulating vote-buying does not

\(^{188}\) See supra Section II.C.2.
\(^{189}\) See supra note 102 and accompanying text.
\(^{190}\) See supra Section II.D.
\(^{191}\) If this is indeed what the Schreiber rule intends courts to capture, certain post-Schreiber interpretations of the test are questionable at best. For instance, IXC’s focus on a “true majority” standard to determine whether the vote-buying transaction “disenfranchised” the remaining shareholders seems like a purely mechanistic test wholly-divorced from the underlying wealth effects on the company shareholders. See supra notes 178–79 and accompanying text.
\(^{192}\) Edward S. Adams & John H. Matheson, A Statutory Model for Corporate Constituency Concerns, 49 Emory L.J. 1085, 1094–95 (2000). While these authors argue that directors should take account of broader constituencies, they acknowledge that “[t]he traditional view of corporate law commands directors to make decisions that will maximize shareholder wealth.” Id. at 1094. This view of corporate law, referred to as the shareholder primacy view, is a central tenet of traditional corporate law. See supra note 16 and sources cited therein.
necessarily imply that the shift away from a per se rule to a case-by-case approach furthers that goal.

To determine whether or not the shift is beneficial requires further consideration of two things; first, the likely uses of vote-buying in a permissive regime, and, second, the ability of courts to separate beneficial and detrimental vote-buying transactions. If vote-buying transactions are never beneficial to shareholders (or, if the potential costs associated with such transactions as a whole far outweigh the benefits), it may be that a per se rejection of such transactions better serves shareholder interests. Even if there are potentially beneficial vote-buying transactions, if courts are poorly equipped to separate them from detrimental transactions, a per se rule may still have advantages, notwithstanding the potential that certain pro-shareholder transactions would be blocked.

This Section discusses various potential uses for vote-buying. It first considers whether permitting vote-buying might allow shareholders to use such transactions to overcome the collective action problems that interfere with the their ability to use voting as an effective means of corporate control, but concludes that this is unlikely. It then discusses the possibility that vote-buying could be employed to facilitate corporate looting or effect coercive wealth transfers between various investor classes. It concludes that both of these represent very real threats, that the potential for such abuses must figure prominently in determining the appropriate regulatory approach, and that, to date, courts have proven incapable of identifying instances in which these harms have occurred.

Finally, this Section discusses what some commentators have advanced as the principal pro-shareholder justification for vote-buying, the possibility that potential acquirers could employ such transactions to overcome certain types of takeover defenses (in particular poison pills) through which management has attempted to insulate itself from the market for corporate control. While in these limited circumstances vote-buying may indeed prove beneficial to shareholders as a group, this Section discusses whether these same benefits could be achieved through the related idea of turnout payments—that is, paying a shareholder to vote, independent of the voting option the shareholder selects. It argues that limiting permissible “vote-buying” to such turnout payments would benefit shareholders by dismantling management’s takeover protection, without exposing shareholders to the potential for abuse present in traditional vote-buying.
A. Vote-Buying as a Response to the Collective Action Problems of Shareholder Voting

As described above, four principal obstacles confront shareholders in developing and using information in corporation governance. First, any shareholder owning few shares will not capture a sufficient amount of gain on any corporate vote to make any but the smallest investment in information worthwhile. \(^{193}\) Second, even large shareholders, who might otherwise have sufficient economic incentives to pursue voting information, may fall prey to the free-rider problem. \(^{194}\) Third, in a voting contest, there are no gains to be made from private information. \(^{195}\) Fourth, coordination problems may prevent an efficient use of any investments in voting information. \(^{196}\)

One possible shareholder response to these difficulties would be use of a vote-pooling arrangement. This would be very similar to the voting trust described earlier. Each shareholder would transfer his or her proxies to a common pool. These proxies would then be voted as a block. In this way, the information costs could be allocated to the members of the pooling arrangement based on the amount of their investment. Because all shareholders would be sharing the costs, there would be no free-rider problem. Finally, there would be no duplicative research costs because the individual shareholders would not be developing the information on their own. The manager of the common pool would have a fiduciary duty to vote the shares in such a manner as to maximize shareholder wealth. \(^{197}\)

Of course, an actual transfer of proxies to the common pool would not be necessary. Each shareholder could simply be a member of an information-pooling cooperative. The members would pay dues based on their share ownership and that money would be used to monitor corporate performance and research voting issues. After all, assuming the pertinent information yields a clear choice on a voting issue, there is no

---

193. See supra Section I.B.1.
194. See supra notes 37–38 and accompanying text.
195. See supra notes 39–44 and accompanying text.
196. See supra Section I.B.3.
197. If the arrangement were a voting trust, the fiduciary duty would arise as a matter of course. See supra note 85. If the shareholders used some other contractual arrangement to achieve the pooling, the fiduciary duty of the person voting the shares could be created and specified in the contract creating the arrangement. See RESTATEMENT (SECOND) OF AGENCY § 376 (1958) (stating that agent’s fiduciary duties specified by contract creating agency).
need for a single trustee to control the actual proxies—the individual shareholders would all presumably vote in the appropriate way.\textsuperscript{198}

If vote-buying were necessary, or even helpful, to such shareholder pooling arrangements, the potential shareholder benefits might justify judicial acceptance of it. It is doubtful, however, that a permissive approach to vote-buying would facilitate such arrangements, because it is unlikely that votes would be “bought.” Rather, shareholders would “buy” memberships in the voting pools or trusts. By sharing the information costs, the shareholders, as equity participants in the corporate endeavor, would collectively benefit. While this sounds like a rational solution to the collective action problem, there are, as a practical matter, very few voting trusts in existence.\textsuperscript{199} Furthermore, those that are in use operate almost exclusively in closely held private corporations.\textsuperscript{200} At first glance this is somewhat anomalous because it would seem that the coordination and free-rider problems would be smaller with respect to close corporations for two reasons.\textsuperscript{201} First, the smaller number of shareholders makes coordination, in general, much easier. Second, closely held corporations tend to be held within families or small groups of close friends. The extra-corporate ties that bind these groups should help to ameliorate any free-rider concerns through the use of social sanctions against the free-rider.

Whatever the benefit of vote-pooling in the close corporation, the question remains—why aren’t there more such arrangements in large public corporations? There are at least three possible answers to that question. First, it may be that voting trusts do not really overcome the

\begin{itemize}
\item \textsuperscript{198} The assertion that they would vote in the same manner assumes (1) that the shareholders share the same level of risk aversion (probably a safe bet in a publicly held corporation where a majority of the shareholders are likely diversified), (2) that no subgroup of shareholders stands to benefit separately from the transaction underlying the vote, and (3) that there is clearly a “right” way to vote from a shareholder wealth maximization perspective.
\item \textsuperscript{199} One commentator, for example, refers to the “absence of voting trusts” and offers reasons to explain it. Barry E. Adler, \textit{Politics and Virtual Owners of the Corporation}, 82 Va. L. Rev. 1347, 1367 (1996). Other commentators have suggested certain disadvantages that have led to “limited use of voting trusts.” William S. Hochstetler & Mark D. Svejda, \textit{Statutory Needs of Close Corporations—An Empirical Study: Special Close Corporation Legislation or Flexible General Corporation Law?}, 10 J. Corp. L. 849, 948, 1011 (1985).
\item \textsuperscript{200} \textit{ECONOMIC STRUCTURE}, supra note 4, at 65.
\item \textsuperscript{201} Of course, coordination and free-rider problems are not the only governance difficulties a corporation might face. Other governance problems may be exacerbated in closely held corporations. See \textit{infra} notes 275-82 and accompanying text; \textit{see also} Clark, supra note 4, at 802 (“Voting trusts and vote pooling agreements are basically devices created to solve the peculiar and troublesome collective action problems of closely held corporations, such as the difficulty of achieving cooperation and avoiding voting deadlocks.”).
\end{itemize}
free-rider problem. Shareholders, after all, cannot be forced to join a voting trust, they must do so voluntarily. With respect to the vote-pooling or voting trust arrangements described above, there would be little incentive to join. Any single shareholder would want others to join the pool, but he would prefer to remain outside it himself. The information costs are borne only by those who are members of the pool. The information developed, however, works to the benefit of both those in the pool and those who are not. This creates a clear incentive against membership.

A second possibility is that membership in a voting trust or similar arrangement would hinder (or even prevent) transfer of the shares. All shares in the trust are legally owned by the trustee. Even if the beneficial owner could transfer her beneficial interest, legal ownership would presumably remain with the trustee. That is, even upon transfer, the shares would remain subject to the voting disability imposed by the trust.

A third possibility is that there are a number of arrangements that approximate voting trusts, but they are not recognized as such. For example, in a mutual fund, the members’ individual funds are pooled and then invested by the fund manager. The fund manager controls the votes of any stocks purchased by the fund. In return for the investment services that the manager provides, he receives a management fee. One of these “services” is usually the voting of the shares. This is, in effect, exactly the pooling arrangement suggested earlier. The “actual shareholders” pool their shares and then pay a third party to vote the shares appropriately. Furthermore, the various funds are rated on their ability to provide shareholders high returns. Thus, to the extent that shareholder


203. A related reason for the lack of voting trusts may be that the trustee is subject to capture. That is, the concentration of voting rights within the trustee means that, if management can capture the trustee, management is entirely insulated from shareholder discipline. See Adler, supra note 199, at 1366–67. Concentrating the voting rights in a trust potentially facilitates such capture to the detriment of shareholders. Of course, the trustee’s fiduciary duty may impose at least some limitations on the extent to which he or she can cooperate with management to the detriment of shareholders.

204. That is, the mutual fund members who are the “beneficial owners” of any profits received.

205. For an example of such a rating service, see the Morningstar rating system, available at www.morningstar.com (last visited July 31, 2001).
voting has an effect on corporate wealth, the fund managers have an incentive to invest the efficient amount in acquiring information and voting the shares.206

Whether or not mutual funds act as collective voting pools,207 it is clear that permitting vote-buying will not, as a general matter, serve as a means of overcoming the collective action problems associated with voting. If shareholders were attempting to use vote-buying for those purposes, the votes should have a negative, not a positive, price. Thus, this understanding of vote-buying simply does not provide an explanation for cases in which charges of vote-buying have arisen (all of which have involved positive prices for votes), nor does it provide guidance as to how courts should handle such cases.

B. Vote-Buying and Corporate Looting

That vote-buying does not facilitate shareholder attempts to overcome the information and coordination problems associated with voting does not, by itself, answer the question of whether vote-buying should be permissible. There may be other “pro-shareholder” results associated with judicial leniency toward vote-buying. In order to further explore this question, assume for the moment that vote-buying were legal. Two important issues immediately come to mind: the identity of the purchasers and the pricing mechanism in the market. An analysis of these considerations suggests two things. First, vote-buying would permit purchasers to exploit shareholder coordination problems and, as a result, shareholders would sell their votes both too often and too cheaply. Second, that purchasers are seeking votes rather than shares suggests the purchasers do not plan to use the resulting voting power to enhance share values.

206. This does not imply that fund managers will be active voters, only that they will be efficient. It may be that the efficient voter does not invest in voting, preferring instead to exercise his or her exit option (i.e., sell the shares). See generally ECONOMIC STRUCTURE, supra note 4, at 88–89.

207. There are limitations on a mutual fund’s ability to do so. In particular, various legal restrictions act to keep the percentage of stock a mutual fund owns in a single company relatively low. For an in-depth discussion of these restrictions, see MARK J. ROE, STRONG MANAGERS, WEAK OWNERS 102–23 (1994). For example, under the Investment Company Act of 1940, 15 U.S.C. § 80a (1994 & Supp. 1999), a mutual fund cannot advertise itself as “diversified” if it owns more than 10% of the stock of a single company, and, if it owns more than 5% of the shares of a company, the fund is treated as an affiliate and an underwriter for securities law purposes. Id. at §§ 80a-2(a), -5(b)(1). Moreover, if the fund does not meet the requirements for diversification, adverse tax consequences follow. ROE, supra, at 106–08.
1. The Price of Votes and the Prisoner’s Dilemma

It is likely that shareholder votes would command a very small price. According to one commentator, in fact, any non-zero price should be sufficient to purchase a vote.\textsuperscript{208} This should not be read as implying that shareholders do not value their vote; studies indicate that non-voting shares generally trade at a 2–4% discount from otherwise identical shares that include voting rights.\textsuperscript{209} Rather, the willingness of voters to accept low prices for their votes can best be explained by reference to collective action problems. Assume that the issue on which the purchaser would like to buy the vote requires a 51% majority of the shares to pass. Any small shareholder knows that his or her vote is unlikely to matter in the election. Thus, the shareholder has little reason not to sell his or her vote. Further, the shareholder realizes that if he or she fails to sell the vote, and the purchaser acquires them elsewhere, the shareholder’s vote will be meaningless but at the same time he or she will have foregone the offered consideration.\textsuperscript{210} Thus, the shareholder is caught in the classic prisoner’s dilemma.\textsuperscript{211}

\textsuperscript{208} Sanford J. Grossman & Oliver D. Hart, \textit{One Share-One Vote and the Market for Corporate Control}, 20 J. FIN. ECON. 175, 177 (1988); see also Easterbrook & Fischel, supra note 5, at 411.


\textsuperscript{210} Note that this is not the case where the only consideration is a reciprocal pledge to vote in the same manner. There, the “consideration” has no value independent of the voting contest, and thus the offeree does not forego anything of value by declining the offer. Thus, traditional voting trusts or vote pooling agreements do not have the coercive effect of vote-buying arrangements.

\textsuperscript{211} Interestingly, the most recent article in favor of vote-buying assumed that the tender offer provisions of the Williams Act would apply to vote-buying. André, supra note 4, at 589 & nn.224–28. The Williams Act, which regulates tender offers, is contained in §§ 13(d)–(e) and 14(d)–(f) of the 1934 Act, 15 U.S.C. §§ 78m(d)–(e) and 78n(d)–(f) (1994), and the regulations thereunder, 17 C.F.R. §§ 240.13d-1 to 13e-4, 240.14a-1 to .14f-1 (2000). It requires that a person making a tender offer for all or some percentage of a company’s shares make it on equal terms to all shareholders. 15 U.S.C. § 78n(d)(7). If the shareholders tender “too many” shares (i.e., more than the stated goal of the tender offer), each of the tendering shareholders has the right to participate in the tender offer in proportion to the shares he tendered. \textit{Id.} § 78n(d)(6). The Williams Act was designed to ameliorate the coercive nature of a tender offer. (A tender offer involves a prisoner’s dilemma because non-tendering shareholders may receive less for their shares than the tender price.) If the Williams Act does indeed apply to an offer to buy votes rather than shares, it may help to nullify the coercive nature of a vote-buying transaction. It is not at all clear, however, that the Williams Act would apply to transactions involving the purchase of votes. Furthermore, in the context of votes, the protection against the prisoner’s dilemma is not as effective. Unlike shares, votes not tendered are completely worthless if a sufficient number are tendered. That is, non-tendering voters have no right of
Of course, while this is true for situations in which share ownership is widely dispersed among numerous small shareholders, it does not hold true when there are concentrated blocks. Imagine for instance that a shareholder has 50% plus one of the shares of voting stock. That shareholder would presumably not be willing to sell even a single vote for a small price.

This suggests the following about vote-buying: to the extent that shareholders participate both as buyers and sellers, permitting vote-buying would tend to concentrate voting power in the hands of the largest shareholders. Small shareholders will in general agree to accept small prices for their votes. Similarly, small shareholders would not be willing to pay much for the votes of others. After all, if a shareholder owns 1% of the outstanding stock, purchasing an additional 1% of the votes would not be likely to make that shareholder’s vote dispositive.

Large shareholders, on the other hand, would have a greater incentive to purchase votes. Assume there are one hundred shares of voting stock in a corporation, and the largest shareholder has forty-nine of them. The shareholder would presumably be willing to pay some amount to purchase the additional two votes necessary to gain control of the corporation, and, because of the prisoner’s dilemma problem noted earlier, the shareholder should have no difficulty doing just that. In general, the larger the size of a shareholder’s current holdings, the more the shareholder would be willing to pay for votes (up to the point where he or she obtains a sufficient number of votes to ensure the desired level of control).

In sum, large shareholders may be able to avoid the coercive aspects of a vote purchase offer. In publicly-held corporations with widely-dispersed share ownership, however, coordination difficulties among the shareholders will permit vote purchases on the cheap.

appraisal. Cf., e.g., DEL. CODE ANN. tit. 8, § 262(a) (1996) (providing that after merger dissenting voters have right of appraisal for their shares).

212. Each of the remaining shareholders must, in effect, compete with the others to sell two votes. Since the value of the remaining votes after the sale is zero, the competitive price should approach zero.

213. Assume a corporation that has 100 issued and outstanding shares. If a shareholder has one share, purchasing an additional vote provides her with 1/50th of the remaining votes needed. If, on the other hand, the shareholder has thirty-five shares, an additional vote represents 1/15th of the remaining voting power required. Thus, the latter shareholder should be willing to pay more. Of course, if the shareholder already holds an outcome-determinative number of votes, he or she would presumably be unwilling to pay anything for the remaining votes.
2. **Who Buys Votes?**

If a potential vote purchaser believes that a certain outcome on a shareholder voting issue will result in increased corporate wealth, the purchaser’s best option is to purchase the shares, not the votes. By purchasing the shares, the shareholder will acquire the entire increase in wealth associated with the purchased shares. If the shareholder purchases only the vote, he or she will merely capture the gain in wealth associated with those shares previously owned. Thus, if a potential purchaser has the option to purchase either the shares or the votes and chooses the latter, it suggests that the purchaser is not intending to use the vote for reasons that would increase corporate wealth, but rather to “loot” the corporation.

Who are likely candidates for looting through vote acquisition? As described previously, one class of vote purchasers could be the large shareholder. It is not clear, however, that this group is likely to use vote buying to engage in looting. Large shareholders extracting wealth from a corporation also hurt themselves in their capacity as shareholders. For example, assume a 40% shareholder extracts $1 from the corporation through his or her purchase of voting control. Even if the extraction is costless, the shareholder’s net benefit is only sixty cents. Forty cents of the dollar already “belonged” to the shareholder. From this sixty cents, the costs of purchasing the votes must also be subtracted. Thus, in general, the greater the percentage stake a shareholder owns, the less the shareholder’s incentive to engage in vote-buying to loot the company.

---

214. “Looting” refers to the improper extraction of wealth from the corporate form, for instance through excessive salaries, sale of corporate outputs or assets at an artificially low price, or purchase of corporate inputs at an artificially high price. See ECONOMIC STRUCTURE, supra note 4, at 129–31; Clark, supra note 4, at 795.

215. There may be some costs, such as “cover up” costs, associated with the transfer. The magnitude of those costs will depend both on the difficulty of concealing the transfer, as well as the legal rules surrounding such transfers.

216. The costs referred to here include both the purchase price for the votes as well as the costs, such as communication costs, associated with making the offer and undertaking the purchase.

217. Indeed, in a permissive vote-buying regime, the large shareholder may often be forced into the role of “reluctant purchaser.” That is, if a potential looter makes an offer for shareholder votes, the large shareholder may find it necessary to bid against the looter in order to protect the value of the shareholder’s stake in the company. While at least one commentator has suggested that the potential for competing bids serves to protect shareholder vote sellers from the coercive nature of the prisoner’s dilemma, see Levmore, supra note 2, at 139, as this example indicates, the possibility of such a bidding war may result in a coercive wealth transfer away from large shareholders.
Even so, a permissive attitude toward vote-buying would provide at least some additional possibilities for looting by large shareholders. While the large shareholder has, in some sense, a pre-existing “claim” to some portion of the corporate dollar, the shareholder would no doubt prefer the entirety to that portion, thereby providing the shareholder an incentive for vote-purchases to solidify control. In a regime that banned vote-buying, by contrast, the large shareholder seeking to obtain greater control to facilitate looting would need to acquire additional shares in order to do so. Requiring the “looting shareholder” to purchase shares rather than votes would likely diminish such looting for two reasons. First, purchasing shares would be more expensive than purchasing votes alone, thus increasing the price of obtaining the desired level of control. Second, the harm imposed on others by the shareholder’s looting would be diminished.\(^\text{218}\) In essence, a ban on vote-buying would force the shareholder to internalize a greater portion of the costs imposed by the looting.

Other corporate constituency groups are in even better position to use vote-buying to engage in looting because they can do so while suffering none of the detrimental effects associated with share ownership in the looted firm. Three prime examples of this are management, vendors, and customers. Management could derive benefits from controlling the vote even absent any ownership in the underlying shares. Because of the role it plays in the corporation, management can often extract benefits to the disadvantage of the shareholders.\(^\text{219}\) By taking the shareholder’s vote and placing it with management, these problems are exacerbated; management simply has greater control and hence a greater ability to loot. Furthermore, by allowing management to control the vote (one of the other monitoring mechanisms), the threat of a hostile takeover is impaired: management can use its purchased shareholder voting power to prevent the takeover transaction. The ability of management to inflict harm on the shareholders is especially pronounced in situations where management purchases the right to vote for an extended period of time.\(^\text{220}\)

\(^{218}\) That is, if the shareholder’s stake increases from 40% to 60%, $0.60, rather than $0.40, of every dollar “looted” already “belonged” to the shareholder. As a result, harms to the remaining shareholders are proportionately reduced.

\(^{219}\) For example, management can raise salaries, create generous retirement plans, or provide extravagant corporate perquisites. See André, supra note 4, at 598–99.

\(^{220}\) For instance, in a successful dual-class recapitalization, management shareholders can acquire effectively complete control for a period of unlimited duration. See id. at 620–21.
However, it still presents a problem even if the management purchases the vote on an issue by issue basis.

Similarly, allowing corporate vendors or customers to control the vote could be detrimental to a corporation. Either of these groups could use the vote to install board members “friendly” to their causes, or use this possibility to gain “converts” among the current board. This possibility could result in the corporation continuing contracts on terms less favorable than could be obtained elsewhere, thus redirecting corporate profits from shareholders to these outside groups.

This analysis suggests the following general rules. Those desiring to profit from control by an increase in share values are unlikely to buy votes; they would prefer to obtain control by purchasing the additional shares rather than merely the votes. The groups that prefer to purchase votes without the underlying shares are likely to do so because they intend to loot the company. Thus, a legal rule legitimizing vote-buying increases the risk of looting, particularly when the vote purchaser is a non-shareholder, and courts should be particularly leery of such transactions.

C. Vote-Buying and Coercive Transfers Between Different Classes of Corporate Participants

Even where the purchasers are shareholders rather than corporate insiders, vote-buying still presents another potential problem. It may be employed to create coercive wealth transfers between different “classes” of corporate participants. This Section examines two important vote-buying cases and concludes that they may well be examples of just such coercive transfers. The possibility that it could be used to effect such transfers argues strongly against permitting unrestricted vote-buying.

There are many different “classes” of corporate participants. Bondholders, shareholders, preferred shareholders, creditors, to name just a few, all derive substantial economic benefit from the existence of the corporation. The nature of the economic benefit varies greatly between these groups, however, with regard to the nature of the financial risk, the payment mechanisms employed, the term of the investment, the available exit options, and many other characteristics. Even among a single group such as shareholders, there may be separate classes,221 each

representing a different claim on the residual assets or earnings of the corporation.

The previous Section addressed the manner in which management and suppliers could use vote-buying to the detriment of the shareholder. However, a closer examination of two cases indicates that the various classes of corporate investors can also use vote-buying to better their position relative to investors in other classes.

I. Extorting Wealth from Bondholders: Lessons from Kass

Kass v. Eastern Air Lines, Inc.\(^{222}\) presents an example of just this type of inter-class conflict. As described previously,\(^{223}\) Kass involved a conflict between corporate shareholders and bondholders. The shareholders desired to undertake a merger in a manner that required the relaxation of certain covenants in some of the outstanding bonds.\(^{224}\) The company sought to purchase with cash and flight coupons the bondholders’ consent to amend the covenants.\(^{225}\) Some of the bondholders attempted to enjoin the offer as impermissible vote-buying, but, using a Schreiber analysis, the court upheld the transaction.\(^{226}\)

The problem with the court’s holding is that it did not reflect an important difference between the economic reality of Schreiber and that of Kass. In Schreiber, although the shares were different classes, the residual interest in the corporation was fairly similar. In Kass, by contrast, the nature of the economic interest between the disputing parties was quite different. Bondholders do not share in increased corporate growth. They have a fixed upside potential, the stated interest rate, but they do share in corporate losses, at least if those losses cause the corporation to default on its indebtedness. Thus, bondholders as a group tend to be risk averse.\(^{227}\) They would prefer a corporation make

---

Every corporation may issue 1 or more classes of stock or 1 or more series of stock within any class thereof, any or all of which classes . . . or series may have such voting powers . . . and such designations, preferences and relative, participating, optional or other special rights, and qualifications, limitations or restrictions thereof, as shall be stated and expressed in the certificate of incorporation . . . .


223. See supra notes 161–72 and accompanying text.


225. Id. at *2.

226. Id. at *5.

“safe” choices because they are not rewarded when risky ventures pay off. Shareholders, on the other hand, have unlimited upside participation. Increases in profits increase shareholder wealth. Thus, they are, in general, more willing to see the corporation adopt a greater risk level.

Bondholders, recognizing this divergence in risk tolerance, generally require covenants in bonds giving them the right to veto certain corporate acts, such as mergers or acquisitions, which can significantly affect the risk of the investment. Kass can be seen as a case where the shareholders, acting through the board, deprived the bondholders of the protections offered by these covenants.

In Kass, each bondholder found itself in a prisoner’s dilemma. Their consent was not likely to influence the outcome of the election. However, if an individual bondholder failed to consent, but the required number of bondholders did, the individual bondholder would be forced to accept the amended covenants without receiving the benefit of the cash offered by the company. The rational choice for each bondholder in this situation was to accept the compensation and give its consent. This was true even if the compensation was not sufficient to offset the additional risk incurred by the bondholders. By allowing the transaction, the court arguably allowed the shareholders to increase their wealth at the expense of the bondholders.

The most important manifestation of the conflict of interest between fixed claimants and shareholders lies in their attitudes toward the optimal level of risk that a firm should take. Shareholders have a powerful incentive to induce their firms to engage in activities that fixed claimants would consider excessively risky. This is because shareholders stand to reap all of the benefits from the spectacular success of a particularly risky activity, but stand to lose only the amount of their initial capital investment. Fixed claimants, in contrast, do no better when their firm performs very well than when their firm garners only a moderate return. For this reason, shareholders generally retain the right to control most details of a firm’s business, subject to the broad contractual protections that fixed claimants extract to protect themselves against default.

See also Thomas E. Stagg & Scott Ferretti, Contractual Protection: An Existing Remedy for Bondholder Distress, 4 St. John’s J. of Legal Comment 245, 253 (1989) (“The bondholder is generally viewed as more risk averse than the stockholder and has a direct interest in preserving corporate capital and earnings in order to maximize firm value.”).

228. See Economic Structure, supra note 4, at 68; Macey, supra note 227, at 181.

229. See John C. Coffee, Jr., Shareholders Versus Managers: The Strain in the Corporate Web, 85 Mich. L. Rev. 1, 19 (1986) (describing shareholders with a diversified portfolio as risk neutral, or even risk preferring); Macey, supra note 227, at 181; Stagg & Ferretti, supra note 227, at 252–55 (explaining why shareholders have a higher risk tolerance than bondholders).

230. Economic Structure, supra note 4, at 68.
It is important to realize that it is not the transfer of wealth per se that was objectionable, but rather the means by which it was achieved. The shareholders, acting collectively through the board, were able to exploit a collective action problem among the bondholders to change the terms of the contract between these two groups. Allowing this type of exploitation results in higher bond rates in general and inefficient pricing of bonds.\textsuperscript{231}

Under the holding in \textit{Kass}, bonds can be converted to covenant-free bonds at a relatively low price.\textsuperscript{232} However, that in turn implies that lenders will provide little discount for accepting covenants in the first place. Because covenants exist in most debt instruments, they must, as a general rule, improve the wealth of both parties to the transaction. By reducing bondholder reliance on such covenants, consent-buying reduces the magnitude of this wealth.

One potential response by lenders could be to insist upon covenants that cannot be altered even by consent. This approach, however, would also result in deleterious wealth effects. Such immutable covenants would in effect result in precommitment strategies without exit options. Thus, even if both lender and borrower truly wanted to change the terms of the agreement, it would be difficult to achieve.\textsuperscript{233}

Another alternative would be for bondholders to form representative committees similar to the board of directors. By forming such committees, the individual bondholder would escape the prisoner’s dilemma because each could be guaranteed concerted action among all of the bondholders. Of course, by adopting a board, the bondholders would incur some of the same agency costs currently experienced by corporate shareholders. And, in any event, creating such a board would increase transaction costs associated with the issuance of corporate debt, once again raising the rates from those in a world where the bondholders could not be coercively deprived of the benefit of their covenants through vote-buying. All in all, it seems likely that ex ante, both shareholders and

\textsuperscript{231} For an argument that bondholders are not necessarily subject to coercion in \textit{Kass}-type consent solicitation, see Royce De R. Barondes, \textit{supra} note 187, at 754. Barondes suggests that bondholders avoid coercion because the consent solicitation process is more accurately modeled as a repeat game rather than a single-shot game. \textit{Id.}

\textsuperscript{232} The court in \textit{Kass} did not undertake an intrinsic fairness analysis to see if the price paid accurately reflected the value of the covenants given up. \textit{See Kass v. Eastern Air Lines, Inc.}, 1986 WL 13008 (Del. Ch. Nov. 14, 1986). Note that in general this could prove to be a very difficult determination for a court to make.

\textsuperscript{233} Conceivably, if both agreed, the borrower could issue the lender a new bond with different covenants and use the proceeds to retire the old debenture. The transaction costs on such a swap (e.g., the costs of complying with securities law requirements) could, however, prove prohibitive.
bondholders would be better off in a world that did not permit shareholders to purchase consents. By failing to recognize the coercive nature of the transaction at issue, the Kass court erred in its decision.

2. Schreiber as a Rubinstein Game.

Even Schreiber can be viewed as an example of the use of vote-buying to effect a coercive wealth transfer between corporate participants. In Schreiber, the corporate bylaws required approval by a majority of each of three voting classes to approve a merger. A single shareholder owned a majority of the shares in one of the classes, and thus could maintain a blocking position on the corporate reorganization.

Everyone, including the blocking shareholder, agreed that the merger would increase corporate wealth, but the blocking shareholder would not agree to it until he extracted a presumably below market rate loan. Although the court did not endeavor to name the price, the favorable terms could, of course, be translated into a cash equivalent. Thus, in reality, Schreiber involved an attempt by one shareholder to extract wealth from other shareholders through use of his blocking position.

Of course, this case is different from Kass in that the economic interests of the shareholders in Schreiber were all the same. That is, all of the shareholders would presumably benefit from an action expected to increase share value. Because the interests among both vote-sellers and vote-buyers were aligned, the coercive transfer must have been accomplished somewhat differently than in Kass. After all, as the interests were the same, at first glance it seems that the other classes of stock could just as easily have demanded that their votes be purchased. In other words, it appears that the other shareholders could have rejected the blocking shareholder’s demand for a loan. If they had, it would presumably have been in his interest to acquire a loan from secondary

---


235. Schreiber, 447 A.2d at 19.

236. Id. at 19–20. Jet Capital’s management stated that a market rate loan was too expensive, so the 5% rate offered by Texas Air was presumably below market rate. The court does not indicate the magnitude of the difference between the terms offered and the market rate. Id. at 19.

237. The key in Kass was that the shareholders (vote-buyers) and the bondholders (vote-sellers) had a divergence in interest. See supra notes 220–26 and accompanying text.
sources and then still vote for the merger. How then was the blocking shareholder able to complete his extraction?

In essence, the negotiations between the shareholders could be modeled as a Rubinstein game. In the basic Rubinstein game, two parties must agree on how to split some sum of money. Here, the increase in corporate wealth resulting from the merger is the sum that must be split between the players. Furthermore, there is pressure on the players to complete the negotiations as quickly as possible. For simplicity purposes, then, \textit{Schreiber} can be modeled as a game involving two players, A representing the blocking shareholder and B representing the remaining shareholders. Both A and B will gain from permitting the merger, the only issue is the extent to which each gains. The fact that in \textit{Schreiber} the blocking shareholder (A) was able to extract the loan suggests a few possible alternatives.

The course of Rubinstein bargaining is strongly affected by the exit options open to the players. Here, however, there is no exit option for splitting the gains. The only exit option open to the players other than a successful negotiation is to forego the gains entirely. The Rubinstein model predicts that in such a splitting game, if there is perfect information, the players will split the gains evenly. This should imply

---

238. Presumably, with a loan, the merger would have a positive value for the shareholder. Thus, unless the “cost” of the loan exceeded the benefit of the merger (see infra notes 242–47 and accompanying text), the shareholder would obtain the loan and vote for the merger.


240. BAIRD ET AL., supra note 239, at 224–32.

241. Due to the speed with which business conditions change, merger negotiations are often very time-sensitive. \textit{See} Ashutosh Bhagwat, \textit{Modes of Regulatory Enforcement and the Problem of Administrative Discretion}, 50 \textit{HASTINGS L.J.} 1275, 1296 (1999) (“Many mergers, especially ones involving publicly traded companies, are often extremely time-sensitive since changes in reported profits or stock prices are likely to unravel any deal over time.”).

242. The remaining shareholders could, of course, also be blocking shareholders. The reason one shareholder was able to assume that position while others were not is examined infra notes 250–51 and accompanying text.

243. There can be no doubt that the blocking shareholder will benefit from the merger after the loan, otherwise he would continue to oppose it.

244. BAIRD ET AL., supra note 239, at 224–32.

245. \textit{Id.} at 224. More formally, Rubinstein predicts that the two players will split the one dollar so that player 1 receives $1/(1+\delta)$, and player 2 receives $\delta/(1+\delta)$, where $\delta = 1/(1+r)$ and $r$ is the per period interest rate of the players. \textit{Id.} at 223–24. \textit{See also} ROBERT GIBBONS, \textit{GAME THEORY FOR APPLIED ECONOMISTS} 68–71 (1992) (providing inductive mathematical proof of this result). When the time period between offers becomes arbitrarily short, however, $r \to 0$, and thus $\delta \to 0$, meaning both player 1 and player 2 receive $1/(1+1)$ or $\frac{1}{2}$ of the dollar. BAIRD ET AL., supra note 239, at 224.
that in *Schreiber*, the gains would be split on a per share basis, and A would not be able to extract any wealth from the other shareholders.

In reality, however, such a division of the gains may not actually have been “even.” In order to benefit from the transaction, A was required to exercise his warrants, which required obtaining a loan. Thus, A was required to incur a cost\(^{246}\) that B was not.\(^{247}\) Obtaining the loan from B\(^{248}\) might then be seen as a means of splitting this cost in order to arrive at an even split of the net profit from the transaction.\(^{249}\)

If this is indeed the case, the result of the vote-buying transaction may not be coercive. In fact, vote-buying may allow beneficial transactions to go forward that otherwise would not. To see this, assume that if A doesn’t exercise his warrants, the merger will indeed result in negative profits to him personally. However, if A does exercise the warrants, both A and B will benefit by $5. Further assume that it would cost A $6 to obtain the loan. A will not be willing to spend the $6 to benefit by $5, even though in the aggregate the benefits ($10) exceed the costs ($6). By allowing the remaining shareholders to “buy” A’s vote, they can share the $6 cost. Indeed, if such cost sharing is allowed, presumably all transactions that are beneficial in the aggregate would occur. Thus, if this were an accurate assessment of what is taking place, vote-buying may be beneficial.

Yet, it is difficult to ascertain whether this is truly taking place. For instance, the “sharing” could easily be skewed by the presence of asymmetric information. Presumably, both A (the blocking shareholder) and B (the remaining shareholders) knew the projected increase in share value resulting from the merger. Thus, A knew B’s true valuation of the merger. As discussed previously, however, A had costs associated with the transaction that B did not. Furthermore, B was not necessarily in a good position to know or even accurately estimate those costs;\(^{250}\) B may

\(^{246}\) For example, the origination cost of the loan.

\(^{247}\) The other shareholders were engaging in tax-free swaps. *Schreiber v. Carney*, 447 A.2d 17, 19 (Del. Ch. 1982).

\(^{248}\) The loan was obtained from the corporation, which can be viewed for these purposes as the representative of B (the other shareholders).

\(^{249}\) The net profit would be the increase in shareholder wealth minus the costs associated with the loan.

\(^{250}\) Remember, B is in reality a group of shareholders. As discussed previously, there are collective action problems associated with the group acquisition of information. *See supra* Section I.B.
simply have accepted A’s estimate of the costs. If the “net” profits are skewed by A’s overblown estimate of its costs, then B may end up “over-reimbursing” A.

This is a slightly less palatable explanation. If B is in the position of being “forced” to rely on A’s estimate of his costs because of the collective action problems associated with information gathering, then A can extract a certain amount of additional wealth from B by misrepresenting those costs.

Finally, the negotiations in Schreiber could be viewed as a breakdown in the Rubinstein model. The blocking shareholder in Schreiber was a single entity. The other party, by contrast, was a group. The group may have faced collective action problems in the bargaining process that the blocking shareholder did not. For instance, counter-offers may have been very difficult for the group to propose. If counter-offers are relatively more expensive, they are relatively less likely to be made.

In order to avoid this problem, the shareholder group in Schreiber used the board as its negotiating representative. However, this in turn may have created an agency problem. The blocking shareholder, for instance, may have been able to “capture” the board, resulting in less forceful negotiations on behalf of the remaining shareholders. Either way, this could be seen as an example of the blocking shareholder exploiting the collective nature of his “opponent.”

Independent of which of these explanations best describes what took place in Schreiber, it is not clear that vote-buying is desirable in cases like this. By permitting it, the courts allow the opportunity for additional negotiations regarding the distribution of the surplus from a corporate reorganization. These negotiations involve transaction costs including delays in the completion of deals. Against these costs, the only benefit to allowing vote-buying in cases such as Schreiber is to permit transactions to go forward that are beneficial in the aggregate, but detrimental to certain shareholders. It does this by permitting the benefiting parties to share their gains with the injured shareholders.

251. Of course, if this were the case, the shareholders could conceivably have a cause of action against the board for breach of fiduciary duty. See Revlon, Inc. v. MacAndrews & Forbes Holding, Inc., 506 A.2d 173, 182 (Del. 1986) (noting that in merger context, board has duty to maximize shareholder values); Christian C. Day, Corporate Governance, Conrail, and the Market: Getting on the Right Track!, 26 J. CORP. L. 1, 16 (2000) (explaining that in merger context board has duty, typically referred to as “Revlon duty,” to maximize shareholder value). It is likely, however, that even to the extent such a fiduciary duty were to exist, absent gross misconduct, the Business Judgment Rule would make it difficult to proceed on such a claim. See Day, supra, at 15–24 (discussing Business Judgment Rule in context of merger negotiations).
In general, however, this would involve a very small percentage of cases. Shareholders’ interests are usually very closely aligned. That which benefits one (in his or her capacity as a shareholder) generally benefits all. Even in those cases where vote-buying is arguably beneficial, collective action problems and asymmetric information could permit small groups of shareholders in control positions to use it to extract wealth from the remaining shareholders. Thus, even those vote-buying cases involving solely purchases among shareholders, vote-buying is still potentially harmful, and there is reason to suspect that courts are incapable of recognizing and policing the harm when it occurs.

D. Vote-Buying as a Means of Facilitating Control Transactions

As the previous sections demonstrate, vote-buying presents the possibility of corporate looting and/or coercive wealth transfers between corporate participants. Commentators have suggested, however, that the potential benefit of vote-buying as a means to overcome the shareholder costs imposed by “shareholder rights plans” justifies permissive treatment of vote-buying, notwithstanding these identified costs. This Section addresses their arguments, concluding that while its use as a means of overcoming shareholder rights plans may provide some benefits, the same benefits could be obtained through use of turnout incentives, without incurring the potential harms vote-buying presents.

It is widely recognized that the possibility of corporate acquisition is generally beneficial to shareholders. As described above, collective

---

252. See supra notes 214–39 and accompanying text.

253. See generally André, supra note 4, at 585–606 (suggesting that courts should permit vote-buying as a means around anti-takeover statutes); see also Hasen, supra note 1, at 1349–52 (suggesting that vote-buying permits potential acquirers to overcome shareholder rights plans, thereby enhancing corporate efficiency); Levmore, supra note 2, at 138 (suggesting that vote-buying is pro-shareholder). For discussion of a related idea, see Clark, supra note 4, at 806–07 (suggesting that vote-buying should be allowed by shareholders who are planning to recoup their investment by an increase in the share price of the company’s stock).

action problems among shareholders can prevent effective monitoring of corporate managers, leading to entrenched (mis)management and resulting loss in shareholder value. Acquirers seek to purchase control positions in companies suffering from such mismanagement and replace the existing management with better managers, thereby capturing the gains associated with the increased productivity. The monitoring effect created by the potential for such transactions provides a useful tool for keeping the current management focused on creating shareholder value.

During the late 1980s, however, managers (often with the help of state legislatures) had largely succeeded in protecting themselves from the threat of hostile takeovers through the use of sophisticated takeover defenses. These defenses generally fell into two categories: antitakeover statutes and “shareholder rights plans,” the latter of which were often euphemistically referred to by more colorful names such as “poison pills” and “shark repellents.” The defenses were designed to allow incumbent managers to thwart the ability of corporate outsiders to effect changes in corporate control through share acquisition. Basically, the defenses adopt provisions whereby, in the event an “unfriendly” suitor purchases a threshold number of shares in the company, some adverse consequence follows. For example, the newly-acquired shares might

255. See supra notes 17–47 and accompanying text.
256. See, e.g., Fischel, supra note 254, at 9.
258. For an in-depth description of the poison pill defense, see Matheson, supra note 9, at 726–33.

The most powerful tool to incumbents in impeding bids was the poison pill. The pill provided managers with a highly effective, easy, and costless way to impede an unwelcome bid. Unlike other defensive tactics (such as a target company’s defensive acquisition of other companies) that have some effect in the world other than stopping a bid, the pill is an artifice whose only upshot is impeding a bidder.

260. For instance, CTS Corp. v. Dynamics Corp. of Am., 481 U.S. 69, 73–74 (1987), involved an Indiana statute providing that whenever a purchaser gained more than 20% control of a corporation, the shareholder would be precluded from voting unless a majority of the disinterested shareholders
Corporate Vote-Buying

not be permitted to vote, or the corporation might issue a share dividend to all but the newly acquired shares, thereby diluting the acquirer’s interest.²⁶¹

Of course, these shareholder rights plans and antitakeover statutes drastically reduced the incentive to attempt a hostile takeover by share acquisition.²⁶² As a result, incumbent managers may remain in place, resulting in continued inefficient use of corporate assets, or, alternatively, the acquirers may be forced to negotiate a side deal with the board prior to acquisition, thereby turning the hostile takeover into a “friendly” acquisition.²⁶³ In essence, the board holds the corporate assets captive, releasing them only if the acquirer pays ransom, thereby reducing the gain available to be shared between the existing shareholders and the acquirer.

Commentators have suggested that vote-buying presents a possible solution to the problem of management that has entrenched itself through use of shareholder rights plans or anti-takeover statutes.²⁶⁴ The basic theory is that acquirers could purchase the vote, and then use it to dismantle the protections by replacing the board with one friendly to the acquisition.²⁶⁵

It certainly seems correct as an initial matter that vote-buying in such situations is not suggestive of corporate looting. The presumption of looting arises due to the purchaser’s choice to purchase votes rather than shares.²⁶⁶ In the face of the adverse consequences under an anti-takeover statute or shareholder rights plan, however, the purchase of the vote does not indicate a preference for the vote over the share. Rather, the vote purchase is but the first step in the share purchase.

approved the voting rights. Moran v. Household Int’l Inc., 500 A.2d 1346, 1348–49 (Del. 1985), involved a shareholder rights plan that allowed purchase of the corporate shares at depressed prices in the event of an attempted takeover.

²⁶¹ CTS Corp., 481 U.S. at 73–74; see also Matheson, supra note 9, at 726–29 (discussing the various types of poison pills companies have adopted).

²⁶² See supra note 9, at 726–29.

²⁶³ See Bebchuck & Ferrell, supra note 259, at 121, stating that: The poison pill, backed by an entrenched management, is extremely formidable. Incumbent management can use this power to prevent an acquisition that they do not want for self-serving purposes (such as saving their jobs). They can also use this power to extract private benefits for themselves, perhaps diverted from what would have otherwise gone to the shareholders, in return for redeeming the pill and allowing the tender offer to proceed.

²⁶⁴ See supra note 253.

²⁶⁵ See generally André, supra note 4, at 585–606.

²⁶⁶ See supra notes 214–20 and accompanying text.
Although facilitating control transactions may be a laudable goal, the efficacy of vote-buying as a means of achieving that goal is questionable for at least three reasons. First, if vote-buying is generally recognized as legal, shareholder rights plans and anti-takeover statutes will undoubtedly adjust to prevent “end runs” accomplished through its use. Indeed, shareholder rights plans typically are triggered by the acquisition of “beneficial” as well as actual ownership of shares, and “beneficial ownership” is typically defined to include the acquisition of a share’s voting right.267 Thus, vote-buying would be no more effective than share acquisition in dismantling such plans.

Second, while the possibility of takeovers is undoubtedly beneficial to shareholders, not all takeover bids are necessarily in the shareholders’ best interest.268 For example, given additional time, a competing bid for a company at a higher price might appear. The choice of whether or not to proceed with a given takeover transaction, then, should be left to the shareholders. If vote-buying is permissible, however, shareholders would be generally incapable of resisting the vote purchase due to the collective action problems they face.269 In light of those problems, the act of tendering the proxy would not necessarily be indicative of the shareholder’s preferred outcome on the takeover transaction, but rather merely a reflection of the coercive nature of the vote purchase offer.

Third, vote-buying might present a new opportunity for corporate outsiders to obtain wealth-diverting transfer payments from corporate management. That is, a purchaser might employ a vote-buying transaction for the stated purpose of acquiring sufficient votes to move forward with an acquisition. As described above, at least in a corporation with dispersed ownership, it should be possible to purchase the votes

267. For example, Stahl v. Apple Bancorp, 16 Del. J. Corp. L. 1573 (Del. Ch. 1990), involved just such a plan. The shareholder rights plan was triggered by a shareholder attaining beneficial ownership of a 15% or greater stake in the company. Id. at 1577. The plan specifically defined beneficial ownership to include acquisition of the voting right of a share. Id. at 1580. The court concluded that under the language of the plan even acquisition of a revocable proxy would constitute beneficial ownership. Id. at 1585–86.


269. See supra notes 234–58 and accompanying text.
Corporate Vote-Buying

quite cheaply. Rather than complete the control acquisition through share purchase, however, the purchaser might instead use the votes (and the threat of a control transaction) to negotiate a transfer payment from the board in exchange for not going forward—that is, extract “greenmail.” In short, while it is indisputable that shareholder rights plans and anti-takeover statutes impose costs on shareholders, vote-buying does not necessarily eradicate those costs, and it may in fact impose additional ones.

A better response to the problem of shareholder rights plans, one that captures the benefits of vote-buying without presenting the harms, would be the use of a variant of vote-buying turnout incentives. Under the turnout incentive approach, the shareholder would be paid for voting, independent of the voting option the shareholder selects. Such payments would increase participation in corporate governance because, as noted above, voting is not costless. Even if the voting information is collected and provided to the shareholders, the cost of reviewing and analyzing that information may cause rational apathy among shareholders. A system that paid the shareholder an incentive fee for returning his or her proxy, independent of the shareholder’s choice of

270. See id.

271. “Greenmail” refers to corporate payments made to potential acquirers in exchange for the acquirer dropping the acquisition plan. For a general discussion of greenmail, see Jonathan R. Macey & Edward S. McChesney, A Theoretical Analysis of Corporate Greenmail, 95 YALE L.J. 13 (1985). While these authors suggest the possibility that greenmail may have some beneficial effects for shareholders, id. at 24–25, it is well accepted that such payments hold at least the potential for harm to shareholder interests. See Jeffrey N. Gordon & Lewis A. Kornhauser, Takeover Defense Tactics: A Comment on Two Models, 96 YALE L.J. 295, 320 (1986) (arguing that the abusive potential of greenmail justifies an outright ban on the practice); Laura Lin, The Effectiveness of Outside Directors as a Corporate Governance Mechanism: Theories and Evidence, 90 Nw. U. L. Rev. 898, 934 (1996). The directors’ fiduciary duties may provide some limits on their ability to protect their positions through use of greenmail. As a general matter, however, courts have upheld such payments in the past as appropriate exercises of directorial discretion. See, e.g., Polk v. Good, 507 A.2d 531, 537 (Del. 1986) (holding that purchase of raider’s block of shares at premium not an abuse of directors’ discretion); Grobow v. Perot, 526 A.2d 914, 927 (Del. Ch. 1987) (dismissing shareholder derivative claims predicated on alleged greenmail payment by General Motor’s board to Ross Perot).

272. For a discussion of the use of turnout incentives in political contests, see Hasen, supra note 1, at 1355–59.

273. Responsibility for the costs of the turnout incentives could be similar to proxy solicitations. The challenging party would bear the cost of paying the incentives, but could recoup those payments out of corporate assets in the event that it was successful, or at the discretion of the directors if it was unsuccessful. See, e.g., Steinberg v. Adams, 90 F. Supp. 604, 607–08 (S.D.N.Y. 1950) (holding that successful proxy opponents can claim reimbursement); Rosenfeld v. Fairchild Engine & Airplane Corp., 128 N.E.2d 291, 293 (N.Y. 1955) (holding that a corporation, at its option, could reimburse a successful rival faction for expenses incurred in a policy fight).
who received the proxy, would increase shareholder participation without coercively depriving the shareholder of her voice. In other words, the shareholder is no longer compelled to tender his or her vote or face the prospect that it will be worthless. Instead, the shareholder’s vote has the same value (the turnout incentive amount) independent of the group for whom it is cast.

In this way, the shareholders have an incentive to review the materials and participate in the voting contest. If the proposed takeover is truly in the shareholder’s interest, the shareholder would presumably still tender his or her vote to the acquirer. Because that tender resulted from a turnout incentive, rather than a vote purchase, however, it would be a reliable indicator of the shareholder’s true preference rather than a reflection of the coercive nature of the transaction.

IV. VOTE-BUYING AS A TOOL FOR PREFERENCE AGGREGATION IN CLOSELY HELD CORPORATIONS

Up until this point, the primary focus of the discussion has been vote-buying in large publicly held corporations. Closely held corporations, by contrast, provide significantly different regulatory concerns. Although the case law draws no distinction between vote-buying in publicly held and closely held corporations, the different nature of the regulatory concerns between the two suggests a more permissive approach to vote-buying with regard to the latter.

In closely held corporations, information and coordination problems tend to be minimized. The blocks of stock are usually much larger on a percentage basis, so the various shareholders have greater incentives to invest in information. Furthermore, most of the shareholders are also

274. This is not to suggest that turnout incentives would provide a panacea. For instance, a rational shareholder may consider it profit maximizing to decide how to vote based on a coin flip. In that way, the shareholder would receive the turnout incentive without incurring the analysis costs associated with reviewing the information.

275. See ECONOMIC STRUCTURE, supra note 4, at 228–52; Moll, supra note 103, at 756–58 (describing differences between close corporations and traditional public corporations and suggesting that “[c]onventional corporate law norms of majority rule and centralized control can lead to serious problems for the close corporation minority shareholder”).


Corporate Vote-Buying

active within the corporation, providing them with better access to information than an outside shareholder might have.\textsuperscript{278} In addition, because the shareholders work together and often have familial or social bonds, the free-rider problem is less prevalent.\textsuperscript{279} Thus, many of the problems facing shareholders in large corporations with widely-dispersed ownership simply do not exist, or are, at any rate, much less acute, in small, closely held corporations.\textsuperscript{280}

At the same time, other problems may replace these concerns. For instance, in closely held corporations, it may be less likely that there are uni-peaked preferences.\textsuperscript{281} This, in turn, can generate instabilities in the shareholder voting process such as path-dependent outcomes or cycling.\textsuperscript{282}

There are a number of reasons to suspect that multi-peaked preferences may exist in closely held corporations. For the typical shareholder in a publicly held corporation, the shareholder’s stake is merely part of the shareholder’s diversified portfolio. Because of the protection offered through diversification, the average shareholder is thus likely to be risk neutral.\textsuperscript{283} In closely held corporations, on the other hand, one might expect a broader spectrum of risk aversion. Many of the shareholders (e.g., the entrepreneur in a start-up, or the manager/owners following a management-leveraged buyout) may have their entire savings, or at least a very significant portion of it, invested in the company. In addition, they may have a significant portion of their human

\textsuperscript{278} Id. at § 1.08.

\textsuperscript{279} Id. at § 1.02; see also Robert B. Thompson, The Law’s Limits on Contracts in a Corporation, 15 J. Corp. L. 377, 392 (1990) (“A close corporation provides more direct opportunity for specific private ordering because there exists no large-numbers problem that can lead to free rider questions or rational apathy.”).

\textsuperscript{280} O’Neal & Thompson, supra note 277, at § 1.08:

Because of this combination of [the decision-making function and the risk-bearing function] in close corporations, such corporations can avoid the sometimes costly agency problem of protecting passive shareholders who need some method of monitoring the performance of the corporation’s managers and ensuring that management’s decisions are aligned with shareholder interests.

\textsuperscript{281} For an explanation of “uni-peaked” preferences, see supra note 15.

\textsuperscript{282} See generally Farber & Frickey, supra note 15 at 38–62 (discussing Arrow’s Theorem). Indeed, Arrow’s Theorem holds that in the presence of varying preferences among individual voters (i.e., non-peaked or multi-peaked preferences), no method of combining preferences (e.g., voting) can satisfy minimum standards of rationality. Id. at 38–39.

\textsuperscript{283} George W. Dent, Jr., The Role of Convertible Securities in Corporate Finance, 21 J. Corp. L. 241, 245 (1996) (“Most shareholders are risk-neutral toward each public company because they diversify away risk by holding a broad portfolio of investments.”).
capital invested in firm-specific capital. Thus, these shareholders would tend toward a higher degree of risk aversion.\textsuperscript{284} Other shareholders, such as venture capitalists investing in the corporation, may have a more diversified portfolio, and thus a better risk tolerance. These risk aversion differences could easily lead to multi-peaked preferences.

Moreover, in a closely held corporation, many of the shareholders may also hold positions within the company. This sets up a potential conflict in preferences between the employee-shareholders, and the non-employee-shareholders with respect to issues such as appropriate compensation and dividend policy.\textsuperscript{285} To further exacerbate the problem, the exit option may be less easily realized because there is rarely a large, liquid market for shares in closely held corporations.\textsuperscript{286} Finally, the same familial and social bonds that serve to minimize the free-rider problem could result in conflicts that disrupt the continuity of corporate control.

Faced with multi-peaked preferences, vote-buying may be a desirable means of preference aggregation.\textsuperscript{287} For example, assume that there are three equal shareholders (A, B, and C) in a closely held corporation. Further, assume that an issue comes to vote before the shareholders, for instance, authorization to pursue acquisition of a competitor. A believes

\begin{itemize}
  \item \textsuperscript{284} Paul Fellows & S.Y. Wu, \textit{The Scope of Legal Interventions in Corporate Affairs}, 15 J. CORP. L. 465, 472–75 (1990) (discussing entrepreneurial risk aversion and its impacts on financing choices). In fact, “[i]t is generally understood that the transition from an owner-operated sole proprietorship to a public corporation is predicated on either a wealth constraint or the risk aversion of the entrepreneur.” Michael Bradley & Cindy A. Schipani, \textit{The Relevance of the Duty of Care Standard in Corporate Governance}, 75 IOWA L. REV. 1, 14 (1989).
  \item \textsuperscript{285} That is not to suggest that publicly held companies do not also have employee-shareholders. As a percentage of total ownership, however, the employee-shareholders typically consist of a relatively small group. That is, while there may be multi-peaked preferences with regard to such issues in the publicly held corporation, the relative magnitudes of the preferences in terms of share ownership are quite different.
  \item \textsuperscript{286} See Moll, supra note 103, at 789–90 (suggesting that absence of market for shares is primary cause of potential shareholder oppression in close corporations). In addition, the shares may be subject to restrictions on transfer, such as vesting or a corporate right of first refusal at a predetermined price. See DEL. CODE ANN. tit. 8, §§ 202(c) (permitting designated categories of transfer restrictions), 342(a)(2) (making the existence of a stock transfer restriction a prerequisite to the election of statutory close corporation status) (1991 & Supp. 2000).
  \item \textsuperscript{287} This Article is not the first to suggest the general possibilities of vote-buying as a preference aggregation tool. In his recent article, for example, Saul Levmore discusses the use of vote-buying for preference aggregation in civic elections. \textit{See generally} Levmore, supra note 2, at 142–58. While he also considers the possibility that similar arguments could support the application of vote-buying to the corporate governance realm, he dismisses the need for preference aggregation there because shareholders share “a single metric, value maximization.” \textit{Id.} at 158. As discussed in the text, while this analysis is correct in the context of publicly held corporations, it fails to recognize the possibility of divergent shareholder interests in closely held corporations.
\end{itemize}
that such an action would result in an expected net return to the company of $10,000, B believes it will result in a $5,000 loss, and C believes that the competitor is priced such that the acquisition would neither result in benefit nor harm (i.e. $0 value).

A would clearly vote for the acquisition, and B against it. The swing vote (C) would have no real basis for deciding one way or the other. However, if vote-buying were permitted, A and B could bid for C’s vote. Because A values the vote more highly, A would presumably buy C’s vote, and the acquisition would move forward.\textsuperscript{288} This results in the option with the highest aggregate shareholder value being chosen.

One possible objection is that this results in C getting more than “his share” of the value. Of course, from a pure efficiency standpoint, such distributional concerns are moot.\textsuperscript{289} In any event, it is unlikely that C would receive a large premium. Assuming A, B, and C all have equal information about the others’ preferences and the associated prices,\textsuperscript{290} B would have very little incentive to engage in an auction for C’s vote because he knows that A would outbid him. Without a competing bidder, C would not really have a credible threat to vote with B. Thus, A should be able to “buy” the vote at a relatively low price. In fact, if C holds out for too high a price, A could simply buy B’s vote. Once B realizes that he would lose a bidding contest (i.e., A’s preference outweighs his), B could easily switch from competing with A for purchase, to competing with C as a seller. This incentive mechanism also serves to keep the price relatively low.

Of course, as a practical matter, many closely held corporations use voting trusts or voting agreements to avoid the control difficulties envisioned here.\textsuperscript{291} Or, they may use multiple classes of stock guaranteeing the competing shareholders control over set percentages of the board, a favorite technique in venture capital deals. Vote-buying, however, with its ability to provide relatively efficient preference aggregation, might prove a useful technique in small corporations.

\textsuperscript{288} Note that A will be willing to pay up to $3,333 (one third of $10,000) for the vote, while B would only be willing to pay half of that (to avoid his perceived loss of one third of $5,000).

\textsuperscript{289} So long as all efficient transactions go forward (i.e., shareholder preferences are maximized), the allocative aspects of the vote-buying transaction do not impact the efficiency of the corporation.

\textsuperscript{290} Given that the example posits a closely held corporation with only three shareholders, this assumption is plausible.

\textsuperscript{291} See O’NEAL & THOMPSON, supra note 277, at § 5.02.
CONCLUSION

The Internet promises to make shareholder voting a more frequently used and more powerful tool for corporate management. A renewed interest in corporate voting, however, also requires a closer look at the voting process, and the appropriate limitations on methods used to compete for votes in the corporate world. One such method is vote-buying.

The current legal rule is that vote-buying is, as a general matter, permissible. Although this change from the earlier common law treatment of such transactions as per se illegal has been welcomed by some, it is by no means clear that such transactions are beneficial. In a public corporation, vote-buying presents possibilities for both looting and coercive wealth transfers. While it also has potential as a means to overcome anti-shareholder “shareholder rights plans,” that potential could be better achieved through use of turnout incentives. Thus, in public corporations, there are reasons to believe that a return to per se invalidity would be beneficial for shareholders.

In close corporations, vote-buying offers greater promise. Here the principal difficulties are not the collective action problems faced by atomized shareholders in publicly-held corporations. Rather, in a close corporation, the shareholders are more likely to face preference aggregation problems. Vote-buying is potentially a useful tool for overcoming those aggregation problems by allowing the competing shareholders to express the relative strengths of their preferences through bidding for the votes of the more neutral shareholders. While this aggregation can be accomplished through other means as well, vote-buying may represent the most convenient, inexpensive, and efficient tool for achieving it.
ECHAZABAL V. CHEVRON: A DIRECT THREAT TO EMPLOYERS IN THE NINTH CIRCUIT

Deborah Leigh Bender

Abstract: Although Congress enacted the Americans with Disabilities Act (ADA) in part to protect disabled individuals from paternalism, the ADA permits employers to adopt a requirement that individuals not pose a direct threat to others in the workplace. The Equal Employment Opportunity Commission (EEOC) has determined that this direct threat defense also protects an employer who discharges or refuses to hire individuals who pose a direct threat to their own health or safety in the workplace. In Echazabal v. Chevron, the Ninth Circuit struck down the EEOC interpretation of direct threat on the ground that it was paternalistic and inconsistent with the purpose of the ADA. This Note argues that the EEOC interpretation was reasonable and consistent with the language and underlying purpose of the direct threat defense. This Note concludes that the Ninth Circuit’s rule will force employers to choose between violating the ADA and knowingly hiring individuals who are at a substantial risk of injury or death in the workplace.

In the Ninth Circuit, a construction company no longer has the right to prevent a construction worker with vertigo from working atop a 50-story high-rise.1 Yet that same construction company is also bound by state safety guidelines that hold the company criminally liable if that employee tumbles to the ground.2 This employer is stuck between a rock and a hard place because of the recent decision in Echazabal v. Chevron USA, Inc.3 In Echazabal, the Ninth Circuit held that the Americans with Disabilities Act (ADA)4 prevents employers from considering the health or safety of a disabled individual when deciding whether that individual is qualified for a job.5 According to the Echazabal court, the ADA sanctions the use of safety-based qualification standards only to prevent direct threats to others in the workplace.6 Thus, the Ninth Circuit concluded that the direct threat defense only includes threats to others.7

1. This scenario is similar to the hypothetical posed by Chevron’s attorneys in their petition for rehearing en banc. Appellee’s Petition for Rehearing with Suggestion for Rehearing En Banc at 13, Echazabal v. Chevron USA, Inc., 226 F.3d 1063 (9th Cir. 2000) (No. 98-55551).
2. See infra Section III.
3. 226 F.3d 1063 (9th Cir. 2000), petition for cert. filed, 69 U.S.L.W. 3619 (U.S. Mar. 9, 2001) (No. 00-1406).
5. Echazabal, 226 F.3d at 1067–68.
6. Id.
7. Id.
The Equal Employment Opportunity Commission (EEOC) regulations, however, expressly state that an employer may also invoke the direct threat defense when the safety of the disabled individual is at risk. Congress gave the EEOC an express grant of authority to issue and enforce regulations implementing the ADA, including this expanded definition of “direct threat.” Despite the fact that two other Circuits had accepted the EEOC’s definition, and despite precedent in the Ninth Circuit establishing an employer’s right to use qualification standards ensuring safe job performance, the Ninth Circuit summarily rejected the EEOC interpretation of direct threat. The *Echazabal* court concluded that the “plain language” of the statute referred to direct threats to others, not threats to the disabled individual. Moreover, on looking at the legislative history of the ADA, the court determined that congressional intent to prevent paternalism in the workplace precluded an interpretation of the ADA that allowed an employer to make hiring and firing decisions based upon the safety of the disabled individual.

This Note argues that the *Echazabal* court should have deferred to the EEOC regulations that define the direct threat defense to include a substantial risk to the health and safety of the disabled individual in the workplace. Part I of this Note explains the purpose and scope of Title I of the ADA and examines the direct threat defense. Part II explores the regulations and case law that developed under the ADA’s predecessor, the Rehabilitation Act. Part III focuses on California’s workplace safety laws that hold employers criminally liable for knowingly subjecting their employees to harm. Part IV illustrates the process that a court must go through before it can reject the kind of agency regulation at issue in *Echazabal*. Part V outlines the facts, procedural history, holding, and rationale of *Echazabal*. Finally, Part VI argues that the *Echazabal* court

---

11. *See* Mantolete v. Bolger, 767 F.2d 1416 (9th Cir. 1984); Bentevegna v. United States Dep’t of Labor, 694 F.2d 619 (9th Cir. 1982).
13. *Id.* at 1067.
14. *Id.* at 1067–68.
erred in determining that the EEOC interpretation of direct threat was inconsistent with the text and purpose of the ADA. As a result, *Echazabal* has forced employers to choose between subjecting at-risk individuals to harm and violating the ADA.

I. TITLE I OF THE ADA PRESERVES AN EMPLOYER’S RIGHT TO SCREEN FOR QUALIFIED EMPLOYEES

Title I of the ADA prohibits employers from discriminating against disabled individuals. The ADA strikes a balance between the legitimate interest of disabled individuals in obtaining employment and the employers’ legitimate interest in maintaining a safe workplace. The text of the Act reflects Congress’s intent to prevent the paternalistic exclusion of disabled individuals from the workplace. Even so, the ADA allows employers to adopt certain qualification standards that tend to exclude disabled individuals from the workplace. Employers may adopt any standard that is job-related and consistent with business necessity, and they may also adopt a specific standard requiring employees not to pose a direct threat to others in the workplace. Although the text of the ADA mentions only direct threats to “others,” EEOC regulations state that the direct threat defense includes threats to the disabled individual as well. A number of circuits have adopted the EEOC interpretation of the direct threat defense.

A. *Purpose and Scope of Title I*

Congress enacted the ADA in 1990 as a comprehensive mechanism to end discrimination against disabled individuals in all aspects of life. Title I of the ADA targets discrimination in both public and private sector employment. Titles II–IV cover discrimination in state and local government services, in public accommodations, and in telecommunications, respectively. Congress found Title I necessary


because existing laws were inadequate to eliminate workplace discrimination.\textsuperscript{20} For example, the Rehabilitation Act\textsuperscript{21} offered protection for disabled individuals in public sector jobs, but no federal law protected disabled workers in the private sector.\textsuperscript{22}

Before the enactment of the ADA, workplace discrimination contributed to staggering levels of unemployment and poverty among persons with disabilities in the United States.\textsuperscript{23} Congress found that 43 million Americans had one or more disabilities\textsuperscript{24} and that these Americans were “severely disadvantaged” in finding and retaining jobs.\textsuperscript{25} Two-thirds of these disabled Americans were without employment, though a large majority wanted to work.\textsuperscript{26} Disabled Americans who were lucky enough to have jobs received unequal pay for their efforts—in 1988, disabled men and women were earning between thirty-six and thirty-eight percent less than their non-disabled counterparts.\textsuperscript{27} In its findings, Congress stated that this unequal treatment was a result of “stereotypic assumptions not truly indicative of the individual ability of such individuals to participate in, and contribute to, society.”\textsuperscript{28} Furthermore, Congress found that if it did not act, this “unnecessary discrimination and prejudice” would continue to “[deny] people with disabilities the opportunity to compete on an equal basis and to pursue those opportunities for which our free society is justifiably famous.”\textsuperscript{29}

Congress enacted Title I of the ADA to address workplace discrimination based on both real and perceived disabilities.\textsuperscript{30} Congress included perceived disabilities under the ADA’s protection, recognizing that social perceptions and myths about disabilities could be as disabling as the underlying conditions themselves.\textsuperscript{31} Title I protects qualified

\textsuperscript{23} Id. at 32, reprinted in 1990 U.S.C.C.A.N. at 314.
\textsuperscript{25} Id. § 12,101(a)(6).
\textsuperscript{27} Id. (internal citation omitted).
\textsuperscript{28} 42 U.S.C. § 12,101(a)(7).
\textsuperscript{29} Id. § 12,101(a)(9).
\textsuperscript{30} Id. § 12,102(2) (defining “disability” as “(A) a physical or mental impairment that substantially limits one or more of the major life activities of such individual; (B) a record of such an impairment; or (C) being regarded as having such an impairment”).
\textsuperscript{31} See id.
disabled individuals by prohibiting discrimination in the terms, conditions, and privileges of employment—including hiring, advancement, discharge, job training, and benefits. The prohibition of discrimination applies to public and private sector employers, employment agencies, labor organizations, and joint labor-management committees.

B. The ADA Discourages Paternalistic Employment Decisions

The language and legislative history of the ADA and general employment discrimination case law state that adverse employment decisions should not be based on paternalistic concerns about the employee’s welfare. Paternalism is the deliberate interference with an individual’s freedom of choice for “his own good.” In the ADA’s findings section, Congress suggested that paternalism was a cause of discrimination, resulting in “overprotective rules and policies” that exclude the disabled. For example, Senator Edward Kennedy commented that the direct threat defense must not enable employers to make paternalistic decisions about what is “best” for HIV positive job applicants. Finally, the Committee on Education and Labor noted in its report to Congress that “[p]aternalism is perhaps the most pervasive form of discrimination for people with disabilities.”

Although the United States Supreme Court has not yet addressed the issue of paternalism under the ADA, the Court has rejected paternalistic justifications for discrimination under Title VII of the Civil Rights Act of 1964. Specifically, the Court has held that a bona fide occupational qualification (BFOQ) cannot exclude all members of a protected class.

---

32. See infra Section I.C.1.
33. Mullen, supra note 19, at 186.
34. 42 U.S.C. § 12,112(a).
35. Id. § 12,111(2).
40. 42 U.S.C. §§ 2000e to 2000e-17 (prohibiting employers from discriminating on the basis of race, color, religion, sex, or national origin). The Echazabal court suggested that Supreme Court dicta stating that Title VII prohibits employers from making paternalistic decisions also applies to the ADA. Echazabal v. Chevron USA, Inc., 226 F.3d 1063, 1068 (9th Cir. 2000), petition for cert. filed, 69 U.S.L.W. 3619 (U.S. Mar. 9, 2001) (No. 00-1406).
out of a paternalistic concern for that protected class’s welfare. A BFOQ is a blanket rule that excludes all members of a protected class because they are either unable to perform an essential function of the job, or because it is impracticable to distinguish among members of the protected group. For instance, under Title VII, a synagogue may require that its rabbi be Jewish, despite the fact that this requirement discriminates against non-Jews based on their religion. The text of Title VII explicitly provides that employers may utilize such BFOQs. Unlike Title VII, however, the text of the ADA does not provide for the BFOQ as an employer defense. Thus, it is unclear whether the Title VII case law condemning paternalism applies to the ADA.

C. The ADA Permits Employers to Utilize Certain Requirements Regardless of Their Effect on Disabled Individuals in the Workplace

1. Requirements that Are Job-Related and Consistent with Business Necessity

In certain circumstances, an employer may defend the use of discriminatory qualification standards by using the ADA’s business necessity defense. While requiring employers to include disabled individuals in the workforce, the ADA’s employment protections extend only to individuals who are “qualified” for a given job. The statute defines a “qualified individual with a disability” as a disabled individual who can perform the “essential functions” of a job with or without reasonable accommodation. Without expressly defining “essential functions,” the statute indicates that courts will consider the employer’s judgment as to which functions are essential. Moreover, if an employer

42. BLACK’S LAW DICTIONARY 168 (7th ed. 1999).
44. But see 42 U.S.C. § 12,117 (stating that enforcement mechanisms outlined in Title VII are applicable to ADA); Muth v. Cobro Corp., 895 F. Supp. 254, 255 (E.D. Mo. 1995) (stating that Title VII procedures—specifically, enforcement mechanisms—are applicable to ADA actions).
45. 42 U.S.C. § 12,111(8).
46. Id.
47. Id.
prepares a written job description before advertising the position or interviewing applicants, that description carries evidentiary weight as to the “essential” job functions.48

Employer discretion has limits: an employer may only use qualification standards that screen out the disabled if the employer shows such standards are “job-related for the position in question and consistent with business necessity.”49 This requirement gives rise to the business necessity defense under the ADA. Before taking advantage of the defense, an employer must either make a “reasonable accommodation”50 for the applicant in question or show that accommodating the employee’s disability would impose an “undue hardship” on the business.51 Thus, the ADA combats qualification standards that are based on stereotypes but permits hiring criteria based on the actual attributes of a disability when those criteria are job-related and consistent with business necessity, and when the disability cannot be reasonably accommodated.52 So far, one circuit has held that an employer may defend the use of safety requirements under the business necessity defense.53

2. Requirements that Employees Not Pose a Direct Threat to Others in the Workplace

The direct threat provision is one example of the kind of permissible qualification standards envisioned by the ADA’s drafters. This provision allows employers to require that workers not pose a “direct threat to the health or safety of other individuals in the workplace.”54 The ADA

48. Id.
49. Id. § 12,112(b)(6).
50. Id. § 12,111(9) (stating that “reasonable accommodation” may include altering existing facilities or restructuring job to accommodate disabled individuals).
51. Id. § 12,111(10)(A) (stating that “undue hardship” is action requiring significant difficulty or expense); id. § 12,111(10)(B) (describing factors considered when determining if accommodation would pose “undue hardship”); id. § 12,112(b)(6) (stating that the term “discriminate” includes the use of qualification standards that screen out or tend to screen out individuals with disabilities unless such qualification standards are job-related and consistent with business necessity).
53. EEOC v. Exxon, Corp., 203 F.3d 871, 875 (5th Cir. 2000).
54. 42 U.S.C. § 12,113(b).
defines direct threat as “a significant risk to the health or safety of others that cannot be eliminated by reasonable accommodation.”

The legislative history of the ADA explains that the direct threat defense codified the standard adopted by the United States Supreme Court under the Rehabilitation Act of 1973. In *School Board of Nassau County, Florida v. Arline*, the school board fired an elementary school teacher because it feared that her recurring bouts of tuberculosis would endanger the health of her students. The Court faced the issue of whether Arline was “otherwise qualified” for the job of elementary school teacher in spite of her possibly contagious tuberculosis. The Court held that under the Rehabilitation Act a disabled individual is not otherwise qualified for a job if he or she poses a significant risk of spreading an infectious disease to others in the workplace and if that risk cannot be eliminated with a reasonable accommodation. Additionally, the Court explained that whether or not one posed a direct threat was a question of fact requiring a trial court to conduct an “individualized inquiry.” Thus, the direct threat defense involves a significantly higher burden of proof than does the business necessity defense.

Like the *Arline* case, the legislative history to the ADA characterizes the direct threat defense only in terms of threats to others, not mentioning threats to self. However, the legislative history suggests that employers could defend a qualification standard requiring employees not to risk their own health or safety in the workplace. In its discussion of medical examinations, the House Committee on Education and Labor explained that if a medical exam revealed a “high probability of substantial harm if the candidate performed the particular functions of the job in question, the employer could reject the candidate, unless the employer could make

55. Id. § 12,111(3).
58. Id. at 276.
59. Id. at 275.
60. Id. at 287–88.
61. Id.
63. See Wong, supra note 52, at 1148–51.
The phrase “substantial harm” appears to include harm to self. As an example of the kind of “substantial harm” that a medical exam might reveal, the Committee mentioned exposures to toxic and hazardous substances that could have a “negative effect” on the individual worker. Thus, Congress seems to have contemplated an interpretation of the direct threat defense that included threats to the disabled individual.

D. The EEOC Interpretation of the Direct Threat Defense

1. EEOC Regulations

Following Congress’s mandate, the EEOC has issued substantive regulations implementing Title I of the ADA. In these regulations, the direct threat defense covers threats to the health or safety of the disabled individual and others. To defend the use of exclusionary safety requirements, employers must utilize the direct threat defense, rather than the lenient business necessity defense. The employer must make an “individualized assessment of the individual’s present ability to safely perform the essential functions of the job.” The assessment must be based on either the best available objective evidence or a reasonable medical judgment. When assessing this evidence, courts must consider the duration of the risk, the nature and severity of the potential harm, the likelihood that the harm will occur, and the imminence of the harm. With such strict requirements, the EEOC aims to prevent employers from

65. Id. at 74, reprinted in 1990 U.S.C.C.A.N. at 357.
67. 29 C.F.R. § 1630.15(2) (2000) (stating that “qualification standard” may include the “requirement that an individual shall not pose a direct threat to the health or safety of the individual or others in the workplace”); id. § 1630.2(r) (stating that “direct threat means a significant risk of substantial harm to the health or safety of the individual or others that cannot be eliminated or reduced by reasonable accommodation”).
68. Id. Pt. 1630 App. § 1630.15(b) & (c).
69. Id. § 1630.2(r).
70. Id.
71. Id.
using the direct threat defense as a loophole for unlawful discrimination.\footnote{See Equal Employment Opportunity for Individuals with Disabilities, 56 Fed. Reg. 35,730 (July 26, 1991) (codified as amended at 29 C.F.R § 1630.2(r)).}

2. **Case Law Interpreting the EEOC Regulations**

Prior to *Echazabal v. Chevron*,\footnote{226 F.3d 1063 (9th Cir. 2000), petition for cert. filed, 69 U.S.L.W. 3619 (U.S. Mar. 9, 2001) (No. 00-1406).} the courts of appeals that had considered the direct threat defense agreed that it included threats to self as well as to others.\footnote{Two cases analyzing the “direct threat defense” have been published since the Ninth Circuit rendered its decision in *Echazabal*. In *Borgialli v. Thunder Basin Coal Co.*, 225 F.3d 1284 (10th Cir. 2000), the Tenth Circuit utilized the EEOC definition of direct threat in holding that an employee who “wasn’t sure if he was going to hurt someone or hurt himself” was not “qualified” under the ADA. \footnote{Id. at 1290. In *Kalskett v. Larson Mfg. Co.*, No. C99-3079 MWB, 2001 WL 630024 (N.D. Iowa June 1, 2001), a district court in Iowa held that, although the direct threat defense only includes threats to others, an employee who poses a direct threat to himself or herself is not “qualified” under the ADA. \footnote{Id. at *21-22. Because this Note argues that *Echazabal’s* analysis was wrong in light of the authority existing at the time of the decision, these cases are beyond the scope of this Note.}} The First, Fifth, and Eleventh Circuits have all relied upon the EEOC regulation that defines the direct threat defense to include threats to the individual.\footnote{See *LaChance v. Duffy’s Draft House, Inc.*, 146 F.3d 832, 835 (11th Cir. 1998); EEOC v. Amigo, Inc., 110 F.3d 135 (1st Cir. 1997) (upholding trial court’s determination that suicidal employee was not qualified to perform essential job function of administering and monitoring residents’ medication); Moses v. Am. Nonwovens, Inc., 97 F.3d 446, 447 (11th Cir. 1996); Daugherty v. City of El Paso, 56 F.3d 695, 698 (5th Cir. 1995).} Finally, the Fifth Circuit has held that safety requirements are permissible qualification standards under the ADA.\footnote{EEOC v. Exxon, 203 F.3d 871 (5th Cir. 2000).} In *Moses v. American Nonwovens, Inc.*,\footnote{Id. at 447.} the Eleventh Circuit held that the direct threat defense included both threats to self and threats to others. Moses, an epileptic man, was discharged because his employer deemed an assembly-line job to be hazardous to an employee with a seizure disorder.\footnote{Id. at 447–48. The job required Moses to perform a number of tasks: as a “product inspector,” Moses would sit above fast-moving press rollers; as a “web operator,” he sat beneath a conveyor belt; as a “Hot Splicer Assistant,” he worked beside extremely hot exposed machinery. \footnote{Id. at 447.}} The trial court granted summary judgment for the employer on the ground that Moses could not prove that he was not a direct threat.\footnote{Id. at 447.} The court of appeals affirmed this decision, holding that
an employer could discharge a disabled employee if the disability posed a direct threat to his or her own health or safety on the job.\textsuperscript{80}

Another Eleventh Circuit case, \textit{LaChance v. Duffy’s Draft House, Inc.},\textsuperscript{81} involved a man with epilepsy who was discharged from his job as a line cook because he could not perform the job safely.\textsuperscript{82} The job involved cooking on a gas flat top grill, using a “fryolater” filled with hot grease, and operating slicing machines.\textsuperscript{83} The district court granted summary judgment in favor of the defendant, finding LaChance unqualified because he could not do the job without risking harm to himself or others.\textsuperscript{84} Affirming the district court’s decision, the Eleventh Circuit Court of Appeals agreed that LaChance’s seizures posed a danger both to himself and to others, rendering him a direct threat.\textsuperscript{85}

As in \textit{Moses} and \textit{LaChance}, Fifth Circuit case law regarding the direct threat defense accepts the EEOC guidelines as a correct interpretation of the law. In \textit{Daugherty v. City of El Paso},\textsuperscript{86} the Fifth Circuit found, as a matter of law, that an insulin-dependant diabetic was not qualified to drive a city bus because the disability posed a “genuine substantial risk” of injury to himself or others.\textsuperscript{87} While the court was not confronted with a situation where a disabled individual posed a threat only to his or her own safety, the court suggested that it would recognize an employer defense on this basis.\textsuperscript{88} The court reasoned that because the Rehabilitation Act term “qualified individual” includes a personal safety requirement and because Rehabilitation Act standards apply to the ADA, the ADA necessarily permits employers to have the same personal safety requirement.\textsuperscript{89}

The Fifth Circuit also allows an employer to defend safety-based qualification standards either under the direct threat defense or the more lenient business necessity defense. In \textit{EEOC v. Exxon Corp.},\textsuperscript{90} the Fifth

\begin{itemize}
\item \textsuperscript{80} Id.
\item \textsuperscript{81} 146 F.3d 832 (11th Cir. 1998).
\item \textsuperscript{82} Id. at 834.
\item \textsuperscript{83} Id.
\item \textsuperscript{84} Id.
\item \textsuperscript{85} Id. at 836.
\item \textsuperscript{86} 56 F.3d 695 (5th Cir. 1995).
\item \textsuperscript{87} Id. at 698.
\item \textsuperscript{88} See id.
\item \textsuperscript{89} Id.
\item \textsuperscript{90} 203 F.3d 871 (5th Cir. 2000).
\end{itemize}
Circuit examined a job requirement at Exxon that prevented anyone with a history of substance abuse from working in certain high-risk jobs. First, the court noted that Congress had used permissive language to designate the direct threat defense as an example of one discrete qualification standard permitted by the ADA. As such, the Exxon court suggested that the permissive phrase “may include . . . direct threat[s] to others” would not prevent employers from having qualification standards that mandate personal safety. Second, the court permitted Exxon to use the business necessity defense, as opposed to the direct threat defense, to defend its refusal to hire ex-substance abusers. In doing so, the court reasoned that safety requirements were “not exclusively cabined into the direct threat test.”

Before Echazabal, the only court to hold that the direct threat defense did not include threats to the disabled individual was a district court in the Seventh Circuit. In Kohnke v. Delta Airlines, the court reasoned that the EEOC’s broad interpretation of direct threat would render meaningless the statutory language describing threats to the health or safety of “other individuals.” However, Kohnke did suggest that potential harm to self could be relevant to a defense of business necessity. Specifically, the court suggested that if the employer found that a particular disability gave rise to injuries on the job, that fact could serve as evidence that an exclusionary qualification standard was “job related and consistent with business necessity.”

II. UNDER THE REHABILITATION ACT, EMPLOYERS COULD DISCHARGE EMPLOYEES WHO RISKED INJURY OR ILLNESS ON THE JOB

Prior to the ADA, case law and regulations under the Rehabilitation Act allowed employers to exclude disabled individuals who posed a risk

91. Id. at 872.
92. Id. at 873; (citing 42 U.S.C. § 12,113(b) (2000) (stating that qualification standards “may include” requirement that employees not pose direct threat to others)).
93. Id.
94. Id. at 875.
95. Id. at 873.
97. Id. at 1111 (citing 42 U.S.C. § 12,113(b)(1994)).
98. Id. at 1113.
99. Id. (citing 42 U.S.C. § 12,113(a)(1994)).
of injury to themselves or future illness while working on the job. Because much of the Rehabilitation Act was incorporated in the ADA, courts often look to Rehabilitation Act law to interpret the ADA.

A. The Rehabilitation Act of 1973 Contributed Significantly to the ADA

Enacted in 1973, the Rehabilitation Act\textsuperscript{100} was the predecessor to the ADA.\textsuperscript{101} Unlike the ADA, which prohibits discrimination in both public and private sectors,\textsuperscript{102} the Rehabilitation Act targets discrimination against disabled individuals only in programs and activities run by a federal executive agency, the United States Postal Service, or any organization receiving federal funds.\textsuperscript{103} Like the ADA, the Rehabilitation Act aims to guarantee equal opportunity to disabled individuals.\textsuperscript{104} To achieve these ends, the Rehabilitation Act aims to level the playing field for disabled persons who, despite a disability, have all the qualifications necessary for a given job. Specifically, Section 504 of the Rehabilitation Act prohibits employers from discriminating “solely by reason of . . . disability [against] otherwise qualified individual[s] with a disability.”\textsuperscript{105} Similar to the ADA, the Rehabilitation Act does not force employers to hire or retain disabled individuals who are not “otherwise qualified” for the job.\textsuperscript{106} The Rehabilitation Act’s legislative history provides evidence of Congress’s intent to balance the realities of employment and the needs of the disabled.\textsuperscript{107}

Rehabilitation Act case law applies to the direct threat defense because the ADA expressly requires agencies such as the EEOC to enforce the ADA in a manner that does not conflict with Rehabilitation

\textsuperscript{101} H.R. REP. NO. 101-485, pt. 3, at 26 (1990), reprinted in 1990 U.S.C.C.A.N 445, 449 (stating that ADA “completes the circle begun in 1973 with respect to persons with disabilities by extending to them the same civil rights protections provided to women and minorities beginning in 1964”).
\textsuperscript{104} See id. § 701(b)(1)(F).
\textsuperscript{105} Id. § 794(a).
\textsuperscript{106} See, e.g., S.E. Cmty. Coll. v. Davis, 442 U.S. 397, 406 (1979) (defining “otherwise qualified person” as one who meets all of a program’s requirements in spite of disability).
\textsuperscript{107} See 124 Cong. Rec. 5950 (1978) (statement of Representative Hyde) (“The Congress needs to give thoughtful and wide-ranging consideration to the needs of such handicapped persons, balanced against the realities of economics, public safety, and commonsense.”).
Act standards. The ADA’s legislative history indicates that Congress intended courts and administrative agencies to utilize Rehabilitation Act precedent when interpreting the ADA. The legislative history also indicates that Congress intended that the ADA concept of a qualified individual with a disability be comparable to the definition found in the Rehabilitation Act regulations. Although Congress amended the Rehabilitation Act in 1992 to provide that the same standards apply to both it and the ADA, the amendment did not limit the applicability of Rehabilitation Act precedent to ADA claims. Courts in eleven circuits continue to apply Rehabilitation Act case law when interpreting the ADA.

B. Agency Regulations Define “Otherwise Qualified” Under the Rehabilitation Act

Two executive agencies have determined that employees who risk injury to themselves or others are not “otherwise qualified” for protection under the Rehabilitation Act. Although the Rehabilitation Act itself does not define the phrase “otherwise qualified,” Congress gave all executive agencies an express grant of authority to implement the terms of the Act. Both EEOC and Department of Labor (DOL) regulations

---

112. See, e.g., Gaston v. Bellingrath Gardens & Home, Inc., 167 F.3d 1361, 1363 (11th Cir. 1999); Francis v. City of Meriden, 129 F.3d 281, 284 n.4 (2d Cir. 1997); Crawford v. Ind. Dep’t of Corr., 115 F.3d 481, 483 (7th Cir. 1997); Andrews v. Ohio, 104 F.3d 803, 807 (6th Cir. 1997); Yin v. California, 95 F.3d 864, 867 (9th Cir. 1996); Allison v. Dep’t of Corr., 94 F.3d 494, 497 (8th Cir. 1996); Katz v. City Metal Co., 87 F.3d 26, 31 n.4 (1st Cir. 1996); McDonald v. Pa. Dep’t of Pub. Welfare, 62 F.3d 92, 95 (3d Cir. 1995); Daugherty v. City of El Paso, 56 F.3d 695, 697–98 (5th Cir. 1995); Bolton v. Scrivner, Inc., 36 F.3d 939, 942–43 (10th Cir. 1994); Tyndall v. Nat’l Educ. Ctrs., Inc. of Cal., 31 F.3d 209, 213 n.1 (4th Cir. 1994).
113. See 29 U.S.C. §§ 705(20), 794 (1994 & Supp. 2000). Although § 794 speaks of an “otherwise qualified individual with a disability . . . as defined in section 705(20),” § 705(20) deals solely with the definition of an “individual with a disability.” Id. The “otherwise qualified” language is not defined in this, or any other, section of the statute.
have permitted employers to use safety-based criteria to determine whether an employee is qualified.\textsuperscript{115} The EEOC defined “qualified individual with handicaps” as someone who could perform all essential job functions without risking the health or safety of the individual or others.\textsuperscript{116} Similarly, the DOL guidelines for “job qualifications” permit exclusionary qualification standards where the qualifications are job-related and consistent with both business necessity and “safe performance.”\textsuperscript{117} The phrase “safe performance” is susceptible to a broad interpretation, which can allow employers to consider the safety of both the disabled individual and others. Thus, according to the plain language of these two regulations, the Rehabilitation Act appears to permit employers to mandate personal safety as a job requirement.\textsuperscript{118}

C. The Ninth Circuit Has Recognized a Defense Based on Personal Safety Under the Rehabilitation Act

The Ninth Circuit has twice recognized a narrow defense under the Rehabilitation Act for employers who exclude from the workplace individuals who pose health or safety threats to themselves on the job. In \textit{Bentivegna v. DOL},\textsuperscript{119} the Ninth Circuit examined the DOL regulation permitting job requirements that exclude disabled individuals because of their disability when necessary for “safe performance.”\textsuperscript{120} Bentivegna, a diabetic, repaired buildings for the City of Los Angeles.\textsuperscript{121} Concerned for the personal safety of its workforce,\textsuperscript{122} the City adopted a requirement that all diabetic employees must “control” their blood sugar levels.\textsuperscript{123}

\begin{footnotesize}
\begin{enumerate}
\item[116.] Id. § 1614.203(a)(6).
\item[117.] Id. § 32.14.
\item[118.] See id. §§ 32.14, 1614.203(a)(6). Some agencies discussed employment qualifications without reference to safety, instead focusing on the “essential functions” of the job; however, these agencies never explicitly prohibited employers from considering threats to safety. See, e.g., id. § 85.3; 24 C.F.R. § 8.3.
\item[119.] 694 F.2d 619 (9th Cir. 1982).
\item[120.] Id. at 621 (citing 29 C.F.R. § 32.14(b) (1982)).
\item[121.] Id. at 620.
\item[122.] Id. at 622.
\item[123.] Id. at 620.
\end{enumerate}
\end{footnotesize}
Pursuant to this rule, the City discharged Bentivegna when he failed to control his blood sugar level.\textsuperscript{124} The court concluded that as an employer, the City could consider its employees’ risk of future injury or long-term health problems, but that it must present adequate evidence of a direct connection between the disability and the safety concern to rely on the defense.\textsuperscript{125} In Bentivegna’s case, the court held that no direct connection existed between the control requirement and the employer’s concern for either business necessity or safe job performance.\textsuperscript{126}

Two years later, in \textit{Mantolete v. Bolger},\textsuperscript{127} the Ninth Circuit created a strict legal standard to determine when a particular safety requirement could disqualify a disabled individual. After a pre-employment physical revealed that Mantolete had epilepsy, the Post Office denied her a job operating a letter-sorting machine out of concern that the machine’s flashing lights might trigger seizures and injure her.\textsuperscript{128} The court agreed that a job requirement could screen out a qualified disabled individual on the basis of a possible future injury, but only when the employer established a “reasonable probability of substantial harm”\textsuperscript{129} based on the applicant’s work and medical histories.\textsuperscript{130} Even though the Rehabilitation Act regulations permitted employers to exclude at-risk disabled individuals from the workplace,\textsuperscript{131} the Ninth Circuit limited the scope of the defense by requiring proof of potential harm.\textsuperscript{132} In this way, the court helped ensure that safety-based qualification standards would not be used as a means for employers to discriminate against disabled individuals.

**III. STATE WORKPLACE SAFETY LAWS REQUIRE EMPLOYERS TO PREVENT KNOWN RISKS IN THE WORKPLACE OR FACE CRIMINAL SANCTIONS**

If an employer knows that a disabled employee faces an increased risk of personal injury on the job and that employee is subsequently injured,
the employer could face criminal liability under state workplace safety laws. Nearly all of the states in the Ninth Circuit have strict safety laws that impose harsh criminal sanctions on employers who are in violation of these codes.133 Because Echazabal discussed California’s safety laws,134 this Section focuses primarily on the California Labor Code.

The California Labor Code requires employers to provide safeguards for employees and to do everything else “reasonably necessary” to protect them.135 Because the term “safeguards” is defined broadly, employers are expected to use “any practicable method” to protect employees.136 Moreover, the Code seems to require employers to consider the safety of not just the average employee, but each employee individually.137 An employer’s failure to take necessary precautions to protect its employees’ lives, safety, and health is a code violation.138

The California Labor Code further requires employers to prevent employees from going into or working in an environment that “is not safe and healthful.”139 The Code places the primary responsibility for the

---

133. Alaska, Arizona, California, Hawaii, Montana, Nevada, Oregon, and Washington have adopted the same general workplace safety requirement: employers must furnish employment and a workplace that is safe and free from recognized hazards. ALASKA STAT. § 18.60.075 (Michie 2000); ARIZ. REV. STAT. ANN. § 23-403 (West 2000); CAL. LAB. CODE § 6403 (West 2000); HAW. REV. STAT. ANN. § 396-6 (Michie 2000); MONT. CODE ANN. § 618.375(1) (Michie 1999); OR. REV. STAT. § 654-010 (1999); WASH. REV. CODE ANN. § 49.17.060 (West 2000). In these states (with the exception of Montana), if an employer knowingly violates the general safety requirements and an employee dies as a result, the employer can be convicted of a felony or misdemeanor and is subject to a fine and jail time. ALASKA STAT. § 18.60.095; ARIZ. REV. STAT. ANN. § 23-418; CAL. LAB. CODE §§ 6423, 6432(a); HAW. REV. STAT. ANN. § 396-10; OR. REV. STAT. § 654-991(1); WASH. REV. CODE ANN. § 49.17.190(3).


136. Id. § 6306(b) (2000).

137. Id. § 6309 (stating that administrative agency may take action if it “learns or has reason to believe that any employment or place of employment is not safe or is injurious to the welfare of any employee”) (emphasis added).

138. Id. § 6403:

No employer shall fail or neglect to do any of the following: (a) To provide and use safety devices and safeguards reasonably adequate to render the employment and place of employment safe; (b) To adopt and use methods and processes reasonably adequate to render the employment and place of employment safe; (c) To do every other thing reasonably necessary to protect the life, safety, and health of employees.

139. Id. § 6402.
safety of workers upon the employer. As a result, the law is necessarily paternalistic because it requires that the employer not permit any employee to work in an unsafe environment—even if that employee wants to work there.

An employer who violates California’s Labor Code safety provisions can be charged with a misdemeanor. If an employee dies as a result of such a violation, his or her employer can be charged with involuntary manslaughter. The employer could also be charged with involuntary manslaughter if it places an employee in a position “which might produce death . . . without due caution and circumspection,” and the employee dies as a result. If an employer knows that a disabled employee faces a significant risk of death on the job, these laws may require the employer to keep the employee out of the workplace or face criminal sanctions.

IV. COURTS USE COMMON TOOLS OF STATUTORY CONSTRUCTION WHEN DECIDING WHETHER TO REJECT AN ADMINISTRATIVE REGULATION

The Chevron doctrine, coupled with traditional rules of statutory construction, helps courts decide whether to follow an agency’s interpretation of a statute. In Chevron USA, Inc. v. Natural Resources Defense Council, Inc., the United States Supreme Court created a two-step process by which a court can determine how much deference to give

140. Bendix Forest Prod. Corp. v. Div. of Occupational Safety & Health, 600 P.2d 1339, 1343–44 (Cal. 1979) (holding that Division of Industrial Safety had authority to issue order requiring employer to furnish protective gear at employer’s expense).

141. See CAL. LAB. CODE § 6423 (stating that employer commits misdemeanor by “[k]nowingly or negligently” committing “serious violation” of labor code); id. § 6432 (stating that “serious violation” exists if a “substantial probability that death or serious physical harm could result.”) (internal quotation marks omitted).

142. CAL. PENAL CODE § 192(b) (2000); see People v. Cleaves, 280 Cal. Rptr. 146, 155 (Cal. Ct. App. 1991) (defining involuntary manslaughter as commission of act involving high degree of risk of death or great bodily injury, committed with criminal negligence, and defining criminal negligence as departure from conduct of ordinarily prudent person under same circumstances so as to be incompatible with proper regard for human life); Granite Constr. Co. v. Superior Court, 197 Cal. Rptr. 3, 7–8, 149 Cal. App. 3d 465, 471–72 (Cal. Ct. App. 1983) (finding that corporate employer may be prosecuted for manslaughter in death of employees under existing law).


an agency regulation. In step one of the process, the court decides “whether Congress has directly spoken to the precise question at issue.” The court will proceed to the second step of the Chevron analysis only if it finds an ambiguity in the statute. Step two of the process requires the court to determine whether the agency regulation is “permissible.”

In the first step of the Chevron analysis, a court examines the statute to decide if the language clearly resolves the dispute at issue or if there is an ambiguity in the text. If the court finds that Congress has spoken directly to that issue, both the court and the agency must give effect to that unambiguous statement of Congress and there the matter ends. “A statute is ambiguous if it gives rise to more than one reasonable interpretation.” When assessing such ambiguities, courts must examine “the language of the statute” by reading the statute as a whole, because the meaning of the text, whether it appears plain or not, depends on context.

The circumstances underlying the enactment of a particular piece of legislation may persuade a court that Congress did not intend the statutory language to be taken literally. For example, where a literal interpretation of a statute produces absurd results, courts should avoid such an interpretation. In McNely v. Ocala Star-Banner Corp., the Eleventh Circuit utilized this cannon of construction when it decided that a narrow reading of the ADA produced absurd results. Specifically, the Eleventh Circuit held that the ADA’s prohibition of discrimination “because of” a disability should not be interpreted as prohibiting

---

145. Id. at 842–43.
146. Id. at 842.
147. Id. at 842–43.
148. Id. at 843.
149. Id. at 842.
150. Id. at 842–43.
155. 99 F.3d 1068 (11th Cir. 1996).
156. Id. at 1075.
discrimination “solely because of” a disability. The court reasoned that if the ADA prohibited only discrimination that was “solely because of” a disability, an employer could fire an individual who was both black and disabled without incurring any liability under the ADA.

A court reaches step two of the Chevron analysis only if it has determined that Congress did not directly address the disputed issue. If an ambiguity exists within the statute, the court cannot impose its own interpretation of the statute. Instead, the court is to defer to the agency’s interpretation, as long as that interpretation is “based on a permissible construction of the statute.” If Congress expressly delegated authority to an agency, a court must give agency regulations controlling weight unless the regulations are “arbitrary, capricious, or manifestly contrary to the statute.”

V. IN ECHAZABAL V. CHEVRON, THE NINTH CIRCUIT REJECTED THE EEOC’S DEFINITION OF DIRECT THREAT

The Ninth Circuit, applying the Chevron doctrine, decided not to defer to the EEOC interpretation of the direct threat defense in Echazabal v. Chevron. The court held that neither the direct threat defense nor the business necessity defense allow an employer to exclude disabled individuals who pose a direct threat to themselves on the job. The United States Supreme Court has not yet decided whether it will grant certiorari in this case.

A. Facts and Procedural History

In Echazabal, the Ninth Circuit reviewed a summary judgment ruling that Mario Echazabal had no valid ADA claim against his employer. Echazabal worked as a contract laborer at Chevron’s oil refinery in El...
Segundo, California between 1972 and 1996. In 1992, Echazabal applied to work directly for Chevron in the same capacity and Chevron offered him a job, “contingent on his passing a physical examination.” A pre-employment physical exam revealed a liver condition that was later diagnosed as hepatitis C. Chevron’s doctors determined that by allowing Echazabal to continue working with the liver-toxic solvents and chemicals present on the job, Chevron would be putting his life in danger. Accordingly, Chevron rescinded its offer, though Echazabal remained at the refinery as a contract laborer for the next three years.

Mr. Echazabal again applied for a permanent position in 1995. As before, Chevron extended the same conditional offer, contingent on the outcome of a physical exam. Again, Chevron rescinded its offer on finding that Echazabal’s liver might not be able to handle the toxins present in the refinery. Ultimately, as a result of the exam, Mr. Echazabal lost his job at the refinery.

After losing his job, Mr. Echazabal filed a complaint with the EEOC and brought suit in state court. He alleged that Chevron and his contract employer had discriminated against him based on his disability, and in doing so violated the ADA, the Rehabilitation Act, and California’s Fair Employment and Housing Act. Chevron removed the case to federal district court and moved for summary judgment on all claims. Although the district court granted summary judgment in favor of Chevron, it certified the case for appeal to the Ninth Circuit Court of Appeals.

165. Id.
166. Id.
167. Id.
168. Id. at 1073 (Trott, J., dissenting) (stating that toxins present on the job included “hydrocarbon liquids and vapors, acid, caustic, refinery waste water and sludge, petroleum solvents, oils, greases, and chlorine bleach”) (internal quotation marks omitted).
169. Id.
170. Id. at 1065.
171. Id.
172. Id.
173. Id.
174. Id.
175. Id. at 1065 & n.1.
176. Id.
177. Id.
B. Holding and Rationale

In *Echazabal*, the Ninth Circuit questioned “whether the direct threat defense includes threats to one’s own health or safety.”\(^{178}\) The issue was one of first impression in the Ninth Circuit and, at the time, had received little treatment in other Circuits.\(^ {179}\) The court noted that several cases had stated, in dicta, that direct threats include threats to the protected individual.\(^ {180}\) After briefly mentioning these cases, the court turned to the one case that had explicitly held that direct threats include threats to one’s own health or safety: *Moses v. American Nonwovens*.\(^ {181}\) The court, however, declined to follow *Moses*, on the ground that *Moses* offered no analysis to support its holding.\(^ {182}\)

After rejecting all of the case law on point, the court attempted to resolve the scope of the direct threat defense by examining the language of the statute.\(^ {183}\) The court stated that the language itself was “dispositive” and reasoned that because the statute specified direct threats to others, Congress meant to limit the scope of the defense to others only.\(^ {184}\)

Next, the court invoked the legislative history of the ADA to support its conclusion that the direct threat provision does not include threats to disabled individuals.\(^ {185}\) The court noted that the term direct threat was used hundreds of times in the legislative history, and in almost every case was accompanied by a reference to threats to “others.”\(^ {186}\) The court found additional support in committee reports, which indicated that the direct threat provision was meant to codify the standard set forth in *School Board of Nassau County v. Arline*,\(^ {187}\) a case that involved only threats to others.\(^ {188}\) Finally, the court referred to a comment made by the ADA’s co-sponsor, Senator Kennedy, stating that the ADA “specifically

\(^ {178}\) Id. at 1066.
\(^ {179}\) See id.
\(^ {180}\) Id. (citing LaChance v. Duffy’s Draft House, Inc., 146 F.2d 832 (11th Cir. 1998); EEOC v. Amigo, 110 F.3d 135 (1st Cir. 1997); Daugherty v. City of El Paso, 56 F.3d 695 (5th Cir. 1995)).
\(^ {181}\) Id. (referencing Moses v. Am. Nonwovens, Inc., 97 F.3d 446, 447 (11th Cir. 1996)).
\(^ {182}\) Id.
\(^ {183}\) Id.
\(^ {184}\) Id. at 1066–67.
\(^ {185}\) Id. at 1067.
\(^ {186}\) Id. (internal quotation marks omitted).
\(^ {187}\) 480 U.S. 273 (1987); see supra notes 57–62 and accompanying text.
\(^ {188}\) *Echazabal*, 226 F.3d at 1067.
Echazabal v. Chevron

refers to health and safety threats to others.” Finding no ambiguity under the first prong of the *Chevron* analysis, the court held that Congress clearly intended to limit the direct threat defense to threats to others, and rejected the EEOC’s contrary interpretation.

In addition to striking down a defense based on direct threats to the disabled individual, the court also held that an employer could not use the business necessity defense to justify excluding disabled employees who risked injury or illness. Initially, the court reasoned that the ability to perform a job without posing a threat to one’s own health or safety was not an “essential” function of any job. The court deemed such qualification standards “paternalistic” and inconsistent with the ADA’s underlying policy. The court declined to follow the reasoning of *EEOC v. Exxon Corp.*, which held that an employer could defend general safety requirements using the business necessity defense. The court chose not to follow the case, stating only that the case’s factual scenario involved threats to others, not threats to individuals.

Next, the court refused to extend the Ninth Circuit Rehabilitation Act holdings of *Mantolete v. Bolger* and *Bentivegna v. DOL*, which established that personal safety requirements were valid qualification standards. Because the Rehabilitation Act did not define “qualified handicapped person,” the *Mantolete* and *Bentivegna* courts had applied an EEOC regulation that permitted them to consider safety of the disabled individual as a qualification standard. However, the *Echazabal* court noted that the ADA already had a definition of “qualified individual with a disability” so there was no need to look to Rehabilitation Act regulations or case law defining that term. Noting that the ADA’s definition of “qualified individual with a disability” did

189. *Id.* (quoting 136 CONG. REC. S9684–03, at S9697 (1990)).

190. *Id.* at 1069.

191. *Id.* at 1072.

192. *Id.* at 1071–72.

193. *Id.* at 1068, 1071.

194. 203 F.3d 871, 873–75 (5th Cir. 2000).

195. *Echazabal*, 226 F.3d at 1072 n.11 (citing *Exxon*, 203 F.3d at 873–75).

196. *Id.*

197. 767 F.2d 1416 (9th Cir. 1984); see supra notes 127–32 and accompanying text.

198. 694 F.2d 619 (9th Cir. 1982); see supra notes 119–26 and accompanying text.

199. See *Echazabal*, 226 F.3d at 1071 n.10.

200. *Id.* (internal quotation marks omitted).

201. *Id.* (internal quotation marks omitted).
not mention threats to the individual, the *Echazabal* court concluded that the ADA definition superceded the Rehabilitation Act’s regulatory definition, rendering both *Mantolete* and *Bentivegna* inapplicable to the ADA.\(^{202}\)

Finally, the court dismissed Chevron’s argument that employers would face tort liability if required to hire at-risk disabled individuals, stating that the issue was not properly before it.\(^{203}\) The court briefly mentioned *International Union, UAW v. Johnson Controls*,\(^{204}\) suggesting that state tort law would likely be preempted where it interfered with federal anti-discrimination law.\(^{205}\) The court, however, did not discuss whether state criminal laws would be preempted by its interpretation of the ADA.

In a strongly worded dissent, Judge Trott questioned the court’s “bizarre” holding that Mr. Echazabal was “qualified,” because performing the essential functions of his job could kill him.\(^{206}\) Judge Trott asserted that the court’s prohibition of paternalism in the workplace was “pernicious” when used to displace longstanding workplace safety laws protecting employees.\(^{207}\) Judge Trott further noted that the holding would lead to absurd results: those very workers known to be in danger would receive the least amount of protection because employers would be reluctant to exclude them from dangerous work.\(^{208}\) Chevron has since petitioned the United States Supreme Court for certiorari to resolve the issue of whether the direct threat defense includes threats to disabled individuals.\(^{209}\)

VI. THE *ECHAZABAL* COURT SHOULD HAVE DEFERRED TO THE EEOC’S REGULATIONS

By rejecting the EEOC’s interpretation of direct threat that allowed employers to consider the personal health and safety of disabled

---

\(^{202}\) Id.

\(^{203}\) Id. at 169–70.

\(^{204}\) 499 U.S. 187, 206 (1991) (holding that bona fide occupational qualification cannot exclude all members of protected class out of paternalistic concern for welfare of protected class).

\(^{205}\) *Echazabal*, 226 F.3d at 1069 (citing *Johnson Controls*, 499 U.S. at 210).

\(^{206}\) Id. at 1073–74 (Trott, J., dissenting).

\(^{207}\) Id. at 1074 (Trott, J., dissenting).

\(^{208}\) Id.

\(^{209}\) Id.
applicants, the *Echazabal* court completely disregarded both precedent and workplace safety concerns. The court erred during step one of the *Chevron* analysis by determining that the statute’s “plain language” prevented employers from defending a discrimination claim on the grounds that the plaintiff posed a safety threat to himself or herself. Because the ADA’s language is ambiguous on this issue, the court should have proceeded to step two of the *Chevron* analysis, which requires the court to defer to an agency regulation unless that regulation is arbitrary, capricious, or manifestly contrary to the statute. Because the EEOC interpretation of direct threat as including threats to disabled individuals is a permissible reading of the statute, the court should have deferred to the regulation. Finally, the court’s utter disregard for a job-applicant’s personal safety runs counter to existing workplace safety laws and forces employers to risk violating either the ADA or state criminal and labor laws.

A. The Plain Language of the ADA Is Ambiguous on the Topic of Direct Threats to Disabled Individuals

Step one of the *Chevron* process required the *Echazabal* court to examine the ADA’s text to determine whether or not Congress intended to permit the exclusion of at-risk disabled individuals from the workplace.\(^{210}\) From the statutory text, it is unclear whether Congress meant the direct threat defense to be the ADA’s exclusive safety-based defense, or whether the direct threat defense was simply an example of the kind of qualification standard permitted by the ADA. Had it adhered to the traditional tools of statutory construction, the *Echazabal* court would have discovered this ambiguity and, as a result, would have deferred to the EEOC’s interpretation of direct threat.

The language of the statute suggests that the direct threat provision is an example of one, but not the only, safety-based qualification standard that the ADA allows.\(^{211}\) Initially, the statute provides that employers can adopt exclusionary qualification standards when those standards are job-related and consistent with business necessity.\(^{212}\) Because employers must obey state workplace safety laws that make it a crime to subject

---


\(^{211}\) Appellee’s Petition for Rehearing with Suggestions for Rehearing En Banc at 11, *Echazabal v. Chevron USA, Inc.*, 226 F.3d 1063 (9th Cir. 2000) (No.98-55551).

employees to known dangers on the job, safety is arguably both job-related and consistent with business necessity. Thus, an employer could potentially defend all safety-related requirements under the business necessity defense. The statute lists the direct threat defense as merely an example of one acceptable qualification standard, using the permissive language that employers “may” impose a requirement that employees not pose a direct threat to others. Contrary to Chevron’s interpretation, this example need not prevent an employer from defending other safety-based qualification standards using the business necessity defense.

The Fifth Circuit, in EEOC v. Exxon, recognized the effect of the permissive “may” and concluded that safety requirements are not “exclusively cabined into the direct threat test.” In effect, the Exxon court held that if a safety requirement did not fall within the direct threat defense, an employer could alternatively defend that requirement as a business necessity. Although Echazabal disregarded this case because it dealt with threats to others, the Exxon court suggested its holding should apply to all safety-based qualification standards—those protecting the disabled individual and others.

The Echazabal court avoided this textual ambiguity by failing to follow proper canons of statutory interpretation. First, the court focused entirely on the direct threat provisions to determine the plain meaning of the statute. The plain meaning of the statute must be determined by looking at both the language in question and the entire statutory context. The court should have read the provisions related to the direct threat defense in conjunction with the ADA section describing those qualification standards that are consistent with business necessity. If the court had done so, it could have seen the direct threat defense for what it was—one example of an acceptable qualification standard.

213. See supra Section III.
214. 42 U.S.C. § 12,113(b).
215. 203 F.3d 871 (5th Cir. 2000).
216. Id. at 873.
217. See id.
218. See id. at 874–75.
Second, a restrictive reading of the direct threat defense produces the sort of absurd results that courts are supposed to avoid. For example, the *Echazabal* court’s narrow interpretation would force a peanut farm to hire an individual disabled by a severe peanut allergy to work in its nut-packing facility; a grocery store could not legally discharge a disabled butcher with hemophilia; and a window-washer disabled by vertigo could continue working on skyscrapers hundreds of stories above the ground. In his *Echazabal* dissent, Judge Trott expressed the view that Congress could not have intended absurd results when codifying the ADA’s direct threat provision.

Although the *Echazabal* court read the “plain language” of the direct threat defense to prevent employers from excluding individuals at risk of workplace injury, the court could have just as easily concluded that the ADA preserves for employers the right to discharge such employees as a business necessity. In this way, the plain language of the ADA is ambiguous on the issue of whether an employer could exclude a disabled individual who risked harm to himself or herself on the job.

**B. Under the Arbitrary or Capricious Standard, the Echazabal Court Should Have Deferred to the EEOC Guidelines**

The EEOC interpretation of direct threat is permissible because it reconciles the ambiguity in the ADA’s text in a way that is consistent with the ADA’s legislative history, the Rehabilitation Act regulations, and the case law under both the ADA and the Rehabilitation Act. Furthermore, the EEOC interpretation minimizes paternalistic decisions based on stereotypes. The *Chevron* doctrine requires a court to defer to agency regulations that are based on a “permissible” reading of a statute. Because Congress gave the EEOC an express grant of authority to implement Title I of the ADA, the court must accept the

---

222. See McNely v. Ocala Star-Banner Corp., 99 F.3d 1068, 1075 (11th Cir. 1996); supra notes 155–58 and accompanying text.
223. For a list of similar hypotheticals, see Appellee’s Petition for Rehearing with Suggestions for Rehearing En Banc at 13, Echazabal v. Chevron USA, Inc., 226 F.3d 1063 (9th Cir. 2000) (No. 98-55551).
224. *Echazabal*, 226 F.3d at 1074 (Trott, J. dissenting).
EEOC’s regulations unless they are “arbitrary and capricious.” Since the EEOC regulations are compatible with the ADA’s language, legislative history, and case law, the Echazabal court should have deferred to those regulations.

I. The EEOC’s Interpretation of Direct Threat is Consistent with the ADA’s Text, Legislative History, and Case Law

The EEOC interpretation that the direct threat defense includes threats to disabled individuals is permissible because the ADA allows employers to have exclusionary qualification standards that are “job related and are consistent with business necessity.” Ensuring the personal safety of employees is a business necessity, given that employers are required by state law to maintain a safe workplace. Thus, it was not arbitrary or capricious for the EEOC to allow for a qualification standard that employees not pose a direct threat to self or others in the workplace.

The legislative history supports the assertion that the ADA allows employers to impose personal safety requirements on their employees. Although the legislative history normally mentions direct threats within the context of threats to others, the legislative history does explicitly recognize the existence of threats to self. For example, the House Labor Committee asserted that the ADA permits an employer to reject a candidate if a pre-employment medical exam reveals a “high probability of substantial harm if the candidate performed the particular functions of the job in question.” As an illustration of an acceptable medical exam, the Committee mentioned that employers could test to determine whether exposures to toxic or hazardous chemicals have had a “negative effect” on individual employees. Such a test would reveal ailments affecting primarily the health and safety of each individual. While the Echazabal court dismissed the Committee’s discussion because it did not take place

228. 42 U.S.C. § 12,112(b)(6).
229. *See supra Section III.*
230. *See Echazabal v. Chevron USA, Inc.*, 226 F.3d 1063, 1067–68 (9th Cir. 2000), *petition for cert. filed*, 69 U.S.L.W. 3619 (U.S. Mar. 9, 2001) (No. 00-1406) (stating that throughout the legislative history, the term direct threat is accompanied by a reference to “others” or to “other individuals” in the workplace).
“in the context of . . . the direct threat defense,” this legislative history suggests that employers can consider threats to disabled individuals under the business necessity defense.

Finally, case law in the Fifth and Eleventh Circuits supports the EEOC interpretation of the direct threat defense. In Moses v. American Nonwovens, the Eleventh Circuit explicitly held that the direct threat defense includes threats to the disabled individual as well as to others. In Daugherty v. City of El Paso, the Fifth Circuit explained in dicta that because the standards for liability are the same for both the ADA and the Rehabilitation Act, and because the Rehabilitation Act recognized an employer defense based on threats to disabled individuals, the ADA must also recognize that same defense. At the very least, these cases show that the Fifth and Eleventh Circuits have found the EEOC regulations to be reasonable.

2. The EEOC Interpretation of Direct Threat Is Consistent with the Regulations and Ninth Circuit Case Law Under the Rehabilitation Act

Additional support for the EEOC interpretation of direct threat may be found in Rehabilitation Act regulations and Ninth Circuit case law, both of which recognize that a disabled individual might not be “qualified” if he posed a threat to his own health or safety on the job. Because the ADA’s text and legislative history indicate that Rehabilitation Act precedent applies to the ADA, and because the EEOC regulations are consistent with the Rehabilitation Act, the EEOC interpretation of direct threat is not arbitrary or capricious.

Ninth Circuit case law interpreting the Rehabilitation Act established an employer defense for refusing to hire an employee who posed a threat to his own health or safety in the workplace. In Bentivegna v. United States Dep’t of Labor, a Ninth Circuit panel held that in order to prove

---

233. Echazabal, 226 F.3d at 1068 n.6.
234. 97 F.3d 446 (11th Cir. 1996).
235. Id. at 447.
236. 56 F.3d 695 (5th Cir. 1995).
237. Id. at 697–98.
238. See 29 C.F.R. § 1614.203(a)(6) (2000); Mantolete v. Bolger, 767 F.2d 1416, 1422 (9th Cir. 1985); Bentivegna v. U.S. Dep’t of Labor, 694 F.3d 619, 622 (9th Cir. 1982).
239. 694 F.3d 619 (9th Cir. 1982).
an employee unqualified under the “safe performance” requirement, an employer must show a link between the disability and a possibility of future injury or health problem.\textsuperscript{240} In \textit{Mantolete v. Bolger},\textsuperscript{241} the Ninth Circuit again recognized a defense based on a threat to personal safety but limited the defense by requiring the employer to make, in each individual case, an independent medical assessment of the probability and severity of the potential harm.\textsuperscript{242} Thus, the Rehabilitation Act’s regulations and case law provided that employers would be able to exclude disabled individuals who posed a threat to their own personal safety in the workplace.

The \textit{Echazabal} court chose to disregard this case law, concluding instead that the ADA superceded the Rehabilitation Act.\textsuperscript{243} Specifically, the court distinguished the cases on the ground that unlike the Rehabilitation Act, the ADA defined the term “qualified individual with a disability” and the ADA definition did not mention threats to the disabled individual.\textsuperscript{244} However, the ADA’s legislative history evidences the intent that the ADA should mirror the Rehabilitation Act; in particular, the term “qualified individual with a handicap” was meant to be the same as in the regulations implementing the Rehabilitation Act.\textsuperscript{245}

The EEOC regulations implementing the Rehabilitation Act stated that an employee would not be “qualified” under the Rehabilitation Act if he could not perform the essential functions of the job without injury to himself or others.\textsuperscript{246} Likewise, the DOL stated that an employer could have a job requirement that excluded disabled individuals and still be in compliance with the Rehabilitation Act, if that requirement provided for “safe performance.”\textsuperscript{247}

Moreover, in \textit{Daugherty v. City of El Paso},\textsuperscript{248} the Fifth Circuit stated that the Rehabilitation Act regulations and case law would apply to the

\textsuperscript{240} Id. at 622.
\textsuperscript{241} 767 F.2d 1416 (9th Cir. 1985).
\textsuperscript{242} Id. at 1422.
\textsuperscript{243} Echazabal v. Chevron USA, Inc., 226 F.3d 1063, 1071 n.10 (9th Cir. 2000), petition for cert. filed, 69 U.S.L.W. 3619 (U.S. Mar. 9, 2001) (No. 00-1406).
\textsuperscript{244} Id. (internal quotation marks omitted).
\textsuperscript{246} 29 C.F.R. § 1614.203(a)(6) (2000).
\textsuperscript{247} Id. § 32.14.
\textsuperscript{248} 56 F.3d 695 (5th Cir. 1995).
ADA. Since the cases and regulations under the Rehabilitation Act provided for a personal safety requirement, that same requirement should also apply to the ADA. Echazabal should have deferred to the EEOC interpretation of direct threat because the EEOC definition was consistent with both the Rehabilitation Act regulations and case law.

3. The EEOC Interpretation of Direct Threat Reconciles the Statute’s Ambiguous Text in a Way that Minimizes Paternalism

The EEOC interpretation is reasonable because it fairly balances the competing interests of employers who want workplace safety and disabled individuals who want to work. The EEOC regulation prohibits employers from defending safety-based qualification standards under the more lenient business necessity defense, as the district court in Kohinke v. Delta Airlines and the Fifth Circuit in EEOC v. Exxon Corp. suggested could be done. To plead the direct threat defense, an employer must make an individualized assessment of one’s ability to do a job safely, based on real medical evidence. The employer must also show that the risk to self presently exists, that severe harm will result, and that the harm is both likely and imminent. When defending a general qualification standard, however, the employer need only show that the standard is job-related and consistent with business necessity.

By including threats to self within the direct threat defense, the EEOC interpretation prohibits employers from excluding an entire group of people out of a mere business interest in employee safety, instead forcing them to defend such exclusions on a case-by-case basis. The interpretation is consistent with the Rehabilitation Act precedent in Mantolete v. Bolger, which required an independent medical assessment of the likeliness and severity of harm before an employer

249. Id. at 698.
252. 203 F.3d 871 (5th Cir. 2000).
253. Kohinke, 932 F.Supp. at 1113; Exxon, 203 F.3d at 873.
255. See id.
256. See id.
257. See supra Section II.B.1.
258. 767 F.2d 1416 (9th Cir. 1985).
could label an applicant “unqualified.” The EEOC interpretation reconciles concerns about discrimination based on paternalistic stereotypes with an employer’s legitimate need for safety in the workplace. As a result, the EEOC interpretation limits the number of paternalistic decisions that will be made, and supports, rather than undermines, the ADA’s underlying policies.

C. Echazabal Forces Employers to Choose Between Violating the ADA and Violating State Workplace Safety Laws

Recall the hypothetical at the beginning of this Note: A construction worker with vertigo applies for a job at a California-based construction company that builds skyscrapers. Although the employer could try to accommodate this employee with the standard harness and safety-rope, she is still at a greater risk of falling than the average worker. As shown above, California’s Labor Code requires an employer to do everything “reasonably necessary to protect the life, safety, and health of employees.” If the employer cannot prevent this applicant from encountering vertigo-triggering heights on the job, the only way to adequately protect her “life, safety, and health” is not to hire her . . . or so a clever prosecuting attorney might argue, after her client’s untimely death on the job. Under a standard requiring the workplace to be “safe and healthful” for an employee with vertigo, a jury could find the construction company liable for hiring her in the first place.

By misreading the statute and incorrectly rejecting the EEOC interpretation of the direct threat defense, the Echazabal court has placed employers in an untenable position. After Echazabal, the ADA no longer permits employers to reject individual disabled applicants who are at an increased risk of injury, but state law still requires employers to comply with safety requirements. As a result, employers in the Ninth Circuit

259. Id. at 1422–23.
260. See supra note 1 and accompanying text.
261. See supra note 135 and accompanying text.
262. See supra note 138 and accompanying text.
263. The Echazabal court did not address the issue of whether its interpretation of the ADA would preempt state criminal workplace safety laws. See supra Section II.C. While the issue of preemption is beyond the scope of this Note, it is unlikely that the ADA would preempt criminal workplace safety laws because such laws fall under the state’s traditional police power. Moreover, the fact that the employer hired the disabled individual would defeat an argument that the ADA’s purpose, to open employment opportunities to persons with disabilities, had been thwarted by criminal liability.
who follow the Echazabal ruling and disregard the health and safety of disabled job applicants will risk criminal sanctions under state workplace safety laws.

VII. CONCLUSION

The EEOC’s direct threat provision, which permits employers to weed out those employees who are at a heightened risk of on-the-job injury, strikes a balance between the employer’s interest in workplace safety and the interest of disabled individuals to participate in the workforce. The Ninth Circuit, in Echazabal v. Chevron, disrupted that balance by holding that an employer could not, under any circumstance, exclude a disabled individual from work just to protect that person’s health or safety. Under the Chevron doctrine of deference to agency regulations, the court should have deferred to any interpretation of “direct threat” that was reasonable in light of the ADA’s text and legislative history. In its attempt to rid the workplace of paternalism, the court turned its back on the state and federal safety laws that forbid employers from knowingly subjecting workers to life-endangering conditions. As a result, employers now have an ugly choice to make: risk an employee’s life or risk a discrimination suit under the ADA.
CAN STUDENTS SUE WHEN SCHOOLS DON’T MAKE THE GRADE? THE WASHINGTON ASSESSMENT OF STUDENT LEARNING AND EDUCATIONAL MALPRACTICE

Rebecca R. Glasgow

Abstract: Washington’s Academic Achievement and Accountability Statute (AAA Statute) creates a statewide system of school accountability. It also requires that all students pass the tenth-grade level of the Washington Assessment of Student Learning standardized test (WASL) to receive a diploma. Unfortunately, when this graduation requirement takes effect in 2008, many students will not receive diplomas because they will be unable to pass the WASL before graduation. Some of these students will have met all local graduation requirements, so the only graduation requirement they will not be able to meet will be the statewide requirement that they pass the WASL. Their WASL failure will show that their school districts failed to educate them to statewide standards either by inadequately instructing them or by allowing them to pass local assessments even though they lacked essential skills. This Comment explores the viability of educational malpractice claims against school districts for compensatory damages for remedial education when students are denied diplomas based only on failing WASL scores. Such claims, based on the AAA Statute, should succeed under the Washington Supreme Court’s most recent test for determining the existence of an implied private statutory cause of action. These claims should also overcome other legal and policy barriers that have led Washington and other state courts to reject educational malpractice causes of action.

Shelby1 has attended public school in the same school district since kindergarten. Although her grades have rarely been above average, she has never failed a grade or a class. When Shelby took the Washington Assessment of Student Learning standardized test (WASL) in her tenth-grade year, she passed only the reading and listening sections, failing the math and writing sections. Shelby’s failure shows that, even though she met local graduation requirements, she did not acquire the skills necessary to meet statewide standards. Beginning in 2008, students like Shelby, who cannot pass the WASL before graduation, will not receive a diploma.

The WASL is a result of Washington’s extensive attempts to improve the quality of its public education system. The Washington Legislature passed the Academic Achievement and Accountability Statute (AAA

---

1. Hypothetical created by the author for illustrative purposes.
Statute) in 1999. This statute creates both a statewide requirement that all students pass the tenth-grade WASL before graduation and a statewide accountability system for schools that fail to produce students who can pass the WASL. Washington’s public high school class of 2008 must pass all sections of the tenth-grade WASL to receive a diploma. In the 1999–2000 school year, eighty percent of the state’s tenth graders failed at least one section. Given these failure rates, a large percentage of students in the class of 2008 is likely to meet all local requirements yet be unable to pass the WASL before the 2008 graduation date.

4. Id. § 28A.655.060(3)(i)(i).
5. Although the legislature did not specifically explain its reasoning, the graduation requirement does not apply to private school students. Id. § 28A.195.010(6). “The legislature hereby recognizes that private schools should be subject only to those minimum state controls necessary to insure the health and safety of all the students in the state and to insure a sufficient basic education to meet usual graduation requirements.” Id. § 28A.195.010.
6. The Superintendent of Public Instruction must craft the following additional sections before the graduation requirement takes effect in 2008: Science, Social Studies, Art, and Health. Id. § 28A.655.060(3)(b)(iii).
10. Students will probably have several chances to pass the Washington Assessment of Student Learning test (WASL) between their tenth-grade year and their expected graduation date. Telephone Interview with Rosemary Fitton, Coordinator for Special Projects, Office of the Superintendent of Public Instruction (Apr. 3, 2001). Remediation services will probably be provided to students who were unable to pass the WASL on their first attempt. Id. However, some educators have expressed concern that many students in the class of 2008 who scored poorly on their fourth grade WASL will be unable to pass the tenth-grade WASL by graduation. Hunter T. George, Delay New Requirements for Diplomas, Lawmakers Told, SEATTLE TIMES, Feb. 20, 2001, at B4. Furthermore, sixteen is both the legal dropout age in Washington and the age at which students will first learn of their tenth-grade test results. See WASH. REV. CODE §§ 28A.225.010(1)(e), 28A.655.060(3)(c) (2000). Students who fail the test may drop out, believing they will not receive a diploma.
Educational Malpractice

Should students who meet all local graduation requirements, but fail to pass the WASL, recover compensatory damages from their school district to pay for remedial education?11 Student suits for educational malpractice may arise when a school district fails to educate students to WASL standards by inadequately instructing them or by allowing them to pass local assessments even though they lack essential skills.12 In order to establish a successful claim for educational malpractice based on a state statute that contains no express cause of action, a plaintiff must show that a cause of action was implied by the law.13 In addition, courts have traditionally insisted that plaintiffs show a workable standard of care against which a court could measure school performance, a concrete and assessable harm, and a causal connection between the school district’s failure and the injury claimed.14 Finally, the plaintiff must overcome traditional policy barriers, including hesitation to interfere with the daily judgment of school officials15 and fear of opening floodgates of litigation against school districts.16

This Comment argues that students in the class of 2008 and beyond who are denied a diploma solely because they are unable to pass the WASL should be able to sue their school district and recover compensatory damages for educational malpractice under the AAA Statute. Part I describes the AAA Statute, which creates both the graduation requirement and a system through which the state evaluates school performance based on its students’ WASL scores. Part II outlines the test that Washington courts use to determine whether an implied individual cause of action exists under a state statute. Part III describes traditional law and policy shields that courts have used to protect school districts from educational malpractice causes of action. Part IV argues that the AAA Statute creates an implied private cause of action for individual students and that those students should also be able to

11. Any aggrieved party may maintain an action against a school district for an injury arising out of a district’s act or omission. WASH. REV. CODE §§ 4.08.110–120 (2000).
12. A student may choose to sue the State of Washington under a variety of other theories including a challenge to the validity of the WASL test. Such claims are outside the scope of this Comment.
16. Peter W., 131 Cal. Rptr. at 861.
overcome traditional legal and policy barriers to educational malpractice claims.

I. THE AAA STATUTE CREATES A SYSTEM THAT MEASURES STUDENT PERFORMANCE AND HOLDS SCHOOLS ACCOUNTABLE FOR STUDENT LEARNING

The Washington State Constitution mandates that the state fund basic education for all children in Washington, but the Washington Supreme Court has not interpreted the mandate to require that public education rise to a particular level of adequacy. However, since the early 1990s the Washington State Legislature has been developing a system of school accountability that relies on statewide standardized tests. The AAA Statute measures school performance by the WASL scores and holds school districts accountable for student failure as measured by those scores. Furthermore, the AAA Statute requires that all students pass the tenth-grade WASL as a prerequisite for graduation.

A. The Washington Constitution Supports the AAA Framework by Creating a Student Right to a Public Education

Article IX of the Washington Constitution establishes a student right to a public education. It declares that the state has a “paramount duty” to make “ample provision for the education of all children” by creating a “general and uniform system of public schools.” The Washington Supreme Court has interpreted this to mean that the state must sufficiently fund the public school system. Furthermore, the “paramount duty” language of Article IX imposes a judicially-enforceable affirmative state duty to provide a public education. That duty creates a “correlative right” for all Washington children to receive a

19. Id. § 28A.655.060(3)(c).
20. Because the U.S. Supreme Court has held that education is not a fundamental right, a student’s right to public education must be based on a state constitution. San Antonio Indep. Sch. Dist. v. Rodriguez, 411 U.S. 1, 35 (1973).
21. WASH. CONST. art. IX, §§ 1–2.
22. Seattle Sch. Dist., 90 Wash. 2d at 518, 585 P.2d at 95.
23. Id. at 482, 585 P.2d at 76.
public education. Recently, the Washington Supreme Court reaffirmed that all children have a “constitutionally paramount” right arising out of the state’s duty to make ample provision for the education of all children.

B. The Washington State Legislature Has Progressively Developed Both a Statewide Graduation Requirement and a System of School Accountability Based on Student Test Performance

Since 1992, the legislature has progressively amended education statutes, making schools increasingly accountable for student performance. In 1992 and 1993, the legislature made significant amendments to Washington’s education statutes. The amendments laid the framework for the current AAA Statute by emphasizing specific school district duties and corresponding school accountability.

The 1992 amendment created only a basic plan for developing an accountability system. The legislature found a need to increase student accomplishment by holding “schools accountable for their performance based on what students learn.” The statute created the Washington Commission on Student Learning (WCSL), charged with both identifying what elementary and secondary school students need to know and developing an assessment system to measure those skills. The assessment system would be used as a tool for measuring student achievement, evaluating instructional practices, and initiating academic

24. Id.
25. Tunstall v. Bergeson, 141 Wash. 2d 201, 221, 5 P.3d 691, 702 (2000). One commentator has suggested that an educational malpractice plaintiff might be able to state a claim based on the Washington Constitution. See Kelly Thompson Cochran, Comment, Beyond School Financing: Defining the Constitutional Right to an Adequate Education, 78 N.C. L. REV. 399, 419 (2000). However, Washington courts have denied attempts to establish such a claim. Id. at 418–19.
29. Ch. 141, § 1, 1992 Wash. Laws 574.
32. The Superintendent of Public Instruction eventually developed the WASL to serve as the assessment tool. See WASH. REV. CODE § 28A.655.030(1)(b) (2000).
support for students who did not master essential skills. The amendment also declared that students must pass the high-school-level assessment to receive a certificate of mastery, which will become a statewide requirement for graduation in 2008. Thus, the 1992 law laid the foundation for a statewide graduation requirement and a school accountability system.

The 1993 amendment added more specific structure to the accountability system. It set deadlines for the development of both “essential academic learning requirements” and the statewide student assessment system. The legislature also specified that the certificate of mastery needed for graduation should be attainable by age sixteen. Finally, the amendment required the WCSL to develop assistance programs for schools whose students fail the statewide assessment and a state intervention system to hold consistently failing schools accountable. These amendments reflect the early development of Washington’s school accountability and student assessment systems.

37. Education Reform-Improvement of Student Achievement Act, ch. 336, § 202(3)(a), 1993 Wash. Laws 1296. The essential academic learning requirements are very specific benchmarks based on the following general learning goals:

[T]o provide opportunities for all students to develop the knowledge and skills essential to: (1) Read with comprehension, write with skill, and communicate effectively and responsibly in a variety of ways and settings; (2) Know and apply the core concepts and principles of mathematics; social, physical, and life sciences; civics and history; geography; arts; and health and fitness; (3) Think analytically, logically, and creatively, and to integrate experience and knowledge to form reasoned judgments and solve problems; and (4) Understand the importance of work and how performance, effort, and decisions directly affect future career and educational opportunities.

C. The Current AAA Statute Completes the Development of Washington’s Assessment and Accountability Systems

In 1999, the legislature passed the current AAA Statute, which retains the salient features of the 1992 and 1993 laws. The new law consolidates existing accountability statutes, amends past assessment and accountability systems, and adds additional detail. The current AAA Statute creates a stricter, more focused accountability structure based on WASL scores.

The statute reveals a specific legislative purpose to protect the educational opportunities of individual students. The purpose of Washington’s accountability system is to improve student learning so that “each individual student” has the opportunity to develop the skills necessary to become a responsible and successful citizen. Furthermore, the legislature based the accountability system on the “fundamental principle” that all public school students should have access to curriculum and instruction aligned with state standards.

The AAA statute delegates duties to both the state Superintendent of Public Instruction (SPI), and a new state commission. The statute requires that the SPI maintain and continue to develop the statewide assessment system, the WASL, which students now take in the fourth, seventh, and tenth grades. The tenth-grade WASL is the high school level assessment that students must pass to receive a diploma. The statute also creates an Academic Achievement and Accountability Commission (A+ Commission) whose primary purpose is to oversee the state accountability system. The A+ Commission must identify schools and school districts “in which significant numbers of students

41. WASH. REV. CODE §§ 28A.655.005–.902; Ch. 388, 1999 Wash. Laws 2142.
42. WASH. REV. CODE §§ 28A.655.060(3)(c), .060(3)(i) (retaining accountability system and graduation requirements).
44. WASH. REV. CODE § 28A.655.005.
45. Id. The state standards are the Essential Academic Learning Requirements. See supra note 37.
46. WASH. REV. CODE § 28A.655.070(3).
49. Id. § 28A.655.020(2).
persistently fail to meet state standards,\footnote{Id. § 28A.655.030(1)(d).} using WASL scores to evaluate school district performance.\footnote{Id. § 28A.655.035(1)(b).} The A+ Commission must also develop a system for state intervention in identified schools.\footnote{Id. § 28A.655.060(3)(i)(iii). The A+ Commission has recommended interventions including withholding of funds, reorganization of district personnel, removal of schools from district jurisdiction, abolition or restructuring of school districts, and authorization of student transfers. Washington State Academic Achievement and Accountability Commission, Accountability System Recommendations 5 (Nov. 2000). The Washington Legislature had not passed these suggested interventions into law as of July 31, 2001.} The ultimate goal of education reform in Washington is a statewide school accountability system that uses WASL scores to evaluate school performance.

Finally, as required by the legislature,\footnote{WASH. REV. CODE § 28A.655.070(3).} the SPI has created regulations for AAA implementation.\footnote{See, e.g., WASH. ADMIN. CODE § 180-51-063 (2000).} The 2007–08 school year is the first year in which high school seniors must have a state certificate of mastery to graduate.\footnote{Id. § 180-51-063.} Therefore, children who are in the sixth grade during the 2001–02 school year will not graduate in 2008, even if they meet all local graduation requirements, unless they pass all sections of the tenth-grade WASL.\footnote{Id. § 180-51-063; see also supra note 10.} However, the AAA Statute contains neither a remedial provision, nor an express cause of action, for students who are denied diplomas based on failing WASL scores.

\section{Washington's Test for Implied Statutory Private Cause of Action Is More Favorable to Plaintiffs Than the Federal Test}

A student seeking redress for a denied diploma based on a failing WASL score should be able to establish an implied cause of action under the AAA Statute. To determine the existence of an implied private cause of action under a statute, the Washington Supreme Court uses a test first articulated in \textit{Bennett v. Hardy}.\footnote{113 Wash. 2d 912, 920–21, 784 P.2d 1258, 1261–62 (1990).} The court initially modeled its analysis after the U.S. Supreme Court's \textit{Cort v. Ash} test.\footnote{422 U.S. 66, 78 (1975).} Although federal courts have since become reluctant to allow implied private causes of action
under federal statutes, the Washington Supreme Court evaluates state causes of action more favorably to plaintiffs. This divergence is due to the Washington court’s refusal to isolate legislative intent as the sole consideration in determining whether a cause of action exists under a state statute. In fact, the court presumes that the legislature is aware of the doctrine of implied statutory causes of action, thereby finding a cause of action when the legislature is silent on the subject.

A. The U.S. Supreme Court Has Narrowed the Cort Test for Determining the Existence of Implied Federal Statutory Causes of Action

In 1975, the U.S. Supreme Court applied a four-factor test to determine the existence of an implied private cause of action under a federal statute. The Court asked whether Congress created the statute for the plaintiff’s special benefit, whether Congress explicitly or implicitly indicated an intent to create or deny a private remedy, and whether allowing the plaintiff a private remedy was consistent with the statute’s underlying purpose. Finally, the Court determined whether the action was one traditionally relegated to state law, thus making a federal implied cause of action inappropriate.

In subsequent cases, the Court narrowed the test by reducing it to a primary consideration of legislative intent. For example, in Touche Ross & Co. v. Redington, the Court treated the question of legislative intent as controlling. Because Congress was silent as to the existence of an implied private cause of action, the Court refused to find one. Two years later, in Middlesex County Sewerage Authority v. National Sea

59. Bennett, 113 Wash. 2d at 920–21, 784 P.2d at 1261–62.
61. Cort, 422 U.S. at 78.
62. Id.
63. Id.
66. Id. at 568.
67. Id. at 571.
Clammers Ass’n, the Court stated that it must consider statutory language and legislative history to determine legislative intent. Because the statute in Middlesex provided other remedies and the court found no intent to provide a private damages claim, the Court held there was no implied statutory private cause of action for damages. Based on this strict test, federal courts rarely find an implied private right of action under a federal statute.

B. The Washington Supreme Court’s Bennett Test More Readily Permits Plaintiffs To Establish Implied Causes of Action Under Washington Statutes

In Bennett v. Hardy, the Washington Supreme Court articulated its test for determining whether an implied cause of action exists under a state statute. The court retained the first three factors of the Cort test, but, unlike the U.S. Supreme Court, it did not require any affirmative legislative intent to create a private cause of action. Thus, it considered (1) whether the plaintiff was one for whose “special benefit” the statute was created; (2) whether legislative intent explicitly or implicitly supported creating or denying a remedy; and (3) whether implying a remedy was consistent with the underlying purpose of the legislation. Most significantly, legislative silence as to the presence or absence of a private cause of action did not preclude such an action. The court reasoned that when a statute has created a right by requiring certain conduct, a court should devise a remedy for injured persons though the statute specifies none. Therefore, a plaintiff can more easily establish an individual private cause of action under a Washington statute than under a federal statute.

---

69. Id. at 13.
70. Id. at 13–14.
71. 113 Wash. 2d 912, 784 P.2d 1258 (1990).
72. Id. at 920–21, 784 P.2d at 1261–62.
73. The fourth factor of the Cort test, determining whether action is traditionally relegated to state law, is not relevant to state claims. Id. at 921 n.3, 784 P.2d at 1261 n.3.
74. Id. at 920–21, 784 P.2d at 1261–62.
75. Id. at 919–20, 784 P.2d at 1261.
76. Id. at 920, 784 P.2d at 1261 (quoting RESTATEMENT (SECOND) OF TORTS § 874A (1979)).
In *Tyner v. State Department of Social & Health Services*, the Washington Supreme Court applied the *Bennett* test to Washington’s Child Abuse Statute. If the Department of Social and Health Services (DSHS) receives a report of suspected child abuse, it must investigate, issue a report, and, when necessary, refer such a report to the court. The *Tyner* court held that the Child Abuse Statute provided an implied private cause of action for a father who suffered lack of contact with his children resulting from negligent investigation by DSHS officials.

First, the court reasoned that the purpose section of the statute established accused parents as part of the class for whose special benefit the statute was created. The statutory language emphasized the importance of the bond between parent and child, stating that any intervention into a child’s life is also an intervention into the parent’s. Therefore, the plaintiff met the first prong of the *Bennett* test. Second, the court found that the legislature supported implying a remedy. Notwithstanding that the statute and legislative history were “silent as to this point,” the court presumed the legislature was “aware of the doctrine of implied statutory causes of action.” The court concluded that by recognizing the deep importance of the parent-child relationship, the legislature intended a judicial remedy for parents if the DSHS invaded that interest. Finally, the court held that an implied cause of action for damages for parents was consistent with the underlying purpose of the Child Abuse Statute. The Child Abuse Statute has two purposes: to protect children and to preserve the integrity of the family. Although the state argued that a cause of action for accused parents

---

77. 141 Wash. 2d 68, 1 P.3d 1148 (2000).
78. WASH. REV. CODE § 26.44.050 (2000).
79. *Tyner*, 141 Wash. 2d at 82, 1 P.3d at 1155.
80. *Id.* at 78, 1 P.3d at 1154.
81. *Id.* (quoting WASH. REV. CODE § 26.44.010 (2000)).
82. *Id.* at 80, 1 P.3d at 1154.
83. *Id.* at 80, 1 P.3d at 1154–55.
84. *Id.* at 80, 1 P.3d at 1155.
86. *Id.*
87. *Id.*
88. *Id.*
would create conflicting caseworker responsibilities, the court found that the statute creates a hierarchy of interests and that hierarchy allows for accomplishment of both purposes. Therefore, finding all three prongs of the Bennett test satisfied, the court held that under the Child Abuse Statute the DSHS owes a duty of care to parent suspects when investigating allegations of child abuse.

Justice Talmadge and two other Justices dissented, complaining that the majority failed to discuss the procedural protections for accused parents defined in the Child Abuse Statute. The dissent argued that the legislature did not create a private cause of action because procedural protections—including written notification of allegations and investigative findings, opportunity for written response, agency review, and adjudicative hearings—all safeguarded parents while recognizing that the primary goal of the statute was to ensure protection of abused children. Accordingly, no general duty should exist towards accused parents, but only a specific duty to follow the statute’s procedural provisions. Furthermore, the dissent argued that a general duty would create a conflict for investigators because sensitivity to parental interests would deter caseworkers from aggressively investigating child abuse. Therefore, the dissenting justices concluded that the legislature did not intend to create an implied statutory private cause of action for parents.

Tyner is the Washington Supreme Court’s most recent determination of the existence of an implied statutory private cause of action. The court continues to consider Bennett’s three factors and presumes the legislature

89. *Id.* at 81, 1 P.3d at 1155. The state argued that the duty to protect children from abusers and the duty to protect parents from being falsely accused created conflicting responsibilities. *Id.*
90. *Id.* at 79, 1 P.3d at 1154.
91. *Id.* at 82, 1 P.3d at 1155.
92. *Id.* at 93, 1 P.3d at 1161 (Talmadge, J., dissenting). Justices Guy and Smith joined the dissent. *Id.* at 89, 1 P.3d at 1159. Although the majority in *Tyner* did not discuss procedural protections, the Washington Supreme Court noted in an earlier case that a court should not find an implied cause of action where the statute has provided an adequate remedy. *Cazzanigi v. Gen. Elec. Credit Corp.*, 132 Wash. 2d 433, 445, 938 P.2d 819, 825 (1997).
94. *Tyner*, 141 Wash. 2d at 95–96, 1 P.3d at 1163 (Talmadge, J., dissenting).
95. *Id.* at 96, 1 P.3d at 1163 (Talmadge, J., dissenting).
96. *See id.* at 98, 1 P.3d at 1164 (Talmadge, J., dissenting).
97. *Id.* at 98–99, 1 P.3d at 1164 (Talmadge, J., dissenting).
98. *Id.* at 96, 1 P.3d at 1163 (Talmadge, J., dissenting).
Educational Malpractice

is “aware of the doctrine of implied statutory causes of action.”99 The Tyner dissenters believed Washington courts should continue to consider the presence or absence of statutorily-defined procedural protections for the plaintiff in deciding whether to find a cause of action for damages.100 Therefore, a court could consider all of these factors when determining the existence of an implied private cause of action for damages under a Washington statute.

III. WASHINGTON AND OTHER STATE COURTS HAVE DENIED EDUCATIONAL MALPRACTICE CLAIMS FOR BOTH LEGAL AND POLICY REASONS

State courts nationwide have almost uniformly denied educational malpractice claims.101 Educational malpractice plaintiffs wishing to sue under a state statute must first show that the legislature implied an individual cause of action.102 Then, the plaintiff must show a workable standard of care, a failure to meet that standard of care, a comprehensible and assessable harm, and a causal connection between the school district’s failure and the resulting harm.103 Finally, a plaintiff must overcome traditional policy concerns including hesitation to substitute the court’s judgment for that of school officials, reluctance to interfere with the day-to-day workings of schools, and fear that a flood of litigation could result.

100. Id. at 93, 1 P.3d at 1161 (Talmadge, J., dissenting).
101. Only Montana has allowed a claim for educational malpractice to survive summary judgment. See B.M. v. Montana, 649 P.2d 425, 427 (Mont. 1982) (holding that state owes duty of care to child in testing and special education placement).
103. These are the basic elements of a prima facie case of educational malpractice. However, courts uniformly treat them as threshold policy barriers to establishing an implied cause of action. See, e.g., Peter W. v. S.F. Unified Sch. Dist., 131 Cal. Rptr. 854, 859 (Cal. Ct. App. 1976); Donohue v. Copiague Union Free Sch. Dist., 391 N.E. 2d 1352, 1353–54 (N.Y. 1979).
A. Washington Courts Have Denied Educational Malpractice Claims Because Individual Student Plaintiffs Were Unable to Establish a Private Cause of Action

Student plaintiffs have twice attempted to establish an educational malpractice cause of action under Washington statutes. Division One of the Washington Court of Appeals denied both claims. In each case, the court determined that the relevant statutes created no cause of action or duty to individual students; thus, the court was wholly unwilling to allow these educational malpractice claims to advance.

1. The Court of Appeals Denied a Private Cause of Action for Educational Malpractice Because Plaintiffs Did Not Show They Had an Individual Right to an Adequate Education

In Camer v. Brouillet, an unpublished case, plaintiffs sued on the theory that the Superintendent of Public Instruction (SPI) had failed to implement the Student Learning Objectives Law correctly. In applying the legislative intent prong of the Cort test, the court recognized an additional requirement for implying a private cause of action: the plaintiffs must show legislative intent to create a school district duty owed to individual students. This requirement arose out of a “traditional rule” that a statutorily imposed duty on public officials is owed only to the public as a whole except when the legislature shows an intent to protect certain people. Those people may bring an action in tort for a violation of the statute.

104. Camer, 52 Wash. App. at 532, 762 P.2d at 358; Brouillet, No. 10227-3-I, at 4.
105. Camer, 52 Wash. App. at 538, 762 P.2d at 360; Brouillet, No. 10227-3-I at 5.
106. Camer, 52 Wash. App. at 537, 762 P.2d at 360; Brouillet, No. 10227-3-I at 5.
108. Id. at 2; cf. WASH. REV. CODE § 28A.58.092 (1981).
111. Bjomrud, supra note 110 at 1074–75.
112. Brouillet, No. 10227-3-I at 4.
Educational Malpractice

To determine whether the statute in question created an individual cause of action, the court examined legislative history. 113 The purpose of the Student Learning Objectives Law was to promote efficient management of quality education and guide student attainment of objectives without consequences for individual students. 114 Accordingly, the court reasoned that the language and legislative history were devoid of any legislative intent to create a duty to individual students. 115 The court rejected the plaintiff’s attempt to establish an implied private cause of action because the statutory duty was owed only to the general public. 116

2. Plaintiffs Have Been Unable To Establish a Cause of Action for Educational Malpractice Based on Other Washington Education Statutes

In *Camer v. Seattle School District*, 117 the court rejected an educational malpractice claim based on an implied cause of action under Washington statutes 118 because the plaintiff could not meet the elements of the Cort test and alternative administrative remedies were available. 119 At that time, one statute required students to study the Washington Constitution as a prerequisite for high school graduation. 120 Another mandated a one-semester course in state history and government. 121 The Seattle School District had provided neither course of study to the individual plaintiffs. 122 In this pre-*Bennett* case, the court determined that no private cause of action existed under these statutes. 123

---

113. *Id.* at 4–5.
114. *Id.* at 4.
115. *Id.* Because the *Brouillet* case was pre-*Bennett*, the court applied the *Cort* test without the presumption that the legislature was aware of the doctrine of implied rights of action. *Id.*
116. *Id.* at 5; *accord* Peter W. v. S.F. Unified Sch. Dist., 131 Cal. Rptr. 854, 862 (Cal. Ct. App. 1976) (holding that statutory enactments were directed to attainment of optimal education results but not safeguards against particular injury); Donohue v. Copiague Union Free Sch. Dist., 391 N.E.2d 1352, 1353 (N.Y. 1979) (holding that state constitution provided general directive but was never intended to impose duty flowing directly from local school district to individual pupils).
121. *Id.* § 28A.05.050.
123. *Id.*
The court used the federal Cort test to determine whether a private cause of action existed that would allow students to sue when school districts failed to provide statutorily-mandated courses. The test was identical to that applied later in Bennett, but it did not include the presumption that the legislature was aware of the doctrine of implied causes of action. The court assumed, arguendo, that the legislature enacted the statute for the special benefit of the plaintiffs. However, the court then concluded that the statutory language was devoid of any express or implied legislative intent to create a private cause of action for damages and that the student plaintiffs did not establish legislative intent to the contrary through legislative history. Finally, the court found that an implied cause of action would not be consistent with the purposes of the legislative scheme. The court reasoned that the legislative purpose was to establish general guidelines for producing the constitutionally-mandated “ample” education for Washington children, to be administered within the discretion of local school districts.

In the course of its application of the Cort test, the court noted that the administrative process provided a proper chain of accountability and that administrative sanctions adequately eliminated a need for an individual right of action. In addition, the legislature had limited judicial review to persons who had been specifically aggrieved by a decision or order of a school official or board. Therefore, existing administrative remedies and legislative limitation of judicial review implied that an individual cause of action for damages was inappropriate.

124. Id. at 536–37, 762 P.2d at 360 (quoting Cort v. Ash, 422 U.S. 66, 78 (1975)).
125. Id.
126. Id. at 537, 762 P.2d at 360.
127. Id.
128. Id.
129. Id. See also supra Part I.A.
131. See id. (citing WASH. ADMIN. CODE 180-16-195(3) (1986)).
B. Early Cases in California and New York Provide Examples of Other Legal and Policy Reasons Why Courts Have Denied Educational Malpractice Claims

Two seminal educational malpractice cases illustrate other state courts’ reasons for barring implied private causes of action for educational malpractice. In *Peter W. v. San Francisco Unified School District*, a California high school graduate sued his school district because he had been allowed to graduate from high school though he could only read at a fifth-grade level. However, the court found that the plaintiff failed to state a cause of action based on either common law negligence or the California Code. Likewise, in a New York case, *Donohue v. Copiague Union Free School District*, a graduate who was unable to read sufficiently to complete an employment application sued his school district under both the state constitution and common law negligence. Although the New York court was careful not to eliminate the possibility that a properly argued educational malpractice claim could survive, this particular claim was dismissed based on policy considerations.

These seminal educational malpractice cases reveal a consistent set of legal and policy concerns that an educational malpractice plaintiff must overcome to establish a cause of action. Courts have been unable to create a definite standard of care for educators. They have recognized the difficulty in proving a comprehensible and assessable educational harm

134. Id. at 856.
135. Id. at 861.
136. Id. at 862. The court rejected claims based on the California Code because, it reasoned, a statute could impose liability only when it is designed to protect against a “particular kind of injury,” and the court did not consider failure of educational achievement to be a particular injury under tort law. Id. Also, the court determined that the statutes did not create a duty to individual students because they were generally administrative and not protective of individuals. Id.
138. Id. at 1353 (“The legislature shall provide for the maintenance and support of a system of free common schools, wherein all the children of this state may be educated.” (quoting N.Y. CONST. art. XI, § 1)). The court dismissed the claim because this clause was intended as a “general directive” but not intended to impose a duty “flowing directly from a local school district to individual pupils.” Id.
139. Id.
140. Id. at 1354.
and have acknowledged that schools often are not solely to blame for a student’s failure to learn. Furthermore, policy concerns have included the judiciary’s hesitation to become involved in day-to-day school administration, its reluctance to interfere with the professional judgment of administrators and educators, and its unwillingness to open a floodgate to costly litigation. Therefore, even if a Washington plaintiff could establish an implied private cause of action for educational malpractice under the AAA Statute pursuant to the Bennett test, that student must also overcome these legal and policy barriers.

1. To Establish Educational Malpractice Claims, Plaintiffs Must Show a Workable Standard of Care, a Concrete and Assessable Harm, and a Causal Connection Between School Failure and the Student Harm

Courts have cited several legal reasons, although couched in public policy terms, for denying educational malpractice causes of action. First, courts have rejected educational malpractice claims based on the lack of a workable standard of care. In Peter W., the court noted that the challenged classroom methodology afforded no readily acceptable standard to assess against the duty owed to students. The court held that there existed “no conceivable ‘workability of a rule of care’” against which the school’s alleged misconduct could be measured, because a variety of conflicting theories existed that suggest how or what a child should be taught. However, the Donohue court noted that creation of a standard by which to measure school performance would not pose an “insurmountable obstacle.”

Courts also have found it very difficult to establish a comprehensible or assessable educational harm. In Peter W., the court observed that on occasions when courts have sanctioned new areas of tort liability, the wrongs and injuries involved were both comprehensible and assessable. Peter W.’s claim that he had failed to achieve academically

142. Donohue, 391 N.E.2d at 1354.
145. Id. at 860–61.
146. Id. (quoting Raymond v. Paradise Unified Sch. Dist., 31 Cal. Rptr. 847, 851 (Cal. Ct. App. 1963)).
147. Donohue, 391 N.E.2d at 1353.
148. Peter W., 131 Cal. Rptr. at 860.
Educational Malpractice

simply did not offer the concrete and assessable injury required. Thus, the Peter W. court could not be certain that the plaintiff had suffered an injury within the law of negligence. Furthermore, though it did not elaborate, the Camer court also believed that the children did not suffer actual damage or injury when they could not take classes in Washington government.

Finally, courts have denied educational malpractice claims because they have been concerned that inadequate instruction and student promotion despite lack of skills are not the only causes for student failure. The Peter W. court noted that “physical, neurological, emotional, cultural, [or] environmental” factors could influence a child’s learning, such that no perceptible causal connection existed between the defendant’s conduct and the injury suffered. Still, the Donohue court admitted that, although the causation element might be difficult to prove, “it perhaps assumes too much to conclude that it could never be established.”

2. Public Policy Concerns Have Led Courts To Deny Educational Malpractice Causes of Action

Courts also have rejected educational malpractice claims based on public policy concerns. Courts have been reluctant to allow a private cause of action if the court must impose its own judgment on school authorities or interfere with the everyday administration of a school by determining which instructional practices or curricula are effective. The Peter W. court emphasized that “pedagogy itself is fraught with different and conflicting theories of how or what a child should be taught.” Similarly, the Donohue court recognized that to entertain a cause of action for educational malpractice, courts would not merely make broad educational policy judgments, but more importantly, would also review their day-to-day implementation. This involvement would cause blatant interference with the administration of the public school

149. Id.
150. Id. at 861.
152. Peter W., 131 Cal Rptr. at 861.
154. Peter W., 131 Cal. Rptr. at 860–61.
155. Donohue, 391 N.E.2d at 1354.
Due to such concerns, the *Camer* court concluded that the internal administrative review was the proper forum for the plaintiffs’ complaint. Nevertheless, the *Donohue* court conceded that extreme situations may exist where a court must intervene.

Finally, courts have feared the possible flood of litigation that might result from finding an implied private cause of action for damages based on educational statutes. For example, the *Peter W.* court worried that holding schools to an actionable duty of care when a student failed to achieve would expose them to real or imagined tort claims of disaffected students. Parents and students “in countless numbers” might burden school and society “beyond calculation.”

Legal and public policy issues have played a crucial role in denying implied private rights of action for educational malpractice claims. Washington courts have, and will likely continue to consider the barriers set forth in these seminal educational malpractice cases. As a result, even if a Washington plaintiff could establish an individual statutory private cause of action for educational malpractice under the AAA Statute, that student would also have to overcome these legal and policy barriers.

IV. WASHINGTON COURTS SHOULD ALLOW EDUCATIONAL MALPRACTICE CLAIMS FOR STUDENTS WHO MEET LOCAL GRADUATION REQUIREMENTS BUT FAIL THE WASL

Unfortunately, a large number of students in the class of 2008 are likely to fail the tenth-grade Washington Assessment of Student Learning (WASL). Some of these students will find that passing the WASL is the only graduation requirement they cannot satisfy. These students should be able to establish an implied private cause of action against their school district for compensatory damages to pay for remedial education. By applying the *Bennett* test to the Academic

---

156. *Id.*
157. *Camer v. Seattle Sch. Dist.*, 52 Wash. App. 531, 537, 762 P.2d 356, 360 (1988) (“Courts and judges are normally not in a position to substitute their judgment for that of school authorities, nor are we equipped to oversee and monitor day-to-day operations of a school system.” (internal citations omitted)).
158. *Donohue*, 391 N.E. 2d at 1354.
159. *Peter W.*, 131 Cal. Rptr. at 861.
160. *Id.*
161. See supra note 8 and accompanying text.
Achievement and Accountability Statute (AAA Statute), these students should be able to assert a private cause of action and also overcome the traditional legal and public policy barriers to educational malpractice claims.

A. Individual Students Who Fail the WASL Should Be Able To Establish an Educational Malpractice Cause of Action Under Washington Law

Applying the Bennett v. Hardy test to the AAA Statute should imply a cause of action because the legislature created the statute especially for student benefit, legislative history does not preclude an implied cause of action for students, and implying a private cause of action is consistent with the underlying purpose of the AAA Statute: school accountability. Furthermore, the AAA statute provides no administrative remedy for students who do not receive a diploma because of their inability to pass the WASL and neither students nor parents have any procedural rights under the statute that otherwise would protect them. Consequently, students who satisfy all local graduation requirements but cannot pass the WASL should have an implied private cause of action against their school districts under the AAA Statute for damages to pay for remedial education.

I. Student Plaintiffs Should Be Able To Establish an Individual Cause of Action for Educational Malpractice by Applying the Bennett Test to the AAA Statute

Students should be able to apply the Bennett test to the AAA Statute to establish a cause of action. First, the purpose statement of the AAA Statute reveals that the legislature created the statute for the special benefit of Washington students. The purpose of Washington’s accountability system is to improve student learning so that each student has the opportunity to develop the skills necessary to become a

162. 113 Wash. 2d 912, 784 P.2d 1258 (1990).
164. See supra Parts I.B., I.C.
165. See WASH. REV. CODE § 28A.655.005.
166. See id. § 28A.655.060(3)(c).
167. See infra Part IV.A.1.
responsible and successful citizen. The legislature also declared that the accountability system should be based on “a fundamental principle” that all public school students have access to curriculum and instruction that is aligned to state standards. This intent to benefit individual students through the development of specific skills was notably absent in the statute requiring Washington government coursework, on which the plaintiffs relied in Camer v. Seattle School District No. 1. The legislature intended the AAA Statute to especially benefit individual students, so the statute should satisfy the first prong of the Bennett test.

Second, the legislative intent behind the AAA Statute does not expressly preclude an individual implied cause of action for students. As established in Tyner v. Department of Social and Health Services, when a statute and its legislative history are silent on this point, Washington courts presume that the legislature is aware of the doctrine of implied causes of action. Like the Child Abuse Statute in Tyner, there is no mention in the AAA Statute’s legislative history of either allowing or precluding a private cause of action for damages for educational malpractice. Unlike the pre-Bennett cases, Brouillet and Camer, when a current court interprets the AAA Statute, the legislature’s silence raises a presumption that it did not intend to preclude private causes of action for students. Furthermore, Washington courts should find that the AAA Statute creates a school district duty to educate individual students adequately because the avowed purpose of the AAA Statute is to create a school accountability system, ensuring that both individual students and schools meet state standards. A student

168 WASH. REV. CODE § 28A.655.005.
169 Id.
172 141 Wash. 2d 68, 1 P.3d 1148 (2000).
174 Id.
right to an adequate education does not appear to arise from the Washington Constitution. However, a student right to an education that meets state standards logically follows from the AAA statutory duty just as the “paramount duty” to provide for a system of public schools created a “correlative” student right to a public education. Thus, a court could infer, as it did in Tyner, that the legislature recognized that an individual cause of action could protect this right.

Third, the underlying purpose of the AAA Statute supports the establishment of a student’s implied cause of action for educational malpractice. The legislature enacted the AAA Statute as an attempt at education reform through school accountability. The entire accountability system emphasizes the importance of school district responsibility for the quality of education that it provides for each student. This emphasis on school accountability to benefit individual students through the development of specific skills did not underlie the statutes mandating Washington government coursework on which the Camer plaintiffs depended. Nor did this emphasis inspire the Student Learning Objectives Law on which the Brouillet plaintiffs relied. In contrast, the AAA Statute’s underlying emphasis on school accountability for individual student skill level logically supports school accountability to individual students through implied individual causes of action, thereby satisfying the third prong of the Bennett test. Accordingly, individual student plaintiffs who are unable to pass the WASL but would otherwise receive a diploma based on local graduation requirements should meet all three prongs of the Bennett test.

2. Under the AAA Statute, School Districts Have a Duty To Educate Individual Students to WASL Standards

The school district’s duty to provide an education adequate to meet state standards does not merely extend to the general public like the

---

178. WASH. CONST. art. IX, § 1.
179. Seattle Sch. Dist. v. State, 90 Wash. 2d 476, 482, 585 P.2d 71, 76 (1978). Even if a school district were to argue that its duty is to the Superintendent of Public Instruction and the A+ Commission, Tyner revealed that Washington courts are willing to recognize duties to more than one entity within a single statute. Tyner, 141 Wash. 2d at 79, 1 P.3d at 1154; see also infra Part IV.A.2.
180. See supra Part I.C.
182. See id. §§ 28A.02.080, 28A.05.050 (1987).
Student Learning Objectives Law in Brouillet.184 It extends also to individual students. The essential academic learning requirements, referenced throughout the AAA Statute, are based on student learning goals that focus on specific, individual abilities that each school district must achieve for “all students.”185 In addition, the statute was created to ensure that “each individual student will be given the opportunity to become a responsible citizen and successfully live, learn, and work in the twenty-first century.”186 Furthermore, the legislature declared a fundamental principle that “all public school students” have access to educational opportunities that are aligned with the standards.187 Finally, the legislature created a very individual consequence when it decided to make the WASL a high-stakes test by denying a diploma to a student who fails,188 indicating the legislature intended that each student should be educated to a level such that he or she can pass the WASL.

Early Washington educational malpractice cases should not deter a court from finding that the AAA Statute creates an individual implied cause of action. The plaintiffs in Camer and Brouillet did not suffer an individual consequence like the loss of a diploma.189 Furthermore, the constitutional duty to provide ample education discussed in Camer is far more general than the particular statutory duty to educate individual students adequately for them to pass the WASL.190 In addition, this case can be distinguished from Peter W. v. San Francisco Unified School District,191 in which the court called California’s education statutes generally “administrative” but not “protective” of individuals because they did not safeguard against an individual injury of any kind.192 Unlike the plaintiff in Peter W., a Washington child would suffer an individual, concrete, and easily proven harm if denied a diploma. Furthermore, the language of the AAA statute is more focused on results for individual

186. Id. § 28A.655.005 (emphasis added).
187. Id. (emphasis added).
188. Id. § 28A.655.060(3)(c).
190. Compare supra notes 128–130 and accompanying text with supra notes 184–188 and accompanying text.
192. Id. at 862. This holding rested on the court’s characterization of harm as a student’s failure to achieve academically. Id.
Educational Malpractice

students than the clause of the New York constitution studied in
Donohue v. Copiague Union Free School District.193 The AAA Statute is
distinguishable from earlier bases for student claims because it focuses
on the individual student. As a result, Washington courts should find that
the school district’s duty to educate adequately extends to the individual
student.

A school district might contend that its only duty under the AAA
Statute is to provide an opportunity for students to achieve all of the
essential academic learning requirements.194 From this perspective, the
WASL merely measures whether the student has absorbed and can
reiterate the information taught. If a child fails the WASL, the school
might argue that the student failed to learn the required materials even
though the school fulfilled its duty by adequately providing an
opportunity to learn those materials.195 However, a student could reply
that her ability to achieve all local graduation requirements reflects that
she either mastered, to the satisfaction of the school district, the skills
actually taught, or that the school district allowed her to pass despite a
lack of skill. In either case, the school district should be responsible
under the AAA Statute for failing to develop the student’s individual
abilities adequately for her to pass the WASL.196

3. Because the AAA Statute Offers Neither Alternative Remedies Nor
Procedural Protections to Students, an Implied Cause of Action for
Educational Malpractice Should Exist

Neither administrative remedies nor statutorily defined procedural
protections would preclude an educational malpractice claim based on an
implied private cause of action under the AAA Statute. While the
majority did not address alternative remedies in Tyner, the Camer court
mentioned the existence of administrative remedies as a reason for
denying a cause of action in Washington courts.197 The A+ Commission’s recommended intervention strategies do not include

194. See supra note 37.
195. The Fifth Circuit has held that competency exams resulting in diploma sanctions must fairly
reflect what is taught. Debra P. v. Turlington, 644 F.2d 397, 403–06 (5th Cir. 1981). However, no
federal or state court has imposed this restriction in Washington.
196. See Arthur L. Coleman, Excellence and Equity in Education: High Standards for High-Stakes
administrative remedies for students who have completed all local graduation requirements but have been denied a diploma. The AAA Statute requires the A+ Commission to create a system to intervene in schools and school districts in which significant numbers of students persistently fail to meet WASL standards. Yet unlike Camer, in which a positive administrative ruling requiring Washington government courses might benefit one of the Camer children, any administrative action intended to improve a school district would not help a student who had already been denied a diploma. Even if a school were to improve, the remedy would simply be too late; it could not impact a student who no longer attends the school.

Additionally, the Tyner dissent relied heavily on the fact that procedural protections were defined in the Child Abuse Statute. Even though the dissent argued that adequate procedural rights provided to children or parents in a statute might preclude the finding of an implied cause of action, neither the AAA Statute nor the A+ Commission provides specific procedural protections for individual students like those for accused abusers in the Child Abuse Statute. Furthermore, in Tyner, accused parents like the plaintiff were not intended to be the primary beneficiaries of the Child Abuse Statute, but merely secondary

198. WASHINGTON STATE ACADEMIC ACHIEVEMENT AND ACCOUNTABILITY COMMISSION, supra note 52, at 5.
201. Even though the Office of the Superintendent of Public Instruction has acknowledged that it will provide remedial educational services to eleventh and twelfth-grade students who initially fail the tenth-grade WASL, those remedial services do not extend to students whose graduation date has passed. Telephone Interview with Rosemary Fitzon, Coordinator for Special Projects, Office of the Superintendent of Public Instruction (Apr. 3, 2001).
202. No one has suggested that a school improved by the AAA system would invite deprived students to return to receive an improved education. Telephone Interview with Rosemary Fitzon, Coordinator for Special Projects, Office of the Superintendent of Public Instruction (Apr. 3, 2001).
204. See id.
205. Although the “Accountability Policies” section of the AAA Statute mentions that the A+ Commission must consider procedural issues in its deliberations, this section specifically deals with state strategies for intervening in failing schools and school districts. See WASH. REV. CODE 28A.655.035(1)(c) (2000). Therefore, it is likely that the procedural considerations mentioned here would apply to institutional intervention into schools and not to individual students.
206. See supra note 93 and accompanying text.
beneficiaries within the hierarchy of interests.\textsuperscript{207} However, the legislature explicitly intended students to be the primary beneficiaries of the AAA Statute,\textsuperscript{208} making an even stronger case that the legislature intended to protect students by allowing tort remedies. Thus, because no statutorily-defined procedural rights exist for students, the primary beneficiaries of the AAA statute, a court should not allow an argument like the one proposed by the \textit{Tyner} dissent to prevent students from seeking compensatory damages for remedial education.\textsuperscript{209}

\textbf{B. A Student Who Has Satisfied All Local Graduation Requirements Should Overcome Legal and Policy Barriers to Establishing a Cause of Action for Educational Malpractice}

An individual student who has met all local graduation requirements but has failed to pass the WASL should overcome traditional legal and policy barriers to educational malpractice claims. The AAA Statute creates a standard of care against which courts can measure school performance. Additionally, a denied diploma is a concrete and assessable harm, and a student who has met all local graduation requirements can show a causal connection between the denied diploma and the school district’s failure to educate adequately. Furthermore, WASL scores could prevent the court from having to interfere with the judgment of administrators because they provide a reliable measure of school performance. Finally, a court could impose limitations that would alleviate concern that a flood of litigation might result.

\textbf{1. A Student Who Has Met All Local Requirements Could Establish a Standard of Care, a Concrete Harm, and a Causal Connection Between School Failure and the Denied Diploma}

A student who has completed all local requirements could show all of the elements necessary to establish an educational malpractice claim. The

\begin{itemize}
\item \textsuperscript{207} \textit{Tyner}, 141 Wash. 2d at 92–93, 1 P.3d at 1161 (Talmadge, J., dissenting).
\item \textsuperscript{208} See \textit{WASH. REV. CODE} § 28A.655.005.
\item \textsuperscript{209} One author has suggested protections to eliminate the risk of inappropriately denying post-graduation opportunities based on test scores. These protections include: (1) establishing compensatory supports to ensure adequate opportunities to master tested materials, (2) providing multiple opportunities to take the test, and (3) considering academic factors other than test scores that may affirm or challenge high stakes conclusions derived from test scores. Coleman, \textit{supra} note 196, at 109–10.
\end{itemize}
Donohue court admitted that a court might create a standard of care, but in Washington the court need not do so because the legislature has concluded that the WASL is a workable standard. The WASL sets a measurable standard of care by reliably assessing a student’s mastery of the essential academic learning requirements. The state itself relies on the WASL to determine the adequacy of the education provided by school districts, establishing the WASL’s validity as an assessment tool. A Washington student who has met all local graduation requirements could present failing WASL scores, indicating the school’s failure to adequately educate to a level mandated by statewide standards. Unlike the plaintiff in Peter W., a Washington student would not have to establish a “correct” pedagogy against which the court would measure the school district’s performance because the school district’s failure, as defined by the legislature, would be evident in the student’s failing WASL scores.

An assessable educational harm will exist when a student does not pass the tenth-grade WASL and therefore does not graduate. The harm of not receiving a high school diploma places a student at a serious disadvantage for future employment opportunities. Individuals with high school diplomas make as much as nineteen percent more per hour than those without. This harm has a more concrete and provable impact than either the failure to achieve academically, claimed in Peter W., or the failure to learn about Washington government, claimed in Camer. Accordingly, a student who is denied a diploma based solely on his or her WASL scores should be able to establish a concrete and assessable harm sufficient to overcome this legal barrier.

Courts have resisted recognizing an implied statutory cause of action for educational malpractice out of concerns that inadequate education is

---

211. WASH. REV. CODE § 28A.655.035(1)(b)(i).
212. Id. § 28A.655.030(1)(b).
213. Id. § 28A.655.035(1)(b)(i).
216. Id. Courts have recognized the impact of a failure to graduate by declaring a property interest in a high school diploma. See, e.g., Debra P. v. Turlington, 644 F.2d 397, 403 (5th Cir. 1981) (quoting Goss v. Lopez, 419 U.S. 565, 574 (1975)).
217. Peter W., 131 Cal. Rptr. at 862.
not the only cause for student failure. However, as the *Donohue* court noted, causation in an educational malpractice case is not impossible to prove. Because a student could satisfy all local graduation requirements, the court could be certain that one of two things has happened. In one instance, the student will have passed all local graduation requirements despite “physical, neurological, emotional, cultural, [and] environmental” factors, thus impugning the school’s inadequate instruction. Alternatively, the school district will have allowed the student to pass local requirements despite a lack of skill, without remedial or special education assistance geared toward the WASL. In either case, the school district will be responsible for the student’s failure to learn, making compensatory damages for remedial education appropriate.

A court could also alleviate the causation dilemma by allowing the “substantial factor” test to prevail. That is, if a court, like the *Peter W.* court, feels that both school and external factors have caused the student’s failure, the court could determine that “two causes concur to bring about an event and either one of them operating alone could have been sufficient to cause the result.” A school district could then be found fully liable for compensatory damages for remedial education because the school district was a “substantial factor” in a child’s educational injury, even if a school district was not the most significant or sole cause. Accordingly, the legal concern that a student would be unable to show causation should not preclude Washington courts from allowing students to establish an implied cause of action under the AAA Statute.

221. *Peter W.*, 131 Cal. Rptr. at 861.
224. *Id.*
225. Additionally, Washington follows a rule of comparative negligence, allowing juries to allocate responsibility for an injury between the plaintiff and the defendant. Wash. Rev. Code § 4.22.005 (2000). Therefore, a court or jury could attribute fault fairly between the student and the school district. See also 6 Hon. George T. Shields, Wash. Pattern Instructions, *in Washington Practice* 106–07 (1989) (“[R]educe the total damages you find to have been sustained by the plaintiff, by the percentage of [plaintiff’s] contributory negligence.”).
School districts may also hypothesize that two students could receive an identical education, even sitting next to each other in every class throughout the students’ educational careers. If one student can pass the WASL and the other cannot, how can the district be blamed for inadequate education? However, in using the WASL tests to evaluate school performance, the Washington State Legislature has acknowledged that school districts carry some responsibility for actual student learning, not merely teacher output of information. This responsibility includes preparing students with both the substantive and exam-taking skills to complete the tenth-grade WASL successfully. The most basic obligation of educators is to meet the needs of students as they find them, accounting for their varying backgrounds and abilities. Under the affirmative legislative mandates and duties of the AAA Statute, school districts cannot skirt their responsibility to educate each child simply by lamenting learning differences.

2. A Student Who Meets Local Graduation Requirements Should Also Be Able To Overcome Traditional Public Policy Concerns That Have Led Courts To Deny Educational Malpractice Claims

A student who has passed all local graduation requirements but failed the WASL and who seeks compensatory damages for remedial education should be able to overcome traditional public policy concerns. In allowing an implied private cause of action for educational malpractice under the AAA statute, a court would not have to substitute its judgment for that of school authorities or interfere with the everyday workings of school administrators. Because the A+ Commission and the legislature are both willing to make evaluative decisions based on the objective WASL assessment system, courts should not hesitate to rely on that system as a measure of both individual student mastery of the essential academic learning requirements and overall school performance. Unlike the Peter W. and Donohue courts, Washington courts would not have to prescribe specific pedagogy because the AAA Statute and the WASL provide a standard against which to measure school performance. Courts would simply be insisting that school districts meet those standards.

226. See supra notes 44–45 and accompanying text.
227. Coleman, supra note 196, at 84.
imposed by the legislature. Therefore, courts would not need to substitute their judgment for that of school authorities.

Furthermore, Washington courts would not have to interfere with everyday school administration because the plaintiff would already be outside the school system. Presumably, the AAA Statute’s accountability system should continue to reform the day-to-day workings of schools. Relief could be limited to compensatory damages for remedial education and would be appropriate only when the accountability system had not yet succeeded in a specific school district. A court could limit damages even further by insisting that the school district pay fees directly to the institution where the child would attend remediation courses. While an implied private cause of action would encourage school districts to reform a failing school quickly, a private remedy for students already outside the school system would in no way require the court to dictate exactly how a district should improve its failing school. A cause of action would simply provide a separate remedial scheme that would avoid the problems anticipated by the Donohue and Camer courts.

Finally, courts have been concerned with the flood of litigation that could result from an implied cause of action for educational malpractice. Presumably, a Washington court would be just as reluctant as the Peter W. court to allow recovery for a student’s lost employment opportunities resulting from an inadequate education. However, awarding only costs of compensatory education would offer a fair remedy that would not unreasonably burden school districts. Thus, by limiting damages to funding for compensatory education, courts would reduce the number of plaintiffs willing to sue. Disaffected students who solely sought retaliation against their districts would have no motivation to sue unless they actually wished to meet WASL standards. Furthermore, only

230. In other contexts, courts have relied on test scores to make conclusions that the court could not reach without outside evaluation. See Chapple v. Ganger, 851 F. Supp. 1481, 1497 (E.D. Wash. 1994) (using test scores to determine extent that student’s diminished performance was result of automobile accident). In addition, one commentator has suggested that educational malpractice claims should follow medical and legal malpractice standards: courts should accept evidence from experts in the profession to prove causation and harm. Lush, supra note 141, at 112.

231. See supra notes 46–52 and accompanying text.

232. Washington State’s community college system might be an appropriate setting for a remediation program. Although some might suggest that the public school system could provide post-twelfth grade remediation, it is important to remember that the court is providing the remedy because the school has failed in its duty to adequately educate.

233. See supra notes 154–158 and accompanying text.

students who could satisfy all other graduation requirements would be able make out an educational malpractice cause of action. Therefore, the causation requirement itself could limit the number of students who could successfully bring the implied cause of action.

Finally, though the graduation requirement does not take effect until 2008, if WASL test scores do not improve significantly before that deadline, there may be a large number of students who could claim that the school district caused their failure. This Comment demonstrates the type of liability that a school district could face if it fails to solve this problem before 2008. School districts might begin to alleviate the problem by insisting that teachers honestly assess a student’s abilities when determining passing or failing grades for a particular course. Furthermore, if the number of students who satisfy local requirements far exceeds the number who can pass the WASL, the Superintendent of Public Instruction may want to reconsider the standards or provide for multiple alternative methods of assessment.235

V. CONCLUSION

In Washington State, the legislature has shown dedication to improving the skill level of Washington students by introducing a new and innovative school accountability system. As part of this system, the legislature has chosen to deny high school diplomas to students who cannot pass a statewide assessment test. Unfortunately, students often are not solely to blame for their failure. The need for school accountability necessarily acknowledges that many schools fail to educate their students adequately. Therefore, students should be able to seek the only remedy that will do them any good: compensatory damages from their school district for remedial education.

These students should be able to achieve this goal by establishing an implied private cause of action for educational malpractice under the AAA Statute, which is responsible for creating the WASL assessment and accountability system. The AAA Statute expresses a clear school district duty to individual students while insisting on school accountability. The Washington Supreme Court applies the *Bennett* test to determine the existence of an implied private cause of action. Also, a denied diploma is a very concrete and individual harm that cannot be remedied by later school improvement. Thus, the unique convergence of

235. *See supra* note 211.
the AAA statute, the *Bennett* test, and the absence of insurmountable legal or policy barriers should allow a student to successfully sue his or her school district for damages to fund a remedial education.
IN THE LITIGATION BUSINESS: INSURANCE COMPANY LIABILITY FOR ACTS OCCURRING IN THE COURSE OF LITIGATION UNDER THE WASHINGTON CONSUMER PROTECTION ACT

Kasey D. Huebner

Abstract: Insurance companies generally have much greater bargaining power and resources than individual insureds. When a claim by an insured against an insurance company fails to settle amicably and is followed by a lawsuit, the insured has few options should the insurance company behave unfairly or deceptively in the course of the litigation. The Washington Consumer Protection Act protects consumers from deceptive and bad faith acts by businesses, including insurance companies. Although Washington courts have created a general exception disallowing CPA suits for acts occurring in the course of litigation, Washington case law has not directly or clearly addressed whether this litigation exception applies in the insurance context. This Comment argues that the CPA allows for a cause of action by insureds against insurance companies for bad acts occurring after a suit has been filed. The rationale behind the litigation exception to the CPA does not apply to the insurance industry, and Washington statutes, regulations, and case law support allowing CPA suits for unfair or deceptive acts by an insurance company occurring in the course of litigation.

Dana was driving alone in her car when another motorist, Tom, hit her from behind while driving at 35 miles per hour. As a result, Dana suffered a broken sternum, knee injuries, and a concussion, and her car was totaled. Tom admitted to Dana that he was uninsured, so Dana made a claim with her insurance company under her uninsured motorist policy. The insurance company denied Dana’s claim, alleging in good faith that Tom was insured at the time of the accident. Dana continued to assert that Tom was uninsured and sued her insurance company under her uninsured motorist policy. After Dana’s suit had been filed, but before final verdict or settlement, the insurance company discovered that Tom was indeed uninsured, but continued to argue against Dana’s uninsured motorist claim. Dana later discovered that the insurance company had knowledge that Tom was uninsured while it continued to fight her claim. Should Dana be able to assert a claim against her insurance company under the Consumer Protection Act (CPA)?

Washington courts have not addressed whether or not Dana would have a claim against her insurance company under the CPA for deceptive acts occurring in the course of litigation. The Washington State Legislature enacted the CPA to protect consumers from unfair and

1. Hypothetical created by the author.
deceptive business practices. Under the CPA, consumers can bring private suits against individuals and businesses that engage in unfair or deceptive business practices that affect the public. These private CPA suits allow for recovery of up to treble damages, as well as attorney’s fees. Both the Insurance Commissioner of Washington, pursuant to a legislative grant of authority, and Washington courts have determined that private CPA suits may be brought against insurance companies.

However, the Washington Court of Appeals has characterized litigation between a consumer and a business as a private dispute between individuals. The court has held that because litigation is a private dispute, acts by a business occurring after a consumer has filed a suit cannot subject the business to liability under the CPA. The court has also held that there is an exception to the litigation exception for acts that the Legislature has determined to be per se unfair or deceptive. In many cases in which a business acts deceptively during litigation, consumers must rely upon Court Rule 11 (CR 11) to provide a means for recovery.

This Comment discusses whether insureds should be able to bring CPA claims against insurance companies for acts occurring in the course of litigation and argues that the litigation exception to the CPA should not apply in the insurance context. Legislative and judicial findings specific to the insurance industry suggest that an insured should be able to fulfill the requirements of a CPA cause of action against an insurance company for acts occurring in the course of litigation. Even assuming that the litigation exception does apply in the insurance context, insureds should still be able to bring suit against insurance companies under the exception to the litigation exception because the Insurance Commissioner has outlined specific acts which are per se unfair or deceptive in the business of insurance. Judicially and legislatively created policies regarding the insurance industry, and the fact that CR

---

3. See infra notes 26–35 and accompanying text.
4. WASH. REV. CODE § 19.86.090. Treble damages, however, are capped at $10,000. Id.
5. Id.
8. Id.
10. See infra Part II.B.
11. See infra Part IV.A.
12. See infra Part IV.B.
13. See infra Part IV.C.
Suing Insurers Under the CPA

11 does not provide insureds with sufficient remedies, lend further support to the argument that insurance companies should be liable under the CPA for their acts occurring in the course of litigation.

This Comment argues that the CPA establishes a cause of action against insurance companies for deceptive acts occurring after a suit has been filed by an insured. Part I describes the history of the CPA and includes a discussion of the five requirements for a private claim under the Act. Part II discusses the litigation exception to CPA causes of action and alternate means of recovery for deceptive acts occurring after a suit has been filed. Part III explains how the CPA applies to the insurance industry. Part IV argues that the litigation exception to the CPA should not apply to the insurance industry because the justification for the exception does not apply in the insurance context, current policies regarding the insurance industry support allowing suit in this situation, and current remedies under CR 11 are not sufficient.

I. THE WASHINGTON STATE CONSUMER PROTECTION ACT

The Legislature enacted the Consumer Protection—Unfair Competition and Acts forty years ago to protect consumers from unfair or deceptive business practices. In 1986, the Washington Supreme Court established the current five-part test for a private cause of action under the CPA. This test requires a complainant to show an unfair or deceptive act or practice, occurring in the course of trade or commerce that affects the public interest and causes harm to the consumers’ business or property. Subsequent cases have acknowledged the liberal construction afforded the CPA due to the substantial public interests involved.

14. See infra Part IV.D.
17. Id.
18. See infra notes 36-40 and accompanying text.
19. See infra notes 61-64 and accompanying text.
A. The Enactment, Purpose, and Application of the CPA

In 1960, when the Attorney General of the United States held a national conference on consumer protection for state attorneys general,20 few states had promulgated any significant systems of consumer protection laws.21 Estimates in Washington at the time determined that consumers in the state lost approximately $40 million per year to unfair and deceptive business practices.22 Washington Governor Albert D. Rosellini established the Citizen’s Advisory Council on Consumer Protection (the Council) to aid the government of Washington in establishing consumer protection legislation.23 Relying upon the Council’s study and recommendations,24 the Washington State Legislature enacted the CPA in 1961 to deter deceptive and fraudulent acts by businesses and to encourage fair competition.25

The CPA provides a means of recovery against companies for unfair business practices.26 The CPA regulates all businesses in Washington, unless another governing board directly exempts specified business acts from coverage.27 Under the CPA, unfair or deceptive business practices and unfair competition are unlawful.28 The Legislature limited the business acts subject to the CPA by declaring that reasonable business acts and practices and business acts that do not injure the public interest are not prohibited by the CPA.29

Either the State Attorney General30 or private individuals31 may bring a CPA suit against a business that employs unfair or deceptive practices,

20. WASHINGTON CONSUMER ADVISORY COUNCIL, CONSUMER PROTECTION IN THE STATE OF WASHINGTON: A REPORT TO GOVERNOR ALBERT D. ROSELLINI 1 (1960).
21. Id. at 2 (finding that California, New York, Texas, and Wisconsin were among the few states with strong and comprehensive consumer protection laws in place at the time).
22. Governor Rosellini’s Press Files, July 1, 1960 (on file with the Washington Law Review) [hereinafter Press Files]. The Governor’s historical files were searched by staff in Olympia; the resulting documents were sent to the author labeled only as “Governor’s Rosellini’s Press Files” with the date.
23. Id.
24. See id. at 1–2.
25. 1961 WASH. LAWS 216 (codified at WASH. REV. CODE §§ 19.86.010–.920 (2000)).
29. Id. § 19.86.920.
30. Id. § 19.86.080.
Suing Insurers Under the CPA

in order to enjoin the business from further unfair practices and to recover damages from the business. The CPA allows plaintiffs to recover from companies that engage in unfair or deceptive business practices. A consumer suing under the CPA may receive attorney’s fees and up to ten thousand dollars in treble damages. The Legislature implemented the attorney’s fees and treble damages provisions in an attempt to encourage private citizens to bring suit under the CPA.

When the CPA was enacted, the Legislature directed courts to interpret its provisions liberally. Courts in Washington have repeatedly applied this legislative mandate to support a variety of claims under the CPA. Due to this liberal construction, courts have allowed CPA suits between a variety of parties, including an insurance company against a chiropractor involved in insurance fraud, a physician against a drug company, and a class of consumers against a cruise line.

31. Id. § 19.86.090.
32. Id. §§ 19.86.080–090.
34. WASH. REV. CODE § 19.86.090 (providing that plaintiffs may recover “reasonable attorney’s fee[s], and the court may in its discretion, increase the award of damages to an amount not to exceed three times the actual damages sustained”); see also Press Files, supra note 22, March 20, 1961 (“[P]rivate parties may recover damages for injury caused by violations. The measure gives the court discretion to award triple damages.”).
35. See Lybeck, supra note 33, at 278; Koontz, supra note 27, at 937.
36. WASH. REV. CODE § 19.86.920 (2000) (requiring the CPA to be “liberally construed that its beneficial purposes may be served”); see also Press Files, supra note 22, March 20, 1961 (“The act is to be liberally construed to compliment the judicial interpretation of the federal acts . . . . It will not forbid reasonable business practices.”) (internal quotation marks omitted).
B. The Five Elements of a CPA Cause of Action

In *Hangman Ridge Training Stables, Inc. v. Safeco Title Insurance Co.*, the Washington Supreme Court determined that to establish a CPA claim, a complainant must prove the following: (1) the business engaged in an “unfair or deceptive act or practice; (2) occurring in trade or commerce; (3) public interest impact; (4) injury to plaintiff in his or her business or property; and (5) causation.” Failure to demonstrate any of the five necessary requirements precludes a cause of action under the CPA.

To demonstrate the first element of a CPA claim—an unfair or deceptive act or practice—the complainant must establish that an act or practice has the capacity to deceive the general public or, alternatively, that the act is per se unfair or deceptive. Courts do not require a showing of an intent to deceive in order to establish an unfair or deceptive act or practice; rather, a complainant may show that an act or practice has the “capacity to deceive” a significant portion of the general public. In the alternative, if a statute or regulation describes an act or practice as unfair or deceptive, a complainant can establish the unfair or deceptive requirement per se. For example, the Insurance Commissioner has declared specific acts by insurance companies in the

---

43. *Hangman Ridge*, 105 Wash. 2d at 780, 719 P.2d at 532–33.
44. *Id.* at 785–86, 719 P.2d at 535–36.
45. *Id.*
46. *Id.* (emphasis omitted).
47. *Id.*
course of claim settlement to be per se unfair or deceptive acts for purposes of the CPA.\textsuperscript{48}

In cases where the consumer demonstrates that the business used a standardized contract that has a capacity to deceive, a single act between an individual consumer and a business can fulfill the unfair or deceptive act requirement.\textsuperscript{49} In \textit{Nelson v. National Fund Raising Consultants, Inc.},\textsuperscript{50} the court held that the failure of a company to disclose full contract terms to a single consumer until after the contract was executed had the capacity to deceive the general public.\textsuperscript{51} In \textit{Nelson}, the operator of a business failed to disclose the percentage of markup to be paid to the business until after the consumers were contractually bound.\textsuperscript{52} The court found that this failure to disclose the full terms of the contract inherently had the capacity to deceive.\textsuperscript{53}

When evaluating the second element of a private CPA claim, courts have found myriad activities to constitute trade or commerce under the CPA.\textsuperscript{54} The Legislature defined trade or commerce as “the sale of assets or services, and any commerce directly or indirectly affecting the people of the state of Washington.”\textsuperscript{55} When applying the statutory definition of trade or commerce, courts have generally held that the definition should

\textsuperscript{48} \textit{WASH. ADMIN. CODE} § 284-30-330 (2001).


\textsuperscript{50} 120 Wash. 2d 382, 842 P.2d 473 (1992).

\textsuperscript{51} \textit{Id.} at 392–93, 842 P.2d at 478.

\textsuperscript{52} \textit{Id.} at 393, 842 P.2d at 478.

\textsuperscript{53} \textit{Id.}


\textsuperscript{55} \textit{WASH. REV. CODE} § 19.86.010(2) (2000).
be construed liberally.\textsuperscript{56} For example, in \textit{Salois v. Mutual of Omaha Insurance Co.},\textsuperscript{57} the court determined that the phrase “trade or commerce” was intended to encompass more than just acts aimed at inducing a sale and held that the term also includes post-sale acts by businesses.\textsuperscript{58} In addition to acts meeting the statutory definition of trade or commerce, the Washington Supreme Court has held that any act or practice that is per se unfair,\textsuperscript{59} as defined by statute or regulation, also fulfills the trade or commerce element of a CPA claim.\textsuperscript{60}

A complainant can establish the third element—that the acts affect the public interest—either by finding a per se public interest or by using the test established in \textit{Hangman Ridge}\.\textsuperscript{61} A legislative declaration that an act is illegal or affects the public interest fulfills the public interest requirement per se.\textsuperscript{62} If plaintiffs do not meet per se the public interest requirement, they can meet this requirement by finding a pattern of business conduct likely to be repeated or with the potential of affecting more than one member of the public.\textsuperscript{63} Courts use a five-part test to establish whether the public interest requirement is met by a pattern of business conduct. Courts look to whether the conduct was part of the defendant’s business, whether the acts are part of a general course of conduct, whether the repeated acts took place prior to the act involving the plaintiff, whether there is real potential that the act will be repeated after the act involving the plaintiff, and whether, if the act was a single transaction, many consumers were harmed.\textsuperscript{64}

The fourth and fifth elements for a private CPA claim, the injury to property or business and causation requirements, are closely linked. The

\begin{itemize}
\item \textsuperscript{57} \textit{90 Wash. 2d 355, 581 P.2d 1349 (1978)}.
\item \textsuperscript{58} \textit{Id. at 359–60, 581 P.2d at 1351}.
\item \textsuperscript{59} See supra notes 44–48 and accompanying text. An act is per se unfair or deceptive when it is specifically defined as an unfair or deceptive act by a state statute or regulation. \textit{Hangman Ridge Training Stables, Inc. v. Safeco Title Ins. Co.}, 105 Wash. 2d 778, 785–86, 719 P.2d 531, 535–36 (1986).
\item \textsuperscript{60} \textit{Hangman Ridge}, 105 Wash. 2d at 785–86, 719 P.2d at 535.
\item \textsuperscript{61} See generally David J. Dove, \textit{Washington Survey: Washington Consumer Protection Act—Public Interest and the Private Litigant}, 60 WASH. L. REV. 201 (1984) (discussing per se test, as well as the Anhold test, precursor to the \textit{Hangman Ridge} test).
\item \textsuperscript{62} \textit{Id. at 203; see also Hangman Ridge}, 105 Wash. 2d at 791, 719 P.2d at 538.
\item \textsuperscript{63} \textit{See Hangman Ridge}, 105 Wash. 2d at 790–91, 719 P.2d at 538.
\item \textsuperscript{64} \textit{Id. at 790, 719 P.2d at 537–38}.
\end{itemize}
Suing Insurers Under the CPA

injury or harm element\textsuperscript{65} requires that the plaintiff demonstrate some degree of actual damage to his or her business or property.\textsuperscript{66} Emotional damages associated with business or property damage are not recoverable under the CPA.\textsuperscript{67} Finally, the plaintiff must prove that the defendant’s deceptive or misleading acts caused the damage to the plaintiff.\textsuperscript{68} The defendant need not directly contact, solicit, or have business dealings with the complainant to establish causation.\textsuperscript{69}

II. THE CONSUMER PROTECTION ACT GENERALLY DOES NOT APPLY TO ACTS OCCURRING IN THE COURSE OF LITIGATION

The Washington Court of Appeals has found acts occurring in the course of litigation to be private acts that do not support a CPA cause of action. Courts recognize an exception, however, where an individual act constitutes a per se violation. Although courts do not allow CPA claims for non-per se acts occurring after a suit has been filed, courts may sanction the party under CR 11 for deceptive acts occurring in the course of litigation.

A. The Litigation Exception to the CPA

Washington courts have held that, because litigation is a dispute between private parties, actions occurring during litigation generally cannot support a cause of action under the CPA.\textsuperscript{70} Acts occurring purely between private parties do not demonstrate three of the five necessary elements of a CPA cause of action: (1) the unfair or deceptive act requirement; (2) the requirement that the act occur during trade or commerce; and (3) the public interest requirement.\textsuperscript{71} Consumers cannot

\textsuperscript{65} WASH. REV. CODE § 19.86.020 (2000).
\textsuperscript{66} Hangman Ridge, 105 Wash. 2d at 792, 719 P.2d at 539.
\textsuperscript{69} See id.
\textsuperscript{71} See id. (holding that private litigation fails to establish unfair or deceptive, trade or commerce, and public interest elements of CPA cause of action); Hangman Ridge, 105 Wash. 2d at 794, 719 P.2d at 539–40 (finding that private acts support neither the unfair or deceptive nor the public interest requirements).
show that a private act had the “capacity to deceive” members of the general public because private acts, by definition, affect private individuals rather than the public at large.\textsuperscript{72} Private acts, therefore, fail to satisfy the unfair or deceptive act requirement.\textsuperscript{73} The consumer also cannot establish that the trade or commerce element extends to individualized acts aimed at a particular consumer.\textsuperscript{74} Finally, the consumer cannot satisfy the public interest requirement because generally no public impact follows from a private act.\textsuperscript{75} If, however, the unfair or deceptive act, trade or commerce, and public interest requirements can be established as per se violations, a consumer can establish a suit under the CPA for acts occurring in the course of litigation.\textsuperscript{76}

Washington’s Court of Appeals confronted the problem of a non-per se CPA suit in the litigation context in \textit{Blake v. Federal Way Cycle Center}.\textsuperscript{77} \textit{Blake} concerned acts by the defendant corporations Federal Way Cycle Center (FWCC) and Yamaha Motor Corporation (Yamaha) that occurred after the plaintiffs, Mr. and Ms. Blake, had filed suit regarding a Yamaha motorcycle they had purchased from FWCC.\textsuperscript{78} FWCC and Yamaha requested a continuance.\textsuperscript{79} The court granted the continuance upon the condition that FWCC and Yamaha deposit the amount that the Blakes had paid for the motorcycle into an interest-bearing account.\textsuperscript{80} The Blakes alleged that failure to make this deposit violated the CPA.\textsuperscript{81}

The \textit{Blake} court determined that acts occurring in the course of litigation were private disputes that could not support the unfair or deceptive act, trade or commerce, and public interest requirements of the CPA.\textsuperscript{82} In \textit{Blake}, the actions of FWCC and Yamaha did not violate the CPA until litigation was commenced.\textsuperscript{83} The court determined that the

\textsuperscript{72} See \textit{Hangman Ridge}, 105 Wash. 2d at 794, 719 P.2d at 539–40; \textit{Blake}, 40 Wash. App. at 310–12, 698 P.2d at 582–84.
\textsuperscript{73} See \textit{Blake}, 40 Wash. App. at 310, 698 P.2d at 582–83.
\textsuperscript{74} \textit{Id.} at 312, 698 P.2d at 584.
\textsuperscript{75} \textit{Hangman Ridge}, 105 Wash. 2d at 794, 719 P.2d at 539–40.
\textsuperscript{78} \textit{Id.} at 304–05, 698 P.2d at 579–80.
\textsuperscript{79} \textit{Id.} at 305, 698 P.2d at 580.
\textsuperscript{80} \textit{Id.}
\textsuperscript{81} \textit{Id.}
\textsuperscript{82} See \textit{id.} at 312, 698 P.2d at 584.
\textsuperscript{83} See \textit{id.} at 311, 698 P.2d at 583.
acts of FWCC and Yamaha occurring after the complainants had filed suit did not fulfill the unfair or deceptive act requirement because the actions were aimed directly at the Blakes, did not substantially harm the Blakes, did not have the capacity to deceive a significant portion of the general public, and did not appear to have the potential for repetition. The plaintiffs could not meet the trade or commerce requirement because, once they filed suit, the matter was controlled by the courts and constituted a private dispute between the parties. The court further held that the complainants did not meet the public interest requirement as the alleged acts did not have potential for repetition. Because the plaintiffs in Blake could not establish three of the five elements of a CPA claim, the court determined that they did not have a valid cause of action for a non-per se violation of the CPA.

In contrast, the Court of Appeals has held that per se violations of the CPA occurring after litigation has begun can support a CPA cause of action. In Evergreen Collectors v. Holt, the court created an exception to the rule established in Blake that acts occurring after a suit has been filed could not give rise to a CPA claim. Evergreen Collectors involved a collection agency that filed suit against a consumer for costs associated with collecting money from him. Evergreen Collectors threatened to sue Holt for attorney’s fees if he did not settle the case before it went to trial. Holt then filed a CPA claim against Evergreen Collectors for threatening to sue in an attempt to force settlement of the case. The court determined that the complained of acts affected the public interest because Evergreen Collectors engaged in acts that were statutorily defined as unfair or deceptive acts in trade or commerce.

The court further reasoned that because the business of collection agencies involves bringing suit against consumers, actions occurring in the course of litigation should subject collection agencies to CPA claims. Evergreen Collectors held that where the Legislature has

---

84. See id.
85. See id. at 312, 698 P.2d at 584.
86. See id.
87. Id.
89. Id. at 156–57, 803 P.2d at 13.
90. Id. at 152–54, 803 P.2d at 11–12.
91. Id. at 153–54, 803 P.2d at 12.
92. Id. at 154, 803 P.2d at 12.
93. Id. at 156–57, 803 P.2d at 13.
94. Id.
determined particular acts to be per se unfair or deceptive for purposes of the CPA, a consumer can fulfill the unfair or deceptive act, trade or commerce, and public interest requirements even if the business participates in such acts in the course of litigation. Thus, under *Evergreen Collectors*, consumers can bring per se CPA suits against businesses for acts occurring in the course of litigation.

**B. Alternative Recovery Used for Misleading Acts Occurring in the Course of Litigation**

As recovery is frequently unavailable under the CPA for acts occurring in the course of litigation, a consumer can instead request sanctions under Court Rule 11 (CR 11). CR 11 requires that the attorney and the client certify that they have a reasonable and good faith belief that “pleadings, motions and legal memoranda” are truthful and are not being filed to harass, delay, or needlessly increase litigation costs. Because its provisions apply to both attorneys and clients who sign court documents, CR 11, like the CPA, allows for recovery against businesses for dishonest acts.

The purpose of CR 11 is very different than that of the CPA. While the CPA protects consumers from unfair and deceptive business acts, CR 11 protects the judicial system from abuse by attorneys and litigants. The usual target of sanctions demonstrates this purpose. Although sanctions can be brought under CR 11 against both the party and the attorney, courts have determined that CR 11 sanctions serve to deter attorneys, rather than parties, from abusing the judicial system. As a result, courts frequently impose sanctions against only the attorney, and not the party to the action.

---

95. *Id.* at 155–57, 803 P.2d at 12–13.

96. *Id.*


98. *Id.*


Suing Insurers Under the CPA

The theory behind CR 11 sanctions for violation of court rules stands in stark contrast to the theory of liberally construing and applying damages under the CPA. While CR 11 provides for sanctions for failure to have a reasonable, good faith belief in signed pleadings, motions, and legal memoranda, the CPA allows recovery for unfair or deceptive business acts. Although courts are given a certain amount of discretion in determining the dollar amount of a CR 11 sanction, judges must follow court-imposed guidelines to arrive at that amount. In contrast to the liberal allowance of treble damages allowed under the CPA, sanctions under CR 11 must be the smallest amount possible to deter the bad act.

III. THE CPA AND THE INSURANCE INDUSTRY

Insurance relationships are contracts between insurance companies and their insureds. Commentators have recognized, however, that there is a differential in the bargaining power between insurance companies and the individuals they insure. The Legislature empowered the Insurance Commissioner to develop administrative rules declaring particular insurance company acts and practices to be per se unfair or deceptive under the CPA. The resulting regulations, along with other legislative determinations, combine to enable consumers to establish the five elements of a CPA claim against insurance companies.

A. The Business of Insurance

The nature of an insurance contract is a promise, on the part of the insurance company, to provide future benefits to the insured upon the occurrence of a contingency. Although courts recognize no single definition of what constitutes the insurance business, the standard relationship between an insured and an insurance company is based upon

103. WASH. CR 11.
106. Id. at 356, 858 P.2d at 1085.
108. EMEC FISCHER & PETER N. SWISHER, PRINCIPLES OF INSURANCE LAW, xxi (1986).
109. See 1 ERIC MILLS HOLMES & MARK S. RHODES, HOLMES’S APPLEMAN ON INSURANCE 2d §1.4 (1996); FISCHER & SWISHER, supra note 108, at xxi.
risk allocation.110 Claims by insureds for benefits may not settle amicably and litigation may ensue to enforce payment by the insurance company.111

Insureds and insurance companies resolve conflicts through litigation with such frequency that litigation is seen as a normal part of the insurance industry.112 Washington courts have determined that the insurance industry, like the collection business, includes litigation in the ordinary course of business.113 For example, the Washington Supreme Court has declared that it is nearly impossible to determine whether or not a document was prepared in anticipation of litigation for purposes of the work-product doctrine, because so much of an insurance company’s normal course of business includes litigation.114

Some commentators argue that the insurance industry lends itself to unfair business practices because insurance policies are often long and complex, include terms standardized throughout the industry, and offer insureds little opportunity to negotiate.115 As insurance companies have such great power compared to insureds, the possibility of bad faith acts by insurance companies is great.116 Therefore, regulations such as the CPA117 ensure that insurance companies deal with their insureds fairly and honestly.118

110. See HOLMES & RHODES, supra note 109, §1.4.
111. See TASK FORCE OF THE COMMITTEE ON INSURANCE COVERAGE LITIGATION, SECTION OF LITIGATION, AMERICAN BAR ASS’N, MANUAL FOR COMPLEX INSURANCE COVERAGE LITIGATION §5.01 (1993). Such disputes often settle before trial. See id.
112. See, e.g., Heidebrink v. Moriwaki, 104 Wash. 2d 392, 399, 706 P.2d 212, 216 (1985); see also TASK FORCE OF THE COMMITTEE ON INSURANCE COVERAGE LITIGATION, supra note 111, §5.01.
113. See Heidebrink, 104 Wash. 2d at 399, 706 P.2d at 216. For further support of the proposition that litigation is part of the insurance business, see also Eugene R. Anderson & James J. Fournier, Why Courts Enforce Insurance Policyholders’ Objectively Reasonable Expectations of Insurance Coverage, 5 CONN. INS. L.J. 335, 383 (1998) (“[L]itigation is the bread and butter of insurance companies. In large part, litigation is their business.”).
114. Heidebrink, 104 Wash. 2d at 399, 706 P.2d at 216.
115. See, e.g., HOLMES & RHODES, supra note 109, §3.4.
117. See infra notes 120–79 and accompanying text.
118. See LENCSIS, supra note 107, at viii.
Suing Insurers Under the CPA

B. Establishing a CPA Cause of Action Against Insurance Companies

The Legislature empowered the Insurance Commissioner to enact regulations governing the insurance industry.119 Pursuant to that authority, the Insurance Commissioner enacted a regulation that identifies particular claim settlement acts or practices as unfair or deceptive in the business of insurance.120 Violation of this regulation by an insurance company establishes per se both the unfair or deceptive act and trade or commerce requirements of a CPA cause of action.121 In addition, the Legislature has determined that the insurance industry affects the public interest,122 which automatically fulfills the public interest requirement of a CPA claim.123 Finally, courts have established a presumption of the harm and causation elements when insurance companies act in bad faith.124

1. Pursuant to Statutory Authority, the Insurance Commissioner Has Promulgated Regulations that Govern the Insurance Industry Under the CPA

The Insurance Commissioner has expansive authority to regulate the insurance industry.125 A primary duty of the Insurance Commissioner’s office is to protect and inform consumers in their dealings with insurance companies.126 To achieve the goal of consumer protection, the Insurance Commissioner’s office supervises the drafting and revision of the insurance code and also proposes and enforces insurance regulations.127 Regulations enacted by the Insurance Commissioner establish acts and practices that are per se unfair and deceptive under the CPA in the business of insurance with regard to the settlement of claims.128 By establishing these per se unfair and deceptive acts and practices, the

121. See infra notes 136–44, 150–52 and accompanying text.
124. See infra notes 167–178 and accompanying text.
127. Id. § 284-02-020(2).
128. Id. § 284-30-330.
Insurance Commissioner promotes the goal of consumer protection embodied by the CPA.

Courts in Washington have upheld the Insurance Commissioner’s authority to promulgate regulations that make specific acts by insurance companies per se unfair or deceptive.\(^{129}\) In \textit{Leingang v. Pierce County Medical Bureau},\(^{130}\) the Washington Supreme Court found that the Legislature vested valid rule-making authority in the Insurance Commissioner to determine which insurance company acts and practices were deceptive.\(^{131}\) The Court also discussed the rule-making authority of the Insurance Commissioner in \textit{Tank v. State Farm Insurance & Casualty Co.},\(^{132}\) and found that the Commissioner had the authority to determine that a breach of good faith was a deceptive act within the insurance industry.\(^{133}\) In spite of the broad authority of the Insurance Commissioner to promulgate rules, courts have retained the power to interpret and refine the application of those rules.\(^{134}\)

2. \textit{Fulfilling the Unfair or Deceptive and Trade or Commerce Requirements in Suits Against Insurance Companies}

The Washington Administrative Code (WAC) provisions promulgated by the Insurance Commissioner outline acts that are per se “unfair or deceptive acts or practices in the business of insurance” for purposes of claim settlement.\(^{135}\) The Insurance Commissioner and courts, however, have not provided a definition of “claim settlement.”\(^{136}\) The prohibitions...


\(^{130}\) Id. at 151–52, 930 P.2d at 297–98 (stating the general rule “that violations of insurance regulations are subject to the Consumer Protection Act”).

\(^{131}\) Id. at 386, 715 P.2d at 1136.

\(^{132}\) See, e.g., \textit{id.} at 393, 715 P.2d at 1140 (finding that regulations do not apply to third party claimants); \textit{Neigel}, 82 Wash. App. at 786–87, 919 P.2d at 632 (1986) (adopting holding from \textit{Tank} that third party claimants cannot recover under CPA).

listed in the WAC concern misrepresentations by insurance companies, unreasonable acts by insurance companies when dealing with insureds, lack of good faith on the part of insurance companies, and inappropriate or coercive actions taken by insurance companies during the course of claims investigation. Courts have interpreted these regulations to mean that any act by an insurance company specified in the WAC automatically fulfills the unfair or deceptive act or practice element of the CPA. For a per se violation, an insured need not establish that the insurance company has a pattern of violating the WAC. A single violation is sufficient to establish an unfair trade practice.

Although most cases brought against insurance companies allege per se deceptive acts or practices in violation of the WAC, insureds can bring suits for acts that have the capacity to deceive. The CPA explains that actions by insurance companies are subject to the CPA, unless state insurance law expressly allows or requires the acts. In Bowers v. Transamerica Title Insurance Co., the court held that an insurance company practicing law without a license had the “capacity to deceive” the general public. As the acts of the insurance company constituted a deceptive act or practice in their own right, the court did not need to rely upon the WAC to establish a claim under the CPA. Therefore, the Insurance Commissioner need not define a specific act as deceptive in order for the act to be subject to the CPA.

Any act described in the WAC as per se unfair or deceptive fulfills not only the unfair or deceptive requirement of the CPA, but also the trade or

138. Id. § 284-30-330(2)–(5), (16).
139. Id. § 284-30-330(6), (18).
140. Id. § 284-30-330(7)–(15), (17), (19).
143. Id.
146. 100 Wash. 2d 581, 675 P.2d 193 (1983).
147. Id. at 591–92, 675 P.2d at 200–01.
148. Id.
commerce requirement. The WAC describes acts or practices occurring in the “business of insurance” that are per se unfair. Washington courts have held that any act that is legislatively determined to be an unfair trade practice fulfills per se both the unfair or deceptive and trade or commerce requirements.

In addition to meeting per se the trade or commerce requirement for violations of the WAC, the trade or commerce requirement can also be established without use of the WAC. The Legislature has broadly defined trade or commerce to include “any commerce directly or indirectly affecting the people of the state of Washington.” In Salois v. Mutual of Omaha Insurance Co., the court expanded the definition of trade or commerce as applied to insurance companies to include deceptive acts beyond those aimed at inducing a sale. The Salois court held that the Legislature intended for insurance companies to be held liable even where the act was not deceptive at the time the insurance policy was sold. The court reasoned that the CPA created liability for deceptive actions and bad faith in settling insurance claims as part of the trade or commerce of the insurance business.

3. Determining that the Insurance Industry Is Affected by the Public Interest

In 1982, the Washington Supreme Court determined that the public interest requirement is met per se when there is a “specific legislative declaration” of a public interest. The insurance code declares that “the business of insurance is one affected by the public interest, requiring that

149. See, e.g., Hangman Ridge, 105 Wash. 2d at 785–86, 719 P.2d at 535.
152. WASH. REV. CODE § 19.86.010 (2000).
154. Id. at 360–61, 581 P.2d at 1351–52.
155. Id. at 359–61, 581 P.2d at 1351–52.
156. Id. at 360–61, 581 P.2d at 1351–52. The court stated:
The defendant was not unfair or deceptive in the sale of its policy. But what the plaintiffs purchased was not the pieces of paper constituting the policy. They purchased the potential benefits and security of coverage. When defendant should have rendered those benefits, it, according to the special verdict, engaged in acts of bad faith and breached its duty of fair dealing.

Id.

all persons be actuated by good faith, abstain from deception, and practice honesty and equity in all insurance matters." Washington courts have adopted the statutory assertion that the insurance business directly concerns the public interest.\footnote{WASH. REV. CODE § 48.01.030 (2000).}

The Washington Supreme Court has also determined that a quasi-fiduciary relationship exists between an insured and an insurance company\footnote{See, e.g., Lenzi v. Redland Ins. Co., 140 Wash. 2d 267, 277, 996 P.2d 603, 607–08 (2000); Coventry Assocs. v. Am. States Ins. Co., 136 Wash. 2d 269, 276, 961 P.2d 933, 935 (1998); Kirk v. Mt. Airy Ins. Co., 134 Wash. 2d 269, 276, 961 P.2d 933, 935 (1998); Indus. Indem. Co. of the N.W. v. Kallevig, 114 Wash. 2d 907, 917, 792 P.2d 520, 526 (1990); Tank v. State Farm Fire & Cas. Co., 105 Wash. 2d 381, 386–87, 715 P.2d 1133, 1136–37 (1986).} that requires the insurance company to give the insured’s interests the same consideration it gives its own interests in all matters.\footnote{Van Noy v. State Farm Mut. Auto. Ins. Co., 142 Wash. 2d 784, 791–92, 16 P.3d 574, 578 (2001). But see id. at 798–800, 16 P.3d at 582–83 (Talmadge, J., concurring) (finding no fiduciary duty in first-party insurance context).} Consequently, an insurance company has a duty to “disclose all facts that would aid its insureds” in asserting a claim for insurance benefits.\footnote{See id. at 791–92, 16 P.3d 578; Coventry Assocs., 136 Wash. 2d at 280, 961 P.2d at 937–38; Tank, 105 Wash. 2d at 385–86, 715 P.2d at 1136.} The quasi-fiduciary and good faith duties owed by an insurance company to an insured exist at an even higher level in the context of a first-party insured, where the interests of the insured are directly opposed to the interests of the insurance company.\footnote{Van Noy, 142 Wash. 2d at 791, 16 P.3d at 578. But see id. at 799–800, 16 P.3d 582–83 (Talmadge, J., concurring) (declaring that no such duty exists).} In this context, the insured is especially susceptible to harm should the insurer fail to act in good faith.\footnote{Id. at 793 n.2, 16 P.3d at 578 n.2.} The good faith and fiduciary duties insurers owe their insureds further the legislative goal of protecting the general public from deceptive insurance company business practices.\footnote{Id.}

4. Establishing the Causation and Harm Elements in CPA Cases Against Insurance Companies

In \textit{Safeco Insurance Co. of America v. Butler},\footnote{118 Wash. 2d 383, 823 P.2d 499 (1992).} the court recognized a rebuttable presumption of harm when an insurance company acts in bad faith.
faith toward an insured. The Butler court determined that the insureds would have a nearly impossible burden of proving that they were “demonstrably worse off” as a result of the bad faith without this presumption. In addition, the court found the rebuttable presumption to be the best means to protect the societal interests implicated by the insurance industry, such as the fiduciary duty of an insurer to an insured. Courts apply the rebuttable presumption of harm when an insurance company acts in bad faith in defending an insured under a reservation of rights, in failing to provide a defense for an insured, or in delaying in the reservation of rights for an insured.

Although there are many instances when an insurance company’s bad faith leads to a rebuttable presumption of harm, this rebuttable presumption does not apply in all situations. There is no rebuttable presumption of harm when an insurance company, as a result of a bad faith investigation, refuses to provide coverage to an insured, if it is eventually determined that the coverage denial was correct. In this instance, harm can be established only if the insured can show that he or she incurred expenses as a result of the bad faith investigation. Additionally, an insurer that is eventually found liable on an insurance policy is not presumed to have acted in bad faith.

The courts have not discussed directly the causation requirement of a CPA claim as it relates to the WAC. Generally, a consumer who uses the WAC to fulfill the unfair or deceptive act and trade or commerce requirements must establish a causal link between the company’s deceptive act or practice and the harm suffered by the consumer.

168. Id. at 390, 823 P.2d at 504.
170. See id. at 389–92, 823 P.2d at 503–05.
171. Id. at 390–92, 823 P.2d at 504–05.
175. Id.
177. See supra notes 168–77 and accompanying text; see also Hangman Ridge Training Stables, Inc. v. Safeco Title Ins. Co., 105 Wash. 2d 778, 792–93, 719 P.2d 531, 539 (1986); Strother v.
Suing Insurers Under the CPA

cases where a rebuttable presumption of harm exists, courts also seem to imply a rebuttable presumption of causation,178 which fulfills the causation requirement of a CPA claim.

IV. THE CPA EXCEPTION FOR ACTS OCCURRING IN THE COURSE OF LITIGATION SHOULD NOT APPLY TO THE INSURANCE INDUSTRY

Insureds should be able to bring suit under the CPA against their insurance companies for acts occurring in the course of litigation. The litigation exception to the CPA should not apply to insurance companies due to distinctive Washington case law, statutes, and regulations governing the insurance industry. Even assuming, arguendo, that the litigation exception does apply to the insurance industry, Washington courts have recognized that acts meeting per se the unfair or deceptive, trade or commerce, and public interest requirements of the CPA are excluded from the litigation exception to the CPA. As previously noted, a specialized statutory and regulatory system governs the insurance industry; therefore, all unfair and deceptive insurance company acts outlined in the WAC should be exempted from the CPA’s litigation exception. In addition, established Washington policy regarding insurance companies lends further support to the argument that insurance companies can be sued under the CPA for their unfair and deceptive acts occurring in the course of litigation. Finally, CPA claims, as opposed to CR 11 sanctions, are necessary to provide an insured with sufficient incentive to bring suit and protect the public interest.

A. The Reasoning of Blake v. Federal Way Cycle Center Is Inapplicable in the Insurance Context

The court in Blake v. Federal Way Cycle Center179 determined that the unfair or deceptive, trade or commerce, and public interest requirements

178. See, e.g., Coventry Assocs., 136 Wash. 2d at 285, 961 P.2d at 940 (holding that there is sufficient causation for a CPA claim due to presumption of harm for bad faith investigation by insurance company, even though policy excluded coverage); Kirk v. Mt. Airy Ins. Co., 134 Wash. 2d 558, 562-64, 951 P.2d 1124, 1126-28 (1998) (arguing that in insurance context courts must “set aside traditional rules regarding harm and contract damages because insurance contracts are different” and finding that bad faith of insurance companies “causes harm” due to bad faith breach of contract).
could not be met to establish a CPA claim for acts occurring in the
course of litigation. Nevertheless, this litigation exception should not
apply in the insurance context. In Blake, the court reasoned that litigation
was a private conflict between two parties that could not support a CPA
claim. In the insurance company context, however, an insured should
be able to establish, without fulfilling per se any of the categories, the
unfair or deceptive, trade or commerce, and public interest requirements
for acts occurring in the course of litigation. As such, an insured
should be able to sustain a cause of action for acts by an insurance
company occurring during litigation. In addition, Washington courts
have already established a rebuttable presumption of harm, and impliedly
causation, in the insurance context, thus enabling an insured to satisfy
all five elements of a CPA cause of action.

The Blake court’s reasoning—that the unfair or deceptive requirement
is not met by acts occurring during litigation—should not apply in the
context of insurance litigation because an insurance company’s failure to
honor a standardized contract creates the capacity to deceive the general
public. Washington courts have held that a single deceptive act between
a consumer and a business can support a CPA cause of action if the act
involves a standardized sales presentation or form contract. Most
insurance transactions include the use of standardized contracts
employed throughout the industry and offering the same promise: to
provide benefits to the insured in the case of a contingent occurrence.
Therefore, the failure of an insurance company to provide benefits to an
insured, as promised in a standardized insurance contract, creates a
capacity to deceive the general public and satisfies the unfair or
deceptive requirement of the CPA.

The Blake court’s reasoning that the trade or commerce element
cannot be met for acts occurring during the course of litigation should
not apply to insurance companies given the distinctive nature of the
insurance business. The Blake court held that acts occurring after a suit is
filed are under the domain of the courts and, thus, are no longer in the

180. Id. at 312, 698 P.2d at 584.
181. Id.
182. See supra Part III.B.2.
183. See supra Part III.B.4.
185. See Holmes & Rhodes, supra note 109, §3.4, at 159.
realm of trade or commerce.\textsuperscript{186} In the insurance industry, however, disputes frequently arise between insureds and insurers that cannot be or are not settled before a suit is filed.\textsuperscript{187} This occurs to such an extent that Washington courts have declared litigation to be part of the usual business of insurance.\textsuperscript{188} As litigation is a part of the insurance business, acts by insurance companies occurring in the course of litigation satisfy the liberally-construed Washington definition of trade or commerce, which includes acts beyond those aimed at inducing a sale.\textsuperscript{189}

Blake’s reasoning that acts occurring in the course of litigation cannot fulfill the public interest requirement is less persuasive in the insurance context, because the insurance industry has been determined by statute to affect the public interest. The business in Blake, selling motorcycles, has no impact on the public interest beyond the general purview of the CPA.\textsuperscript{190} In contrast, the Legislature has statutorily defined the insurance business as an industry that affects the public interest.\textsuperscript{191} Furthermore, Washington courts have consistently upheld the Legislature’s determination that the insurance industry affects the public interest.\textsuperscript{192} Therefore, an insured should be able to establish the third element of a CPA cause of action.

The Blake court does not directly address whether a consumer can fulfill the harm and causation requirements of a CPA cause of action when the company’s acts occur in the course of litigation.\textsuperscript{193} Other Washington court decisions, however, indicate that an insured can establish the harm and causation requirements against insurance companies for acts occurring in the course of litigation.\textsuperscript{194} The Washington Supreme Court has established a rebuttable presumption of harm when an insurance company acts in bad faith.\textsuperscript{195} This rebuttable

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{186} Blake, 40 Wash. App. at 312, 698 P.2d at 584.
\item \textsuperscript{187} See supra notes 107–14 and accompanying text.
\item \textsuperscript{188} See, e.g., Heidebrink v. Moriwaki, 104 Wash. 2d 392, 399, 706 P.2d 212, 216 (1985).
\item \textsuperscript{189} See supra notes 54–60 and accompanying text.
\item \textsuperscript{190} See generally Blake, 40 Wash. App. 302, 698 P.2d 578.
\item \textsuperscript{191} WASH. REV. CODE § 48.01.030 (2000).
\item \textsuperscript{193} See generally Blake, 40 Wash. App. 302, 698 P.2d 578.
\item \textsuperscript{194} See supra note 167–78 and accompanying text.
\end{itemize}
\end{footnotesize}
presumption of harm has been used by courts to imply a rebuttable presumption of causation.196 Therefore, proof of bad faith insurance company acts in the course of litigation would satisfy both the harm and causation requirements of a CPA claim.197

B. The WAC Allows Insureds to Use the Per Se Exception of Evergreen Collectors v. Holt To Bring Suit Against Insurance Companies for Acts Occurring During Litigation

Assuming, arguendo, that the insurance industry does fall under the purview of Blake’s litigation exception to the CPA, the Blake exception would not apply to the insurance industry for acts that the Legislature or an administrative agency has determined to be per se unfair or deceptive.198 Like the collection business addressed in Evergreen Collectors v. Holt,199 Washington regulations and statutes establish per se the unfair or deceptive act, trade or commerce, and public interest requirements to a CPA cause of action.200 Thus, in the insurance context, a complainant could satisfy per se the required elements of a CPA claim when the insurance company’s acts occur during litigation.

The WAC outlines specific acts by insurance companies, for purposes of “the settlement of claims” that fulfill per se the unfair or deceptive and trade or commerce requirements of the CPA.201 The WAC does not provide a definition for “the settlement of claims,”202 and courts that have applied the WAC have not attempted to explain what acts constitute claim settlement.203 As a result, no controlling definition of claim settlement has been established in Washington.

196. See supra notes 177–178 and accompanying text.

197. The individual attorney must decide whether it would be in the best interests of the client to amend the current complaint by adding a CPA claim for the acts occurring in the course of litigation or to file a separate CPA claim against the insurer.


200. See supra Part III.B.2–3.

201. WASH. ADMIN. CODE § 284-30-330 (2001); see also supra Part III.B.2.

202. WASH. ADMIN. CODE § 284-30-330; see also id. at § 284-30-320 (listing insurance code definitions).

Suing Insurers Under the CPA

Although the Washington courts and Legislature have been silent on the issue of claim settlement, commentators have indicated that insurance claims frequently result in litigation. Because an insurance policy is based in contract, the insured pays the insurer consideration, so that, if a contingent event occurs, an insured can make a claim against the insurer to perform the contract. In addition to regular contractual duties, insurers owe their insureds a duty of good faith and fair dealing as part of the insurance agreement, which requires the insurance company to settle an insurance claim when liability to the insured is apparent. However, instead of settling claims when they are clearly liable, insurance companies often consider litigation just another step in the cost-benefit analysis used in the settlement of claims by insureds. In other words, for insurance companies, litigation is merely an alternate form of claim settlement.

If claim settlement encompasses litigation, the regulation of claim settlement practices should apply to acts occurring in the course of litigation. An insurance company should not be able to avoid its duties to its insureds simply because the insured has filed suit against the insurance company and the insurance company’s acts occur in the course of litigation. Thus, insureds involved in a lawsuit against their insurance companies are owed the same duties as those who are involved in informal settlement negotiations. Therefore, because settlement in the insurance industry carries over into litigation, the WAC’s description of unfair or deceptive acts in “the settlement of claims” should include acts occurring in the course of litigation.

Because filing a suit does not terminate the duties of an insurance company, the WAC may be used to establish a cause of action against an


205. See Linstrom, supra note 204, at 696–97; Anderson & Fournier, supra note 113, at 379–81; see also supra notes 107–108 and accompanying text.


207. See Henderson, supra note 204, at 21.


209. See Alan I. Widiss, supra note 204, at 90 (“The involvement of an attorney on behalf of an insured does not, and should not, affect [the] responsibilities of an insurance company.”).

insurance company for acts occurring in the course of litigation. The WAC establishes, per se, both the unfair or deceptive act\(^\text{211}\) and the trade or commerce requirements,\(^\text{212}\) even in cases where a single deceptive act occurs between an insurance company and an individual insured.\(^\text{213}\) Insureds can also rely upon the legislative pronouncement that insurance affects the public interest\(^\text{214}\) to establish the third CPA element in suits against insurance companies in the litigation context. By demonstrating an insurer’s bad faith during litigation, an insured could prove a violation of the WAC, thereby establishing per se the first three elements of a CPA claim. Fulfilling per se the first three elements of a CPA cause of action would allow the insured to bring suit under *Evergreen Collectors*.\(^\text{215}\)

Thereafter, insureds would need only demonstrate the causation and harm requirements\(^\text{216}\) to bring suit under the CPA for acts committed in the course of litigation.

C. Established Washington Policy Regarding Insurance Companies Supports Allowing CPA Claims for Acts Occurring After a Suit Has Been Filed

Permitting CPA suits for acts occurring once a suit has been filed would recognize the enhanced duty of good faith and quasi-fiduciary duties insurance companies owe their insureds. The Washington Supreme Court has broadly defined the duty of good faith on the part of insurance companies and has not limited this duty to events prior to the development of an adversarial relationship between the insurer and the insured.\(^\text{217}\) The Legislature has determined that the duty of good faith applies in *all* insurance matters.\(^\text{218}\) During litigation, the interests of the insurance company and the interests of the insured are at odds; however, the duty of good faith still applies due to the public interests affecting the insurance industry.\(^\text{219}\) Washington courts have not excused insurance

---

211. *See supra* notes 135–44 and accompanying text.
212. *See supra* notes 149–51 and accompanying text.
218. *WASH. REV. CODE § 48.01.030.*
companies from acting in good faith toward their insureds, even in the adversarial relationship of litigation.220

Not only have courts recognized a duty of good faith in all insurance dealings, they have also found that this duty is enhanced when applied to insurance company actions directed against their insureds.221 This enhanced duty requires that, in all insurance matters, insurance companies give the interests of their insureds the same consideration they would give their own interests.222 Allowing insureds to bring suit against their insurance companies for unfair and deceptive acts occurring in the course of litigation would serve to support the enhanced duty of good faith owed by insurance companies.

The quasi-fiduciary duty of insurance companies to insureds applies to acts in the course of litigation and a breach of such duty should allow for suit to be brought against the insurance company under the CPA. In addition to requiring that insurance companies give the interests of their insureds the same consideration that they would give their own interests,223 this duty also requires insurance companies to give insureds access to all information that would aid the insured in making a claim for benefits under an insurance policy.224 Courts have not determined that these quasi-fiduciary requirements apply only prior to the filing of a suit by an insured.225 Therefore, if an insurance company were to withhold vital information in the course of litigation, the company would violate its duty to insureds and should be liable under the CPA.

D. The Recovery Available Under the CPA is Necessary To Give Insureds an Incentive To Protect the Public from Insurance Companies that Act in Bad Faith

Remedies under the CPA for deceptive acts occurring after litigation commences against an insurance company would serve to protect the


221. Id.

222. Id.


225. See, e.g., id. at 793 n.2, 16 P.3d at 579 n.2 (finding that duty of insurer to insured may be heightened when interests of the two parties are at odds). But see id. at 799–800, 16 P.3d 582–83 (Talmadge, J., concurring).
public from future harm by insurance companies. The Washington Legislature established the CPA to protect consumers, 226 but CR 11 sanctions protect only the judicial system. 227 Allowing suits under the CPA against insurance companies for bad acts occurring in the course of litigation would provide greater protection for the public than CR 11 sanctions alone provide.

CR 11 sanctions do not allow individuals the same liberal recovery as the CPA. CR 11 allows recovery only of the smallest amount possible to deter the bad act, 228 as the rule was not enacted to deter future harm to the public at large. In addition, the language of CR 11 does not require courts to impose sanctions payable to the opposing party, 229 so the insured may not receive any recovery from CR 11 sanctions. In addition, CR 11 sanctions are often payable by the attorney, rather than the represented business, 230 so such fines may not provide insurance companies with an incentive to curb bad acts during the course of litigation.

Recovery under the CPA for deceptive acts occurring during litigation is more likely to have a deterrent effect upon insurance companies than are CR 11 sanctions. Under a private CPA claim, the insurance company may have to pay as much as treble damages and reasonable attorneys’ fees directly to the consumer. 231 Also, the offending insurance company, rather than the attorney, pays recovery for acts taken during litigation that violate the CPA, as the insurance company is the party to the action. 232 When an insurance company’s bad faith occurs only in the course of litigation, recovery under the CPA ensures that an insurance company will be liable to the same degree, and with the same deterrent

| 226. See generally Dove, supra note 61. |
| 229. WASH. CR 11 (2001) (“T]he court, upon motion or upon its own initiative, may impose upon the person who signed it, a represented party, or both, an appropriate sanction, which may include an order to pay to the other party or parties the amount of the reasonable expenses incurred.”) (emphasis added). |
| 231. WASH. REV. CODE § 19.86.090 (2000). Treble damages, however, are capped at $10,000. Id. |
effect, that the company would have been liable for deceptive acts occurring prior to litigation.

The CPA is needed, in addition to CR 11 sanctions, to protect the public from harm that occurs when insurance companies act unfairly or deceptively during litigation. The Washington Legislature has noted, and the courts have repeatedly affirmed, that the business of insurance greatly affects the public interest. The CPA provides financial recovery substantial enough for individual insureds to bring suit against insurance companies that act unfairly and dishonestly in the course of litigation, which serves to protect the public from these harmful acts. Allowing these suits supports the policies behind the CPA, and provides a mechanism to enforce the duties owed by insurance companies to their insureds. By themselves, CR 11 suits are insufficient for this purpose.

V. CONCLUSION

The litigation exception to the CPA should not apply to the insurance industry. Therefore, insureds should be able to sue their insurance companies under the CPA for its bad faith acts occurring in the course of litigation. Blake should not bar suits by insureds against insurance companies for acts occurring in the course of litigation, as the unfair or deceptive, trade or commerce, and public interest requirements should be able to be fulfilled in the insurance context. Alternatively, a cause of action may lie under the CPA where specific deceptive acts of an insurance company fall into one of the categories outlined in the WAC. Established Washington policies regarding insurance companies warrant a CPA cause of action for deceptive acts occurring during litigation. If the claim of an insured is denied due to an insurance company’s deception in the course of litigation, the insured has little recourse against the insurance company. The CPA was enacted to protect consumers in situations where they had little power and little recourse.

235. See supra Part IV.C.
236. See supra Part III.B.3.
237. See supra Part IV.D.
Therefore, the purpose of the CPA would best be served—and the interests of the public would best be protected—by allowing insureds to bring CPA suits against insurance companies for unfair or deceptive acts occurring in the course of litigation.
CLOSING A DISCRIMINATION LOOPHOLE: USING TITLE VII’S ANTI-RETLATION PROVISION TO PREVENT EMPLOYERS FROM REQUIRING UNLAWFUL ARBITRATION AGREEMENTS AS CONDITIONS OF CONTINUED EMPLOYMENT

Sidney Charlotte Reynolds

Abstract: Courts have long viewed mandatory arbitration agreements (MAAs) as contract provisions that employees may accept or decline based on the common law doctrine of employment at-will. However, employees may see such MAAs as attempts to curtail Title VII rights and may refuse to sign them. Title VII prohibits employers from retaliating against employees who oppose discriminatory employment practices. A legal loophole has developed where some employers seek explicitly or implicitly to exempt themselves from Title VII’s provisions by drafting MAAs that eliminate statutory rights and remedies from the arbitration process or deter employees from filing discrimination claims altogether. The U.S. Supreme Court has declared such MAAs to be contrary to the policies and purposes of Title VII. At the same time, courts have broadly construed Title VII’s anti-retaliation provision to protect employees who might not be protected under the plain language of the provision. This Comment argues that Title VII’s anti-retaliation provision should protect employees who have a reasonable belief that the MAAs they oppose are unlawful under Title VII and its policies and purposes.

Mary has worked at XYZ Inc. for five years.1 XYZ, noting a recent employer trend of incorporating arbitration clauses in employment contracts, has decided to require all its employees to sign a mandatory arbitration agreement promising to submit any future discrimination claims to arbitration. However, the agreement contains a provision that the arbitrator may award no more than $5000 in punitive damages if an employee proves intentional discrimination. Mary recognizes the agreement as unlawful and tells this to her employer.2 She refuses to sign the agreement.3 Because Mary’s employment is at-will, XYZ may fire her for refusing to sign the agreement. XYZ fires Mary. This Comment addresses whether Mary should have any legal recourse in this situation.

A mandatory arbitration agreement (MAA) is a contract clause in which parties agree to submit any claim arising from the contract to

1. Hypothetical created by the author.
2. In the case of intentional discrimination, the statutory cap for damages under Title VII of the Civil Rights Act of 1964 (Title VII) is $300,000. 42 U.S.C. § 1981(b)(3) (2000).
3. At least one employee has filed a wrongful discharge claim after being fired for refusing to sign a mandatory arbitration agreement (MAA). See Lagatree v. Luce, Forward, Hamilton & Scripps LLP, 88 Cal. Rptr. 2d 664, 668 (Cal. Ct. App. 1999).
arbitration instead of filing the claim in court. An MAA usually creates a specific arbitration scheme, including how to choose the arbitrators, what law will govern the dispute, and other guidelines, such as remedial limits. Employers often condition the retention of an employee on the signing of an MAA. However, MAAs in employment contracts sometimes contain provisions that violate the policies and purposes of Title VII of the Civil Rights Act of 1964 (Title VII). The U.S. Supreme Court has declared such MAAs unlawful. Sometimes employees dispute the validity of this kind of MAA once they have a Title VII claim they wish to file in federal court. Federal courts have invalidated many such MAAs and have allowed employees to pursue their claims. However, courts have observed that between the time an employee signs the MAA and the time it is invalidated, the MAA undermines Title VII by shielding employers from the full force of its provisions.

This Comment argues that if an employer fires an employee who refuses to sign an MAA because it unlawfully mandates the waiver of Title VII rights and remedies, then the employee should have a viable retaliation claim against the employer. Part I explains private arbitration systems and employment at-will, and tracks the rise of arbitration in the employment context. Part II examines Title VII and its protections against employment discrimination. Part II continues by addressing Title VII’s prohibition against retaliation against employees who complain of employer practices made unlawful by Title VII and the broad construction courts have afforded this anti-retaliation provision. Part III discusses how Title VII and other statutory schemes limit employers’ ability to enforce MAAs that explicitly or implicitly shelter employers

5. Id.
8. See infra Part III.B.
10. This Comment focuses exclusively upon individuals that are already employed. Whether or not job applicants would have a cause of action for retaliation in such circumstances is beyond its scope. For an example of a case addressing whether a job applicant has a cause of action for retaliation, see Ruggles v. Cal. Polytechnic State Univ., 797 F.2d 782, 786 (9th Cir. 1986) (holding that job applicant does have a cause of action for retaliation in a failure-to-hire situation).
11. See infra Part III.B.
Mandatory Arbitration Agreements

from the full force of Title VII. Finally, Part IV argues that Title VII’s anti-retaliation provision should protect employees who oppose MAAs because those MAAs unlawfully require them to surrender rights or remedies under Title VII. Therefore, an employer who terminates an employee for refusing to sign an MAA made unlawful by Title VII should be liable for violating the statute’s anti-retaliation provision.

I. ARBITRATION, EMPLOYMENT AT-WILL, AND THE RISE OF MAAS

Arbitration is a process by which opposing parties resolve their disputes privately, outside the judicial system. Parties forgo a judicial forum and present their evidence to an arbitrator who acts as decision maker in a less formal and less adversarial setting. There are some distinct advantages to arbitration for both parties, such as efficiency, reduced cost, and quicker dispute resolution.

The traditional employment at-will doctrine allows employers contractual freedom in creating MAAs. Employees who do not agree to submit all employment-related legal claims to arbitration instead of to courts may lose their jobs. Within the past ten years, arbitration has become an increasingly popular method for resolving disputes, and employers have included MAAs in their employment contracts more frequently. Although courts historically were reluctant to uphold such agreements, the 1925 Federal Arbitration Act (FAA) compelled a change of judicial attitude.

A. The Employment At-Will Rule

At-will employment represents the traditional common-law relationship between employer and employee. “At-will” means either party may terminate the employment relationship at any time. It also

13. See id. at 5.
14. Hoffman, supra note 4, at 133.
16. Id. at 4.
19. Id.
means that employers may draft employment contracts tailored to their needs, and employees are free to sign the contracts or reject them in favor of other employment opportunities. Absent an explicit contract requiring employers to dismiss only for just cause, or a state or federal statute prohibiting discrimination, a majority of states, including Ninth Circuit states such as Washington, Hawaii, Montana, and California, recognize employment at-will as the default doctrine.

B. Employers’ Increasing Use of MAAs in Employment Contracts

The doctrine of employment at-will allows employers and employees to sign MAAs in which they agree to submit employment-related claims to arbitration rather than to the courts. Because most states follow the at-will rule, employees may accept or decline to sign the MAA, although declining usually means losing their jobs. As the courts and Congress have expanded the scope of protection against employment discrimination, large judgments, lengthy jury trials, and exorbitant legal costs have made arbitration more attractive to employers, and the use of MAAs in employment contracts has increased. However, the private arbitration systems established in MAAs sometimes unfairly favor employers by deterring employees from filing claims or restricting the remedies the arbitrator may award to a prevailing employee.

25. Fadeyi v. Planned Parenthood Ass’n of Lubbock, Inc., 160 F.3d 1048, 1049 n.11 (5th Cir. 1998) (citing numerous cases utilizing the employment at-will doctrine).
27. See id.
C. The FAA and the Evolving Judicial Attitude Toward MAAs

The judicial attitude toward arbitration agreements has evolved from skepticism to endorsement.\(^\text{30}\) Historically, judges did not trust arbitration because the arbitrator takes the judge’s place in the legal proceedings.\(^\text{31}\) In 1925, however, Congress passed the Federal Arbitration Act (FAA).\(^\text{32}\) The FAA is a federal statute governing the enforcement of arbitration agreements in commercial contracts.\(^\text{33}\) It requires judges to enforce pre-dispute arbitration agreements in commercial contracts unless there is a contract law basis (such as duress or unconscionability) for invalidating them.\(^\text{34}\) However, courts have remained cautious about enforcing MAAs that encompass not just contract claims, but also claims that arise under federal statutes.\(^\text{35}\)

The FAA provides in relevant part that a “contract evidencing a transaction involving commerce to settle by arbitration a controversy . . . shall be valid, irrevocable, and enforceable, save upon such grounds as exist at law or in equity for the revocation of any contract.”\(^\text{36}\) With this language, Congress mandated that courts uphold valid arbitration agreements despite judicial mistrust of arbitration.\(^\text{37}\)

For decades after the passage of the FAA, the U.S. Supreme Court remained skeptical of arbitration, particularly when the claim involved statutorily created rights.\(^\text{38}\) Beginning in the 1980s, the Court began to abandon its mistrust of arbitration and to enforce MAAs in the


\(^{31}\) Turner, supra note 28, at 231 n.3.


\(^{34}\) Id.

\(^{35}\) See infra Part III.A.


\(^{37}\) Turner, supra note 28, at 231–32.

commercial context.\textsuperscript{39} The Court’s opinion in Southland Corp. v. Keating\textsuperscript{40} strongly supports judicial enforcement of arbitration agreements in commercial contracts negotiated by parties of relatively equal bargaining strength.\textsuperscript{41} In Rodriguez de Quijas v. Shearson/American Express, Inc.,\textsuperscript{42} the Court removed any remaining doubts as to its view of MAAs in the commercial context.\textsuperscript{43} The Court mandated that MAAs be enforced unless they contain provisions illegal under traditional contract law.\textsuperscript{44}

Whether the FAA governs individual employment contracts was a long-disputed issue finally settled in March 2001.\textsuperscript{45} Section One of the FAA exempts the contracts of workers “in interstate commerce” from the mandate of the FAA.\textsuperscript{46} For years, courts debated about whether only transportation workers were included, or whether this was standard commerce clause language meant to reach any worker in the stream of commerce.\textsuperscript{47} The Supreme Court, however, has recently held that Section One applies only to contracts of transportation workers, and therefore other employment contracts are governed by the FAA.\textsuperscript{48}

II. RIGHTS AND REMEDIES UNDER TITLE VII

Title VII, a major exception to the employment at-will rule, makes it unlawful for employers to discriminate against certain protected classes of employees, both in hiring and firing practices and in the terms and conditions of employment.\textsuperscript{49} Title VII establishes complex enforcement mechanisms, including a federal agency to receive complaints and


\textsuperscript{40} 465 U.S. 1 (1984).

\textsuperscript{41} Id. at 7.

\textsuperscript{42} 490 U.S. 477 (1989).

\textsuperscript{43} Id. at 481.

\textsuperscript{44} Id. at 483. For example, the Court noted that an MAA would still be voidable if it inherently conflicted with the underlying purposes of a statute. Id.


\textsuperscript{48} Adams, 121 S.Ct. at 1306.

\textsuperscript{49} 42 U.S.C § 2000e-2 (2000).
attempt conciliation, a cause of action for private individuals to sue employers, and extensive damage provisions.\textsuperscript{50} It also provides employees with a cause of action to sue employers who fire them in retaliation for opposing employment practices made unlawful by Title VII.\textsuperscript{51} Courts have construed this anti-retaliation provision broadly to effectuate the purposes of Title VII in ending employment discrimination.\textsuperscript{52}

A. Overview of Title VII Structure and Remedies

Title VII makes it unlawful for employers “to refuse to hire or to discharge any individual . . . because of such individual’s race, color, religion, sex, or national origin.”\textsuperscript{53} This prohibition against discrimination covers intentionally discriminatory treatment, such as refusing to hire an applicant because of his or her race.\textsuperscript{54} It also includes employer practices that have a “disparate impact” on employment conditions for protected groups.\textsuperscript{55} Disparate impact refers to the discriminatory effect of a facially neutral employment practice.\textsuperscript{56} For example, an employer might want to categorize employees by giving them an intelligence test. If the test does not screen for specific job qualifications and negatively affects a protected group’s terms or conditions of employment, it has a disparate impact.\textsuperscript{57}

Congress charged the EEOC with responsibility for enforcing Title VII.\textsuperscript{58} After an employee becomes aware of his or her employer’s alleged unlawful employment practice, the employee must file a complaint with the EEOC within 300 days.\textsuperscript{59} The EEOC investigates employee complaints of discrimination and tries to resolve them by “conference, conciliation, and persuasion.”\textsuperscript{60} If successful, an amicable agreement is made between the employee and employer.\textsuperscript{61} If unsuccessful, the EEOC,
the U.S. Attorney General, or the aggrieved employee has 180 days to file a civil suit in federal court.\textsuperscript{62}

If an employee proves either intentional or disparate impact employment discrimination in court, remedies are extensive. They include back pay and other actual monetary damages the employee has incurred,\textsuperscript{63} injunctions and other equitable relief to stop the unlawful employment practice in the future,\textsuperscript{64} attorneys' fees,\textsuperscript{65} and expert fees.\textsuperscript{66} If the employee proves intentional discrimination, the employee may also receive compensatory damages such as money for "pain and suffering"\textsuperscript{67} and punitive damages up to $300,000 to punish the employer for the wrongful conduct.\textsuperscript{68}

B. Title VII's Anti-Retaliation Provision as an Exception to At-Will Employment

Prohibition of discrimination is not the only exception to the at-will rule.\textsuperscript{69} Anti-retaliation provisions prohibit employers from dismissing employees who make complaints about potentially unlawful employment practices.\textsuperscript{70} Title VII contains such an anti-retaliation provision.\textsuperscript{71} Courts have construed the provision broadly to afford employees substantial protection from employers who fire them for complaining about potentially discriminatory practices under Title VII.\textsuperscript{72}

1. History of the Anti-Retaliation Doctrine

Beginning with the rise of unions in the early part of the twentieth century, courts and legislatures developed anti-retaliation laws as an

\begin{itemize}
  \item \textsuperscript{62} Id. § 2000e-5(f).
  \item \textsuperscript{63} Id. § 2000e-5(g).
  \item \textsuperscript{64} Id.
  \item \textsuperscript{65} Id. § 2000e-5(k).
  \item \textsuperscript{66} Id. § 1988(c).
  \item \textsuperscript{67} Id. § 1981(b)(3). Compensatory damages include: “future pecuniary losses, emotional pain . . . and other nonpecuniary damages.” Id.
  \item \textsuperscript{68} Id. §1981a(b)(1).
  \item \textsuperscript{70} Id.
  \item \textsuperscript{71} 42 U.S.C. § 2000e-3(a).
  \item \textsuperscript{72} See infra Part II.B.3.
\end{itemize}
exception to the at-will rule.  

The NLRA includes a provision making it unlawful for an employer to fire or otherwise retaliate against an employee for complaining to a government agency or for giving testimony in proceedings about potential violations of the NLRA. This language became known as an “anti-retaliation” provision. Many federal statutes prohibiting employment discrimination, including Title VII, the Americans with Disabilities Act, and the Age Discrimination in Employment Act, contain anti-retaliation provisions similar to the one in the NLRA.

2. Title VII’s Anti-Retaliation Provision

The anti-retaliation provision of Title VII forbids discrimination against employees who protest allegedly discriminatory employment practices. The provision states that an employer may not “discriminate against any [employee] . . . because he has opposed any practice made an unlawful employment practice by this subchapter.” Title VII prohibits both intentional discrimination and disparate impact discrimination based on an individual’s “race, color, religion, sex, or national origin.” The anti-retaliation provision enables employees to watch over their employers and report unlawful practices without risking their jobs.

To establish a prima facie case of retaliation under Title VII, an employee must show that: (1) he or she was engaged in a “protected activity” (actions against which an employer may not retaliate); (2) he or she was a victim of “adverse employment action;” and (3) there is a

73. Ellis & Rudder, supra note 69, at 253.
75. 29 U.S.C. § 158(a)(2).
76. Id. § 158(a)(4).
79. Id. §§ 12,101–213.
84. Id. § 2000e-2.
85. Ellis & Rudder, supra note 69, at 243.
causal link between the protected activity and the adverse action. Under
the second element, discharge is unquestionably an “adverse
employment action,” an action that negatively affects terms or conditions
of employment. With respect to the first and third elements, courts have
devoted considerable attention to what constitutes “protected activity”
and whether a causal nexus exists.

3. Judicial Interpretations of the Scope of Title VII’s Anti-Retaliation
Provision

Retaliation liability is much broader than discrimination liability. Courts have construed “protected activity” broadly. Thus, most circuit
courts have protected from retaliation white employees who complain
about discrimination against black co-workers, employees who make
informal complaints, and employees who have a reasonable but mistaken
belief that an employment practice is unlawful. Nevertheless, the
judiciary has placed limits on the scope of retaliation claims it will
entertain by requiring plaintiffs to prove a causal nexus between the
activity and the employer’s alleged retaliatory action.

a. The Broad Interpretation of Protected Activity

Courts have broadly construed “protected activity” not only in terms
of the kind of activity protected from retaliation, but also whom the anti-
retaliation provision protects. Title VII’s anti-retaliation provision
generally protects employees who make informal complaints about

86. See Folkerson v. Circus Circus Enter., 107 F.3d 754, 755 (9th Cir. 1997); EEOC v. Avery
Dennison Corp., 104 F.3d 858, 860 (6th Cir. 1997).
87. 42 U.S.C. § 2000e-3; see also Ruggles v. Cal. Polytechnic State Univ., 797 F.2d 782, 786 (9th
Cir. 1986) (holding that closing of job opening and loss of opportunity to compete for position is
“adverse employment decision”).
88. See infra Part II.B.3.
89. See infra Part II.B.3.a.
90. See NLRB v. Scrivener, 405 U.S. 117, 122 (1972) (“[T]extual analysis [of the NLRA’s anti-
retaliation provision] alone . . . reveals . . . an intent on the part of Congress to afford broad rather
than narrow protection to the employee.”); Payne v. McLemore’s Wholesale & Retail Stores, 654
F.2d 1130, 1139 (5th Cir. 1981).
91. See infra Part III.B.3.a.
92. See infra Part III.B.2.
93. See, e.g., Maynard v. City of San Jose, 37 F.3d 1396, 1403 (9th Cir. 1994); EEOC v. Crown
Zellerbach Corp., 720 F.2d 1008, 1012 (9th Cir. 1983).
intentionally discriminatory actions. It also protects third party employees who are not victims of discrimination, for example white employees who assist black employees with Title VII claims. Interestingly, most circuits include complaints about activities that employees reasonably suspect are unlawful, but that turn out to be lawful. Nevertheless, retaliation claims generally demand a fact-specific inquiry.

A classic protected activity is filing a charge with the EEOC alleging employer discrimination barred by Title VII. The statute explicitly protects this activity. However, employees need not contact the EEOC before Title VII’s anti-retaliation provision protects them. Five circuits have defined “protected activity” to include informal complaints to the employer or to third parties. For example, in EEOC v. Crown Zellerbach Corp., the Ninth Circuit held that a letter complaining of discriminatory activities sent to a customer of the employer was protected activity, even though it was not classic opposition to unlawful practices. In a later case, the Ninth Circuit justified the Crown Zellerbach expansion of protected activity by explaining that a narrow interpretation of protected activity would chill the assertion of employees’ right to be free from discrimination and frustrate the purposes of Title VII.

94. Crown Zellerbach, 720 F.2d at 1012.
95. Maynard, 37 F.3d at 1403.
97. See, e.g., Rodriguez-Hernandez, 132 F.3d at 855; Hunt-Golliday, 104 F.3d at 1014; Reed, 95 F.3d at 1178; Moyo, 40 F.3d at 984.
98. See, e.g., Ruggles v. Cal. Polytechnic State Univ., 797 F.2d 782, 787 (9th Cir. 1986).
101. Robbins v. Jefferson County Sch. Dist., 186 F.3d 1253, 1258 (10th Cir. 1999); Barber v. CSX Distrib. Servs., 68 F.3d 694, 702 (3d Cir. 1995); Sumner, 899 F.2d at 209; Rollins v. Florida Dept. of Law Enforcement, 868 F.2d 397, 400 (11th Cir. 1989); EEOC v. Crown Zellerbach Corp., 720 F.2d 1008, 1014 (9th Cir. 1983).
102. 720 F.2d 1008 (9th Cir. 1983).
103. Id. at 1012–13.
104. Trent v. Valley Elec. Ass’n, 41 F.3d 524, 526 (9th Cir. 1994).
Furthermore, an employee may have a claim for retaliation without being a victim of discrimination.\textsuperscript{105} In \textit{Maynard v. City of San Jose},\textsuperscript{106} a white male employee assisted a black co-worker with a discrimination complaint.\textsuperscript{107} The employer harassed the white employee and decreased his responsibilities.\textsuperscript{108} The court held that Title VII’s anti-retaliation provision protected him because allowing the retaliation to occur against the white employee would perpetuate the harmful effects of the unlawful discriminatory treatment of his black co-worker.\textsuperscript{109}

Finally, the broad interpretation given to anti-retaliation laws has led many courts to find that an employee can sustain a retaliation claim even if no discrimination actually occurs but the employee has a reasonable belief that an employer’s practice is unlawful under Title VII.\textsuperscript{110} This expansive rule contravenes the express language of the statute, which protects those employees who oppose practices made unlawful by Title VII.\textsuperscript{111} For example, in \textit{Moyo v. Gomez},\textsuperscript{112} a prison fired a corrections officer for refusing to obey an order to deny showers to black inmates after work shifts.\textsuperscript{113} The officer filed a retaliation claim under Title VII, alleging that the order denying showers constituted an unlawful employment practice under Title VII.\textsuperscript{114} The EEOC had previously declared that inmates were not “employees” for the purposes of Title VII.\textsuperscript{115} Therefore, the practice he opposed was actually a lawful practice under Title VII.\textsuperscript{116} However, the Ninth Circuit held that the officer stated a cause of action because he demonstrated a reasonable belief that the

\begin{footnotesize}
\begin{enumerate}
\item[105.] See, e.g., Lowrey v. Texas A & M Univ. Sys., 117 F.3d 242, 251 (5th Cir. 1997); Maynard v. City of San Jose, 37 F.3d 1396, 1403 (9th Cir. 1994); Skinner v. Total Petroleum, Inc., 859 F.2d 1439, 1446-47 (10th Cir. 1988).
\item[106.] 37 F.3d 1396 (9th Cir. 1994).
\item[107.] Id. at 1400.
\item[108.] Id.
\item[109.] Id. at 1403 (citing Sullivan v. Little Hunting Park, Inc., 396 U.S. 229, 237 (1969)).
\item[110.] See, e.g., Robbins v. Jefferson County Sch. Dist., 186 F.3d 1253, 1259 (10th Cir. 1999); Rodriguez-Hernandez v. Miranda-Velez, 132 F.3d 848, 855 (1st Cir. 1998); Moyo v. Gomez, 40 F.3d 982, 985 (9th Cir. 1994); Drinkwater v. Union Carbide Corp., 904 F.2d 853, 865 (3d Cir. 1990).
\item[112.] 40 F.3d 982 (9th Cir. 1994).
\item[113.] Id. at 984.
\item[114.] Id.
\item[115.] Id.
\item[116.] See id. Such a practice would likely be vulnerable to challenge under another federal statute or the federal constitution.
\end{enumerate}
\end{footnotesize}
practice was unlawful under Title VII. The court found his belief to be reasonable, even though the law concerning the employment status of inmates was clear.

The court construed “reasonable” broadly because Title VII requires employees neither to be familiar with all the factual and legal issues in a potential case, nor to be able to analyze the case law and come to the correct conclusion. Therefore, the court interpreted “reasonable” to be a mixed standard of objective and subjective elements, where the court examines the facts and law from the average employee’s perspective to determine whether his or her belief was “reasonable.” Eight other circuits have likewise held that a reasonable belief that a practice is unlawful is enough to sustain the plaintiff’s claim, even if the practice complained of turns out to be legal.

b. An Employee Must Establish a Nexus Between the Protected Activity and the Alleged Retaliatory Action

Courts’ broad interpretation of “protected activity” does not imply that every employee challenge to an alleged discriminatory employment practice is a protected activity. An employee must establish the third

117. Id. at 985.
118. Id.
119. Id.
120. Id.
121. Rodriguez-Hernandez v. Miranda-Velez, 132 F.3d 848, 855 (1st Cir. 1998); Hunt-Golliday v. Metro. Water Reclamation Dist. of Greater Chicago, 104 F.3d 1004, 1014 (7th Cir. 1997); Reed v. A.W. Lawrence & Co., 95 F.3d 1170, 1178 (2d Cir. 1996); Meeks v. Computer Assocs. Int’l, 15 F.3d 1013, 1021 (11th Cir. 1994); Drinkwater v. Union Carbide Corp., 904 F.2d 853, 866 (3d Cir. 1990); Booker v. Brown & Williamson Tobacco Co., 879 F.2d 1304, 1312 (6th Cir. 1989); Love v. Re/Max of Am., Inc., 738 F.2d 383, 385 (10th Cir. 1984); Sisco v. J.S. Alberici Const. Co., 655 F.2d 146, 150 (8th Cir. 1981); Payne v. McLemore’s Wholesale & Retail Stores, Inc., 121 S.Ct. 1508, 1509 (2001). The Court did not address the propriety of the “reasonable belief” rule, but it held that the employee’s belief that one isolated comment could constitute an violation of Title VII was not reasonable. Id. at 1510.
122. See, e.g., Folkerson v. Circus Circus Enters., 107 F.3d 754, 756 (9th Cir. 1997) (holding that opposition must be to discriminatory practice of employer, not private individual); Roth v. Lutheran Gen. Hosp., 57 F.3d 1446, 1460 (7th Cir. 1995) (holding that plaintiff’s baseless accusations of discrimination provided no grounds for retaliation claim); Jurado v. Eleven-Fifty Corp., 813 F.2d 1406, 1411 (9th Cir. 1987) (holding that disc jockey’s complaints about radio station programming change were not an “activity protected by Title VII” because they related to the desire to maintain the “value of his radio personality”).
prong of a retaliation claim: causation. Causation means that the employer fired the employee as a result of the protected activity rather than for a legitimate reason. Thus, if the employee is disruptive to the workplace—for example, by blocking access to the employer’s place of business—courts will generally find that the dismissal was legitimate. Also, Title VII’s anti-retaliation provision does not protect employees who, in the guise of “protected activity,” make defamatory statements against their employers. For example, a court has held that Title VII’s anti-retaliation provision did not protect an employee who sent a memorandum to her supervisor containing libelous statements about the employer’s allegedly discriminatory employment practices. Other legitimate reasons for dismissal include unsatisfactory job performance, engaging in disruptive or excessively hostile behavior, or refusing to accept additional work duties. These defenses protect law-abiding employers who fire employees for legitimate reasons, even though circumstances may suggest that the employee has a retaliation claim.

III. MAAS MAY NOT LIMIT STATUTORY RIGHTS AND REMEDIES, PARTICULARLY THOSE UNDER TITLE VII

The U.S. Supreme Court has held that when employees agree to submit federal statutory claims to arbitration, they agree only to a change of forum, not a change of substantive rights and remedies under a statute. The Court has been cautious in examining MAAs involving Title VII claims because federal courts play an important role in enforcing Title VII. Nevertheless, employers sometimes unlawfully

123. Folkerson, 107 F.3d at 755.
125. Id. at 803.
127. Id. at 803.
131. Ellis & Rudder, supra note 69, at 244.
use MAAs to limit, explicitly or implicitly, the substantive rights and remedies available to employees under Title VII.\textsuperscript{134}

\textbf{A. The Supreme Court’s Protection of Federal Statutory Rights and Remedies in FAA Cases}

Although the Supreme Court has become less skeptical of arbitration since the passage of the FAA, it has remained cautious about enforcing MAAs that govern federal statutory claims.\textsuperscript{135} The Court has held that despite the FAA, statutory rights and remedies should be available in arbitration just as they would be in court.\textsuperscript{136} For example, an MAA may not contain a provision that prohibits a plaintiff from filing a charge of discrimination with the EEOC, making a claim under a particular federal statute, or recovering monetary damages.\textsuperscript{137}

In \textit{Alexander v. Gardner-Denver Co.},\textsuperscript{138} an employee participated in an arbitration of his Title VII discrimination claims pursuant to a collective bargaining agreement.\textsuperscript{139} After the arbitrator dismissed his claims, the employee filed a Title VII discrimination suit in federal court.\textsuperscript{140} The Supreme Court held that an individual employee does not waive his or her right to file a civil suit under Title VII simply because his or her union has a collective bargaining agreement to arbitrate contract-based employment claims.\textsuperscript{141} The Court emphasized the importance of Title VII’s purpose and declared that “final responsibility for enforcement of Title VII is vested with federal courts.”\textsuperscript{142}

In the 1980s, the Court continued to enforce MAAs encompassing statutory claims as long as the MAA did not reduce statutory rights and remedies.\textsuperscript{143} For example, in \textit{Mitsubishi Motors Corp. v. Soler Chrysler-}

\begin{itemize}
  \item \textsuperscript{134} See infra Part III.B.
  \item \textsuperscript{135} See, e.g., Gilmer v. Interstate/Johnson Corp., 500 U.S. 20, 26 (1991); Mitsubishi, 473 U.S. at 627; \textit{Alexander}, 415 U.S. at 44.
  \item \textsuperscript{136} \textit{Gilmer}, 500 U.S. at 26; \textit{Mitsubishi}, 473 U.S. at 628; \textit{Alexander}, 415 U.S. at 44.
  \item \textsuperscript{137} See infra Part III.B.
  \item \textsuperscript{138} \textit{Rodriguez de Quijas v. Shearson/Am. Express, Inc.}, 490 U.S. 477, 481 (1989); \textit{Mitsubishi Motors Corp. v. Soler Chrysler-Plymouth, Inc.}, 473 U.S. 614, 628 (1985).
  \item \textsuperscript{139} Id. at 39. Because the arbitration clause was contained in a collective bargaining agreement, the Court has subsequently held \textit{Alexander} to apply only in that context and not in the context of individual employment contracts. See \textit{Gilmer}, 500 U.S. at 34–35.
  \item \textsuperscript{140} \textit{Alexander}, 415 U.S. at 42–43.
  \item \textsuperscript{141} Id. at 43.
  \item \textsuperscript{142} Id. at 44.
\end{itemize}
Plymouth, Inc., a consumer filed a claim in federal district court pursuant to a federal antitrust statute. The defendant, an automobile manufacturer, sought to compel arbitration of the claim in accordance with the terms of an MAA previously signed by the consumer. The Supreme Court held that the MAA was enforceable because it sufficiently protected the consumer’s statutory rights and remedies in arbitration. However, the Court emphasized that MAAs may not be a vehicle for the waiver of federal statutory rights and remedies because “a party does not forgo the substantive rights afforded by the statute [in an MAA]; it only submits to their resolution in an arbitral, rather than a judicial, forum.”

In Gilmer v. Interstate/Johnson Lane Corp., the Court re-emphasized the importance of preserving statutory rights in MAAs. In that case, a securities broker signed a registration form containing an MAA. His employer terminated his employment at age sixty-two, in possible violation of the Age Discrimination in Employment Act, so he filed a claim in federal court. The Court held the agreement enforceable and compelled arbitration. However, the Court acknowledged the Mitsubishi line of cases and reaffirmed that arbitration agreements should constitute only a change of forum, not a reduction of statutory rights.

---

145. Id. at 618.
146. Id.
147. Id. at 628.
148. Id.
150. Id. at 26.
151. Id. at 23. The New York Stock Exchange rules require all brokers to sign this registration form if they are to do business with the exchange. Id.
154. Id. at 23.
155. Id. at 26.
B. Employers Sometimes Seek To Have Employees Sign MAAs To Shelter Themselves from Employees’ Title VII Rights and Remedies

The EEOC has recognized that some employers draft MAAs that “flagrantly eviscerat[e] core rights and remedies” of Title VII.156 Considering the extensive remedies available under Title VII, employers may desire to create arbitration schemes in employment contracts that limit the employer’s liability for future discrimination.157 Employers might do this explicitly by restricting the damages available under Title VII to a level far below the statutory limits.158 They might also do so implicitly by requiring employees to pay up-front arbitration costs so high as to deter employees from filing claims altogether.159 Such MAAs are unlawful according to the rule reaffirmed by the Supreme Court in Gilmer: arbitration should be a change of forum, not a change of rights or remedies.160 However, the FAA allows courts to invalidate commercial arbitration agreements only for reasons available under traditional contract law.161 Therefore, many courts faced with MAAs that do restrict substantive rights and remedies under Title VII have either reformed the contract or invalidated it on breach of contract, public policy, or unconscionability grounds.162

1. Cases Involving Explicit Shelters

Employers sometimes use MAAs to restrict the damages available to employees who successfully show discrimination under Title VII.163 In

156. EEOC, POLICY STATEMENT ON MANDATORY BINDING ARBITRATION OF EMPLOYMENT DISCRIMINATION DISPUTES, NO. 915.002, at 15 (July 10, 1997).
157. See Maltby, supra note 12, at 5–6.
158. See infra Part III.B.1 and accompanying text.
159. See infra Part III.B.2 and accompanying text.
160. Gilmer, 500 U.S. at 26. Although Gilmer dealt with a claim under the Age Discrimination in Employment Act, circuits addressing the question have agreed that Gilmer also applies to Title VII actions. See EEOC v. Frank’s Nursery Crafts, Inc., 177 F.3d 448, 448 n.2 (6th Cir. 1999); Koveleskie v. SBC Capital Mkts., Inc., 167 F.3d 361, 364 (7th Cir. 1999); cf. Duffield v. Robertson Stephens & Co., 144 F.3d 1182, 1198 (9th Cir. 1998) (finding that 1991 Amendments to Civil Rights Act reflect Congress’s intent to codify and extend Gardner-Denver rule excluding enforcement of MAAs in Title VII cases).
163. Hooters, 39 F. Supp. 2d at 599; Paladino v. Avnet Computer Techs., Inc., 134 F.3d 1054, 1060 (11th Cir. 1998); Stirlen, 60 Cal. Rptr. 2d at 158–59.
Hooters of America, Inc. v. Phillips, the MAA granted the employer complete discretion in designing the arbitration rules. The rules the employer adopted limited virtually every remedial measure in Title VII to amounts far below what Title VII allows. For example, the employer’s rules capped punitive damages at $13,000 while the Title VII statutory cap is $300,000. Also, the rules precluded the arbitration panel from ordering any injunctive relief, such as an order to stop Hooters from engaging in discriminatory policies or practices, where Title VII explicitly provides for such relief. The district court held that these damage restrictions were inadequate to protect federal statutory rights as required by Gilmer. Therefore, the subsequent rules were unconscionable and violated Title VII’s public policies. The Fourth Circuit Court of Appeals affirmed the district court decision. Although the appeals court made no comment on the district court’s findings of unconscionability and violation of public policy, it did hold that the employer had completely breached its contractual duty to arbitrate by “creating a sham system unworthy even of the name of arbitration.”

Other employers have also unsuccessfully attempted to include restrictions on Title VII damages in MAAs. For example, in Derrickson v. Circuit City Stores, Inc., the MAA provided for only one year of back pay and capped punitive damages. The district court refused to enforce the MAA and denied the employer’s motion compelling arbitration because the arbitration agreement severely limited the employee’s remedies under Title VII. The court explained that

165. Id. at 599.
166. Id. at 599–60.
167. Id.
168. Id.
169. Id. at 622 (referring to Gilmer court’s standards for cases involving arbitration of statutory rights as authority for invalidating MAA at preliminary injunction hearing rather than allowing arbitration to go forward).
170. Id. at 623.
172. Id. at 940.
175. Id.
176. Id.
such a provision sheltered the employer, stating, “[The MAA] shields Circuit City . . . and prevents Plaintiff from effectively vindicating her rights.” Therefore, the court invalidated the provision based on the Supreme Court’s rule in *Gilmer* and the public policies of Title VII. In *Paladino v. Avnet Technologies*, the MAA at issue was even more restrictive, limiting damages to breach of contract only. The Eleventh Circuit, also citing *Gilmer*, held that the district court properly denied enforcement of the provision because it was “fundamentally at odds with the purposes of Title VII.”

2. Cases Involving Implicit Shelters

Some MAAs implicitly deter employees from filing Title VII discrimination claims by requiring the employees to pay arbitrator’s fees. Although in federal court the services of a judge are free to the litigants, arbitrator’s fees can run between $500 and $1,000 per day. Paying some or all of these fees is beyond the means of many employees and may make arbitration prohibitively expensive, thereby discouraging employee claims. The D.C. Circuit addressed this practice as a matter of first impression in *Cole v. Burns International Security Services*. In that case, a former employee filed a federal court complaint for race discrimination under Title VII despite the fact that his employment
contract contained an MAA. The D.C. Circuit court affirmed the district court’s decision to compel arbitration but ruled that the employer could not compel the employee to pay some or all of the arbitrator’s fees. The court took particular notice of the Gilmer requirement that arbitration should constitute a change of forum, not a change of rights and remedies. It then concluded that an MAA requiring the employee to pay some or all of the arbitrator’s fees “would surely deter the bringing of arbitration and constitute a de facto forfeiture of the employee’s statutory rights.” The court found that to enforce such an MAA would be to change the rights afforded employees by Title VII. The Tenth and Eleventh Circuits have also invalidated MAAs that failed the Cole test because they required the employees to pay the arbitrator’s fees.

Recently, in Green Tree Financial Corp.-Alabama v. Randolph, the Supreme Court addressed the arbitration cost issue in a consumer dispute over an arbitration clause. The Court did not cite Cole directly but used similar reasoning to suggest that a party may seek to invalidate an MAA on the basis of prohibitive fees. In Green Tree, the consumer argued that burdensome arbitrator’s fees made the MAA void because deterring Title VII claims violates the public policies behind Title VII. The Court agreed with this theory but found that the consumer failed to offer any evidence as to what type of arbitration system would be used or what the fee assessment procedure was likely to be. Presumably, an employee who could present an MAA that explicitly required the arbitrator to assess fees to the employee, and who could prove that the fees would prevent him or her from filing a claim, would meet the burden that Green Tree established.

187. Id. at 1467.
188. Id. at 1484.
189. Id. at 1481.
190. Id. at 1468. The court upheld the agreement at issue in Cole because it construed the language relating to arbitrator’s fees to require the employer to pay. Id. at 1486.
191. Id. at 1482.
194. Id. at 86.
195. Id.
196. Id.
197. Id.
198. Id. at 87 n.6.
However, an MAA that merely requires sharing arbitration costs may be lawful. Courts have distinguished *Cole* when plaintiffs could not produce evidence that paying one half of the arbitrator’s fees would prevent them from filing a claim.\(^{199}\) A requirement that the employee pay costs that he or she would have had to pay in court does not alter that employee’s access to a forum.\(^{200}\) Thus, these cases are consistent with *Cole*, which is premised on the Supreme Court’s test articulated in *Gilmer*. If the fee burden does not constitute a waiver of forum, and therefore a waiver of rights and remedies, it is probably valid.\(^{201}\)

IV. IF AN EMPLOYEE IS FIRED FOR REFUSING TO SIGN A POTENTIALLY UNLAWFUL MAA, THE EMPLOYEE SHOULD HAVE AN ANTI-RETIHALATION CLAIM

After such cases as *Hooters of America, Inc. v. Phillips*,\(^{202}\) in which the Fourth Circuit denounced the employer’s arbitration rules as “so one-sided that their only possible purpose is to undermine the neutrality of the [arbitration],”\(^{203}\) courts have begun to recognize the potential for abuse in employers’ use of MAAs. The judiciary’s ultimate conclusion is that MAAs seeking to limit Title VII rights or remedies are unlawful and will not be enforced.\(^{204}\) Furthermore, courts have warned that if an employer succeeds in forcing an employee to sign an unlawful MAA, the employee may be deterred from filing a discrimination claim.\(^{205}\) However, if the employee points out the unlawful nature of the MAA and refuses to sign it, the employer may discharge or discipline the employee in the name of employment at-will. This Comment argues that in order to resolve this “Catch-22,” the employee should have a viable cause of action for retaliation against the employer under the anti-retaliatiion provision of Title VII, which protects employees who oppose employment practices made unlawful by Title VII.


\(^{201}\) See id. at 1482.

\(^{202}\) Id. at 1465, 1484 (D.C. Cir. 1997).

\(^{203}\) 173 F.3d 933 (4th Cir. 1999).

\(^{204}\) Id. at 938.

\(^{205}\) See supra notes 182–185 and accompanying text.
Admittedly, this solution may push Title VII’s anti-retaliation provision into uncharted territory. Traditional retaliation claims generally arise when employers discipline or fire employees for complaining about patently discriminatory behavior, such as refusing showers to black workers but not white workers. Furthermore, the statutory language only refers to employees who complain of “practice[s] made unlawful” by Title VII, which Title VII defines as intentional or disparate impact discrimination.

However novel the idea may seem, allowing a cause of action for retaliation for MAA opposition is not without legal support. Although some disputed MAAs may be neither intentionally discriminatory nor have a disparate impact, courts have found MAAs that reduce statutory rights and remedies unlawful due to the strong purposes and policies of Title VII. At the same time, courts have broadly construed the language of the anti-retaliation provision to protect employees in situations outside the bounds of Title VII’s plain language. Therefore, employee opposition to an MAA that explicitly or implicitly shelters an employer from Title VII rights or remedies should give rise to a viable Title VII retaliation claim, even if the MAA in question does not constitute intentional or disparate impact discrimination. Without limiting legitimate employer defenses, recognizing such claims would both prevent employers from circumventing the protections and purpose of Title VII and encourage employers to draft lawful arbitration agreements.

A. An Employee’s Opposition to an MAA Made Unlawful by Title VII or Its Policies Could Give Rise to a Claim for Retaliation

An analysis of the lawfulness of any MAA or the merits of any Title VII retaliation claim requires a highly fact-specific inquiry. Thus, it is

---

206. One employee has filed a wrongful termination claim after his employer fired him for refusing to sign an MAA. Lagatree v. Luce, Forward, Hamilton & Scripps LLC, 88 Cal. Rptr. 2d 664, 668 (Cal. Ct. App. 1999). The employee argued that requiring him to sign the MAA—which he believed to be unlawful—as a condition of employment, was contrary to California public policy. Id. The court declared that allowing a plaintiff to “strike gold” with such a wrongful termination suit was “absurd.” Id. at 681. However, in that case, the court found the MAA in question to be lawful and enforceable. Id.

207. Cf. Moyo v. Gomez, 40 F.3d 982, 984 (9th Cir. 1994).


209. See supra Part III.B.

210. See supra Part II.B.3.a.

211. See, e.g., Moyo v. Gomez, 40 F.3d 982, 985 (9th Cir. 1994).
Mandatory Arbitration Agreements

essential to distinguish the various situations in which an employee fired for opposing an unlawful MAA may have a viable retaliation claim. The four fact patterns below recall Mary (introduced at the outset of this Comment) and serve as useful illustrations of the analysis that follows. The first two situations represent more traditional retaliation claims, where the employee reasonably believes a practice is either intentionally discriminatory or has a disparate impact. The second two situations demonstrate why Title VII’s anti-retaliation provisions should protect employees who oppose MAAs made unlawful by the policies behind Title VII, because these MAAs explicitly or implicitly shelter the employer from employees’ rights and remedies.

First, consider a situation where XYZ presents Mary with an MAA that states, “By agreeing to arbitration, black employees agree not to file any claim of discrimination in federal court.” She tells XYZ that this practice is discriminatory and refuses to sign. XYZ fires Mary for refusing to sign the agreement. This represents a traditional prima facie retaliation case. Mary recognizes that the MAA itself is facially and intentionally discriminatory, has complained to that effect, and has been fired as a result. While few employers would offer such a patently discriminatory MAA, this example illustrates that traditional retaliation claims limit MAAs despite employment at-will arguments.

Second, consider a situation where XYZ presents Mary with an MAA that states, “By agreeing to arbitration, black employees agree not to file any claim of discrimination in federal court.” She tells XYZ that this practice is discriminatory and refuses to sign. XYZ fires Mary for refusing to sign the agreement. This represents a traditional prima facie retaliation case. Mary recognizes that the MAA itself is facially and intentionally discriminatory, has complained to that effect, and has been fired as a result. While few employers would offer such a patently discriminatory MAA, this example illustrates that traditional retaliation claims limit MAAs despite employment at-will arguments.

Second, consider a situation where XYZ presents an MAA that states, “Employees waive their right to file Title VII discrimination claims in federal court and agree to submit such claims to arbitration.” Mary tells XYZ that use of the MAA is a discriminatory employment practice because it will deter her female and minority co-workers from fully enforcing their legal rights and refuses to sign the agreement. XYZ fires Mary. This kind of agreement represents a more common MAA, waiving rights to file a lawsuit in federal court. Federal courts have found that such MAAs are not unlawful as long as they constitute only a change of forum. Nevertheless, Mary has opposed the MAA reasonably, believing it to be discriminatory because it may prevent black co-workers from bringing a claim. In this situation, the dispositive issue is whether Mary’s belief that the MAA is unlawful on a theory of disparate impact is a reasonable belief.

---

212. See supra note 86 and accompanying text.
213. The court’s analysis will be the same regardless of whether Mary complains to her employer, the EEOC, or a third party about the alleged unlawful practice. See supra notes 98–101 and accompanying text.
Third, consider a situation where XYZ presents an MAA that states, “If the employee proves intentional employment discrimination on the part of the employer, the arbitrator may award no more than $1000 in punitive damages.” Mary tells XYZ that this is unlawful because she heard that in successful discrimination cases plaintiffs can get as much as $300,000 in punitive damages. She refuses to sign the agreement. XYZ fires Mary. Such a provision is unlawful because it explicitly shelters XYZ from the remedies of Title VII and therefore contravenes Title VII’s public policies. Courts can avoid enforcing such provisions in MAAs using a variety of contract doctrines, including doctrines of construction, public policy, and unconscionability. However, with respect to the first requirement of a prima facie claim. Title VII’s anti-retaliation provision protects employees who complain about a “practice made unlawful” by Title VII. In other words, “protected activity” includes complaints about a practice of intentional or disparate impact employment discrimination, but nothing in the statutory language refers to protection for employees who oppose practices made unlawful due to the policies behind Title VII. Thus, the critical issue is whether Mary’s opposition to a practice that is unlawful based on Title VII’s policies is “protected activity” for retaliation purposes.

Finally, consider a fourth situation in which the MAA states, “The employee and employer must submit all legal claims arising from employment to arbitration. Any and all arbitrator’s fees shall be paid by the party bringing a claim.” Mary informs XYZ that she will not sign the MAA because the fees may be excessive and prohibit her or her co-workers from vindicating their rights. XYZ fires Mary for refusing to sign. Under the standard set forth in Green Tree Financial Corp.-Alabama v. Randolph, there is nothing inherently unlawful about an MAA of this type, unless Mary could show a court that the fees would be unduly burdensome to her rights under Title VII. Yet, Mary has a reasonable belief that the MAA’s fee provision will preclude her ability to bring a discrimination claim. That is, she believes the MAA implicitly shields XYZ from the force of Title VII. Again, the critical issue is

214. See supra notes 163–172 and accompanying text.
215. See supra Part III.B.
216. See supra note 86 and accompanying text.
219. Id. at 87 n.6.
whether Mary’s opposition is “protected activity” for the purpose of Title VII’s anti-retaliation provision.

B. Title VII’s Anti-Retaliation Provision Should Protect Employees Who Have a Reasonable Belief That an MAA Is an Unlawful Practice Under Title VII or Its Policies

While the first factual scenario encompasses a traditional retaliation claim, where the discriminatory practice complained of is obvious, the other three present more challenging issues likely to arise in the context of an employee opposing an MAA. If an employee believes an MAA is intentionally discriminatory or has a disparate impact, as in the second scenario, Title VII’s anti-retaliation provision should protect him or her if the belief is reasonable, even if the MAA is legal.\(^{220}\) The issue raised in the third and fourth scenarios is whether the provision should protect employees who oppose employment practices made unlawful by the policies behind Title VII, rather than the plain language of Title VII itself. In other words, is an employee’s opposition “protected activity” when the employee opposes a practice that contradicts the policies of Title VII? Retaliation liability is broader than discrimination liability.\(^{221}\) Anti-retaliation provisions extend to a number of circumstances not expressly addressed by Title VII and effectuates Congress’s purpose in protecting employees who report potentially unlawful practices.\(^{222}\) Therefore, courts should extend this same protection to employees who reasonably oppose MAAs that explicitly or implicitly shelter employers from employee rights and remedies in violation of the policies and purposes of Title VII.

I. Title VII’s Anti-Retaliation Provision Should Protect an Employee Who Has a Reasonable Belief That an MAA Is an Unlawful Employment Practice Under Title VII

Title VII’s anti-retaliation provision should protect employees fired for opposing MAAs they reasonably believe are intentionally discriminatory or have a disparate impact on certain classes of employees. According to \textit{Moyo v. Gomez} \(^{223}\) the broad protection of Title

\(^{220}\) See \textit{supra} note 121 and accompanying text.

\(^{221}\) See \textit{supra} Part II.B.3.a and accompanying text.

\(^{222}\) See \textit{supra} Part II.B.3.a.

\(^{223}\) 40 F.3d 982 (9th Cir. 1994).
VII encompasses even those employees who mistakenly believe an employer’s practice is discriminatory.\textsuperscript{224} Title VII itself does not afford such protection, and a rule that protects employees who report practices that are actually lawful seems contrary to the plain language of the statute.\textsuperscript{225} Yet anti-retaliation provision coverage extends to such employees.\textsuperscript{226}

There is no reason that the same broad protection should not apply to employees who oppose MAAs. In an MAA opposition claim, the employee may have a reasonable belief that the MAA is unlawful under Title VII. For example, an employee, such as Mary in the second scenario, could contend that use of the MAA has a disparate impact on racial minorities. If members of this Title VII protected class are more likely to refuse to sign the MAA, these minority employees might be fired. Meanwhile, white employees who may not be as concerned about Title VII protections may sign the MAA and remain employed. Thus, the MAA may disparately impact minority employment.\textsuperscript{227} This disparate impact is analogous to \textit{Griggs v. Duke Power,}\textsuperscript{228} where the Supreme Court found that a standardized test administered to all employees as a condition of employment operated to exclude blacks and was not related to job performance.\textsuperscript{229}

An employee would not have to be correct about this legal theory to be protected from retaliation.\textsuperscript{230} He or she would simply need a reasonable belief that the MAA constitutes an unlawful employment practice under Title VII.\textsuperscript{231} In \textit{Moyo}, the Ninth Circuit used a broad standard for “reasonable,” basing it on the limited factual and legal knowledge available to most employees.\textsuperscript{232} This means that an employee would not need statistical data or a sophisticated understanding of the disparate impact theory to sustain a claim of retaliation, even though he or she would need such data to win a claim of discrimination. Thus, if an

\begin{flushleft}
\textsuperscript{224} See supra notes 110–121 and accompanying text.
\textsuperscript{225} 42 U.S.C. 2000e-3(a) (2000). The language of the statute refers only to practices made unlawful by Title VII.
\textsuperscript{226} See supra Part II.B.3.a.
\textsuperscript{228} 401 U.S. 424 (1971).
\textsuperscript{229} Id. at 431.
\textsuperscript{230} See supra notes 110–121 and accompanying text.
\textsuperscript{231} See supra notes 110–121 and accompanying text. Although it is not discussed in detail, some courts have also mentioned that the employee must have a “good faith” belief. See, e.g., \textit{Moyo v. Gomez}, 40 F.3d 982, 984 (9th Cir. 1994).
\textsuperscript{232} \textit{Moyo}, 40 F.3d at 985.
\end{flushleft}
employee observed a number of black employees quit rather than sign an MAA while white employees signed the MAA, this could be the basis for the reasonable belief that the MAA is unlawful and could sustain a retaliation claim.233

2. Title VII’s Anti-Retaliation Provision Should Protect an Employee Opposing an MAA Because It Violates the Strong Public Policies Behind Title VII

In Title VII retaliation claims, courts should interpret “protected activity” to include not only opposition to intentionally discriminatory employment practices or those with disparate impact, but also opposition to MAAs that explicitly or implicitly shelter employers from Title VII rights or remedies and are unlawful under the policies and purposes of Title VII. The use of MAAs that restrict substantive rights and remedies under any federal statute is unlawful according to the rule in Mitsubishi Motors Corp. v. Soler Chrysler-Plymouth, Inc.234 that arbitration constitutes only a change of forum.235 Gilmer v. Interstate/Johnson Lane Corp.236 and Circuit City Stores, Inc. v. Adams237 recognized that only arbitration systems that fully protect an employee’s statutory rights and remedies are acceptable replacements for courts.238 Thus, if an MAA allows an employee the same rights and remedies he or she would have in court, the employer may legally require the employee to waive the right to file a claim in federal court.239

However, courts routinely invalidate MAAs explicitly or implicitly sheltering employers from compliance with Title VII. Such sheltering includes capping damages at an artificially low amount, preventing employees from filing claims with the EEOC, establishing patently unfair arbitration systems, or requiring employees to pay burdensome

233. Also, because the employee does not have to be a member of a protected class to be safe from retaliation, there is no requirement that the plaintiff show that the MAA restricts his or her own rights and remedies under Title VII. See supra notes 105–109 and accompanying text. Therefore, the employee need only demonstrate a reasonable belief that use of the MAA is an unlawful employment practice as applied to his or her co-workers who may otherwise be protected by Title VII.

235. Id. at 628.
238. Id. at 1313; Gilmer, 500 U.S. at 26.
arbitrator’s fees. Courts have used both contract theory and the policies behind Title VII to invalidate terms that cause employees to “forgo the substantive rights afforded by the statute.”

The FAA requires courts to enforce MAAs unless they are invalid under traditional contract law. Therefore, federal courts have used various contract theories to invalidate MAAs that hinder Title VII. For example, the Supreme Court in Alexander v. Gardner-Denver Co. construed the MAA narrowly to exclude Title VII claims from its scope, while the Cole v. Burns International Security Services court read the MAA to require the employer to pay all arbitrator’s fees to avoid perpetuating an implicit Title VII shelter. In Hooters of America, Inc., v. Phillips, the court found the MAA itself to be lawful, but refused to compel arbitration because the employer had breached the MAA by promulgating patently unfair arbitration rules. The Derrickson v. Circuit City Stores, Inc. and Paladino v. Avnet Technologies courts refused to compel arbitration because the MAAs in both cases contained damage restrictions that violated Title VII’s public policies. Although the legal means employed by courts have varied, the end has always been to invalidate MAAs that explicitly limit employees’ rights and remedies under Title VII.

Concurrent with their invalidation of unlawful MAAs, courts have also construed “protected activity” beyond Title VII’s plain language in order to effectuate Title VII’s policies and purposes. The Fifth Circuit observed that in order for Title VII’s anti-retaliation provision to be effective, “appropriate informal opposition to perceived discrimination

240. See supra Part III.B.
241. Id. at 26.
244. Id. at 42.
245. 105 F.3d 1465 (D.C. Cir. 1997).
246. Id. at 1485.
247. 173 F.3d 933 (4th Cir. 1999).
248. Hooters, 173 F.3d at 940.
250. 134 F.3d 1054 (11th Cir. 1998).
251. Derrickson, 81 Fair Empl. Prac. Cas. (BNA) at 1538; Paladino, 134 F.3d at 1059–60.
252. See supra note 90 and accompanying text.
Mandatory Arbitration Agreements

must not be chilled by the fear of retaliatory action.” To avoid this chilling of opposition, the Ninth Circuit in *EEOC v. Crown Zellerbach Corp.* extended the meaning of “protected activity” to include employee complaints about discrimination to customers of the employer. In *Maynard v. City of San Jose*, the court held that “protected activity” included a white employee’s assistance of a black co-worker because to deny the white employee’s retaliation claim would perpetuate the harmful effects of discriminatory behavior in contravention of Title VII’s purpose. Finally, in *Moyo*, the court used the “reasonable belief” rule to extend “protected activity” to include opposition to practices that are lawful under Title VII. The lesson of these cases is clear: courts should broadly construe “protected activity” beyond the plain language of the anti-retaliation provision when doing so will effectuate the purposes of Title VII.

Therefore, an employee who opposes an MAA that restricts Title VII rights and remedies does so with the Court’s support. Because courts have also sought to effectuate the purposes and prevent the circumvention of Title VII by extending anti-retaliation protection beyond the statute’s plain language, the provision should protect an employee fired for opposing an MAA made unlawful by the policies behind Title VII. Furthermore, because courts have invalidated explicit and implicit shelters alike, this protection should be extended to employees who oppose MAAs that contain implicitly unlawful fee-sharing provisions, such as the MAA in scenario four, as well as those who oppose more explicit violations such as the damage restrictions in scenario three. If the MAA’s end result is a potentially unlawful Title VII shelter, an employee should have the right to oppose the MAA.

This argument should be particularly cogent in the Ninth Circuit and the eight other federal circuits where the “reasonable belief” rule applies. Employees are not expected to be familiar with all the factual and legal issues in a potential case. The average employee may not have the legal sophistication required to discern between employment

254. 720 F.2d 1008 (9th Cir 1983).
255. Id. at 1012.
256. 37 F.3d 1396 (9th Cir. 1994).
257. Id. at 1403.
258. Moyo v. Gomez, 40 F.3d 982, 985 (9th Cir. 1994).
259. See supra notes 110–121 and accompanying text.
practices made unlawful by Title VII’s plain language and those that contravene Title VII’s policies and purposes. The employee’s reasonable belief should prevent outright dismissal of the claim. For example, if scenario four occurred in the Ninth Circuit and Mary filed a retaliation claim, the Moyo rule would apply and the burden would be whether or not she had a reasonable belief that the fee-splitting agreement contradicts the policies of Title VII. Whether a practice that contradicts Title VII’s public policies is a “practice made unlawful” according to the anti-retaliation provision’s plain language would then be subjected to the court’s analysis of “reasonable” in Moyo and subsequent cases. The court would have to determine whether an average employee would have believed that the MAA restricted his or her ability to bring a claim. While Mary may ultimately be mistaken about the provision’s illegality, she can still sustain a retaliation claim with a reasonable belief. Accordingly, an employee’s reasonable belief that an employment practice contradicts the policies of Title VII should trigger the protections of Title VII’s anti-retaliation provision regardless of whether Title VII’s plain language or policies are the source of the illegality, or whether the provision would be enforceable if signed.

3. Protecting Employees Who Oppose MAAs That Contradict Title VII
Will Prevent Employers from Gaining Leverage over Employees’ Title VII Rights

Courts should construe Title VII’s anti-retaliation provision to protect employees who oppose MAAs that contradict the policies of Title VII because the current legal alternatives leave many employees with few legal options with regard to their Title VII rights. Without the provision’s protection, an employee who believes an MAA subverts Title VII has two options: refusing to sign and finding another job or signing and then challenging the MAA if a discrimination claim arises. However, the choice of whether to sign may actually mean choosing between employment and rights, while signing and challenging may place a series of obstacles between an employee and his or her Title VII rights.

While the employment at-will doctrine legitimately compels an employee’s choice between resigning and continuing employment under changed conditions, a closer look reveals that for a class of cases the choice may really be between continued employment and the employee’s

261. See supra notes 110–121 and accompanying text.
Title VII rights. A cause of action for wrongful termination would only apply to current employees. Employers might present MAAs to employees who have complained of discrimination or who might be at a higher risk of filing a Title VII complaint. Thus, requiring an employee to choose between keeping his or her job by signing an MAA that restricts access to Title VII’s remedial scheme and quitting with his or her Title VII rights intact produces the same result: a diminishment of the employee’s Title VII rights. Allowing a cause of action for wrongful termination would prevent employers from using the presumptions of employment at-will to restrict their employees’ Title VII rights.

Alternatively, if the employee signs the MAA in order to continue working, the employer might gain unfair leverage with respect to the employee’s Title VII rights. First, the employee may be deterred from filing a Title VII claim at all. Employees might believe that once they sign the agreement, it is binding regardless of the circumstances under which it was signed. If an employee thinks that the employer’s arbitration scheme is prohibitively expensive or unfair, he or she may decide that filing a claim is unfeasible or futile. This is what the Cole court identified as an MAA with a deterrent effect on the filing of Title VII claims. Also, having to litigate the validity of the arbitration clause in addition to the discrimination claim might increase the time and expense parties must invest in a Title VII case. Such additional costs may ultimately favor employers, who may be better able to absorb additional litigation costs than employees.

Recognizing a cause of action for wrongful termination when a current employee refuses to sign an MAA that contradicts Title VII’s policies could prevent employers from taking advantage of this Catch-22 with regard to an employee’s Title VII rights. Ultimately, Title VII was designed as a limitation on employment at-will that protects employee rights. Employers should not be able to use the presumption of

---

262. See Hoffman, supra note 4, at 153.
263. A few courts have identified and condemned “preemptive retaliation,” which is the practice of firing an employee who the employer fears may make a discrimination complaint in the future. See, e.g., Sauers v. Salt Lake County, 1 F.3d 1122, 1128 (10th Cir. 1993).
264. Basic principles of contract law dictate that parties who sign contracts are presumed to have read and understood the terms. See, e.g., Golenia v. Bob Baker Toyota, 915 F. Supp. 201, 204 (S.D. Cal. 1996). Furthermore, the argument that requiring an employee to sign an MAA to keep his or her job makes the MAAs an illegal contract of adhesion has been rejected by some courts. See id.; Lang v. Burlington Northern R.R., 835 F. Supp. 1104, 1106 (D. Minn. 1993).
266. See supra notes 49–51 and accompanying text.
employment at-will to limit those rights, either by putting employees to a choice between continued employment and their Title VII rights, or by using an MAA as a disincentive on the exercise of those rights. The threat of suit for wrongful termination is one way to restrain the very imbalance in the employment relationship that Title VII seeks to eliminate.

C. Including MAA Opposition in the Scope of “Protected Activity”

Discourages Employers from Drafting MAAs Unfairly in Their Own Favor

There is great temptation for employers to lessen their potential Title VII liability by drafting MAAs to their own advantage. Courts’ recognition of Title VII retaliation claims premised on MAA opposition would reduce an employer’s ability both to draft MAAs thwarting Title VII’s policies and to fire employees who question the practice. In turn, the adverse effects these agreements have on the enforcement of Title VII could diminish. Retaliation claims could motivate those employers who willfully or negligently draft unlawful agreements to redraft them. Recognizing this cause of action should ultimately decrease litigation by producing binding arbitration agreements that fully protect Title VII rights and remedies.

D. An Employer May Defend Against an Employee’s Retaliation Claim by Showing that No Causal Nexus Exists Between the Employee’s Opposition to the MAA and His or Her Termination

To defend against a claim of retaliation, the employer may offer evidence that breaks the causal link between the employee’s protected activity and the dismissal. Were an employee to offer a prima facie case that he or she was dismissed for expressing opposition to a potentially unlawful MAA, the employer could establish legitimate reasons for the dismissal and defeat the claim. For example, it would not matter if the employee engaged in “protected activity” if the employer fired him or her for habitual tardiness. Also, if the employee opposed the agreement

267. Maltby, supra note 12, at 5.
268. See, e.g., McDonnell Douglas Corp. v. Green, 411 U.S. 792, 802 (1973); Folkerson v. Circus Circus Enters., 107 F.3d 754, 755 (9th Cir. 1997); EEOC v. Avery Dennison Corp., 104 F.3d 858, 860 (6th Cir. 1997).
269. See, e.g., Hochstadt v. Worcester Found. for Experimental Biology, 545 F.2d 222, 234 (1st Cir. 1976).

988
in an excessively disruptive or illegal way, and that behavior was the true reason for dismissal, then the retaliation claim should also fail. In these situations, the employer would not be firing the employee for opposing an unlawful practice, but for legitimate, lawful reasons that would justify any employer’s dismissal of any employee.

V. CONCLUSION

With increasing numbers of employers turning to mandatory arbitration for conflict resolution, the number of opportunities for courts to adjudicate employment disputes is decreasing. Therefore, pursuant to Supreme Court mandate, courts must take every precaution to ensure that arbitration agreements provide the same Title VII rights and remedies that employees would have in a courtroom. Congress created Title VII’s anti-retaliation provision to ensure that employers would not find ways to circumvent Title VII’s mandate. Consistent with Congressional policy, courts have construed employers’ retaliation liability more broadly than employers’ discrimination liability. Accordingly, if an MAA undermines Title VII’s discrimination protections, an employee should be able to reject that MAA. If he is discharged, he should have a viable claim under Title VII’s anti-retaliation provision.

HARRY MAYBURY CROSS: IN MEMORIAM

William B. Stoebuck

Just because you were from Ritzville, where Harry Cross was born on August 23, 1913, it would not mean you were “ritzy.” The town is a plain one, named, for a plain man, Philip Ritz, in the plains country of Eastern Washington, the county seat of Adams County. There Harry Cross grew up, graduating from Ritzville High School, valedictorian of his class. His parents were substantial citizens of the town, his father the owner of a title insurance company and a sometime member of the Washington State Legislature. All three of the Cross brothers must have been influenced by their father’s occupation; they probably helped in the business in their spare time. Brother Edward was to become a lawyer in Ritzville, and brother Robert was a future insurance agency owner in Eugene, Oregon. Harry, of course, had an abiding interest in titles and title insurance, working part-time during law school in a Seattle title insurance company and later, as a legal scholar, having a special interest in those areas of law.

From Ritzville High, Harry Cross continued his education at Washington State University. (Yes, the future keeper of Husky mascots was formerly a Cougar!) He graduated in 1936, having been editor of the student newspaper. Though he had sufficient credits to take his degree in journalism and perhaps another subject, he chose to take his B.A. in the “dismal science” of economics. Meantime, he had fallen in love with his wife for life, Mylinn (Linn) Gould, a Montanan, who had chosen to leave that state to operate her beauty salon in Ritzville. They were married on Christmas day of Harry’s senior year in college. Upon graduation, now with a wife to support, he took a job as a newspaper reporter on the Yakima Morning Herald, where he learned to type reporter fashion with two fingers (and never did add more fingers). But by the fall of 1937, Harry was enrolled in the first year of law school at the University of Washington. The Crosses’ first son, Harry, Jr. (“Pete,” a future lawyer, graduate of New York University law school, now believed to be the only patent lawyer in Montana!) had been born by then. So, Harry had to work part-time—as a title officer at Washington Title Company, of course—to support his family. Nevertheless, he excelled as a student, graduating near the top of his class of 1940, one of the Washington Law Review editors, and Order of the Coif.

Thinking probably of becoming a law teacher, Harry obtained a prestigious Sterling Fellowship to do graduate work at Yale Law School, where he spent the 1940–1941 school year. World War II intervened. Harry
entered legal work with the United States Government, first as a lawyer with the Treasury Department in Washington, D.C., in 1941–1942, and then with the Tennessee Valley Authority in Chattanooga during 1942–1943. After Professor Eugene C. Luccock died unexpectedly in early 1943, Harry was appointed assistant professor of law at Washington to replace him. Now he and Linn had two little sons, two future lawyers, Harry, Jr., and Bruce (“Puck,” Harvard Law, 1967, now a partner at Perkins Coie in Seattle); not long after, their third son, Kim, was born.

Harry took over Luccock’s classes—Real Property, Personal Property, Conveyancing, and Community Property—which were to remain the areas in which he taught thousands of law students and became a noted scholar, all during his long career at Washington. Particularly in community property, he became a national authority and certainly the final word in Washington. His articles on community property, first published in *Louisiana Law Review*, then later twice revised and updated in the *Washington Law Review*, the last time shortly before his retirement, were the gospel that saved many law students in his Community Property course, not to mention on the bar examination. He published in other law reviews, too, including his oft-cited article, *The Record “Chain of Title” Hypocrisy*, in *Columbia Law Review*. For Washington practicing lawyers, in addition to his community property law review articles, Cross,—together with his colleague Professor John C. Huston and Judge George Shields—wrote the *Community Property Deskbook*, published by the Washington State Bar Association.

To those of us who knew him closely as his students and as faculty colleagues who labored with him in his areas of law, Harry’s outstanding quality was his incisive and excruciatingly precise mind. It was the mind of a man born to be a property law scholar. Example: “It’s not ‘Shelley’s Rule’; it’s ‘the Rule in Shelley’s Case’.” Meatier example: “No, you’re wrong. You can’t say a grantor reserves a life estate and conveys a remainder. A remainder can follow only the conveyance of a preceding particular estate, usually a life estate [describing ancient principles of English common law]. It’s an executory interest, not a remainder. [His colleague knew that, of course.] Well, I don’t care if courts say ‘reserve a life estate and convey a remainder’; they’re wrong.” And adding as the final clincher: “I’ve taught you everything I know, and you still don’t know anything!” (These examples are not made up; they actually happened.) Harry had a mind that worked both vertically and horizontally: vertically, in that he traced specific principles of law back, back to their origins in larger principles; horizontally, in that he had an incredible capacity to see the interrelationships between specific principles that were on the same “level” with each other. He taught that way in class. He taught by the Socratic method, which in its pure form,
Dedication to Harry Cross

like Plato’s Dialogues of Socrates, uses only questions, to lead the student to discover solutions on his or her own. But he used it only so far, eventually summing up with a brief lecture, which, following the class discussion that preceded it, was at a high level of abstraction, vertical and horizontal. And he was inclined to speak cryptically, with an economy of words, at that point. To the serious student, this was exhilarating learning from a master teacher, but the more lackadaisical student, who wanted to be “spoon fed,” probably did not regard Harry as the teacher from whom he learned the most in law school.

Lest the reader think this is an embellished encomium, it must be said that Harry did have one Achilles heel. It was probably connected with the level at which he thought of legal questions: his conceptualizations were difficult to express in words. Surprisingly, despite his earlier experience in journalism, he was hardly the master of elegant English. The content of his law review articles was authoritative and profound, but the prose did not flow liquid and luminescent from his pen. Editors pored long over his often opaque constructions, to remove the dross without destroying the gold. Once he took a sabbatical leave, to complete the dissertation for his J.S.D., for which he had spent his year in residence at Yale back in 1940 and 1941. Yale wanted to help him complete it—their tribute to his standing as a scholar, actually to facilitate his obtaining the degree long after the normal deadline—and suggested that three connected articles on community property would suffice. But he never got around to doing it, certainly not from lack of knowledge of the subject, but in all probability because of “stage fright” over putting pen to paper.

Harry Cross was taciturn, laconic; that is obvious from what has just been said. One does not think of such men, but, rather, of men with Churchillian loquacity, as being leaders. But Harry was a leader and is best known to most people because of his success in positions of leadership. In the conduct of faculty business, he played the role of wise counselor, a most influential leader. He was personally popular, too, and not only because for years he and Linn sponsored the annual faculty picnics at their “family compound” in Kirkland. After serving as associate dean in the law school from 1975 to 1978, he was its acting dean in 1978 and 1979. The faculty had full confidence in his quiet, fair leadership, and the administration of the law school ran as smoothly as butter. And, before and after his deanships, from 1963 until his retirement in 1984, Harry was the University of Washington’s athletic representative to the PAC-10 Conference, for a time serving as president of the PAC-10. And we must not forget the dog or, rather, the succession of Husky mascots that he and his family kept at their home in Kirkland. For many years, he was chair of the NCAA Committee on
Infractions, a most sensitive and demanding position. Then he became president of the NCAA, the only president who was elevated to that position without having been previously a member of the NCAA Council. It was that position of national prominence that put his name in *Who’s Who in America*.

When he was 67 years old, Harry had to undergo heart bypass surgery. But he snapped back quickly, despite the annoyance of recurring bouts of infection in the leg from which the surgeons had taken veins for the surgery. He quickly returned to his long-time swimming exercise program, even adding some exercise, because his physicians wanted him to walk the mile back and forth for his leg problems. When he had to retire in 1984, it was not because of health but because of the then-draconian university retirement age of 70. Actually, because of the time of year in which his 70th birthday fell in August 1983, the regulations allowed him to teach the following year, so that he retired at age 71. Even then, the faculty would have liked to find some way around the regulations, saying among themselves, “we have lost a great resource, a treasure.” His friends and former students honored him by funding the law school’s first professorship, the Harry M. Cross Distinguished Visiting Professorship in Law.

Once or twice after retiring, he did come back as an “adjunct professor,” to teach Community Property. But mostly he retired to his Kirkland home, to enjoy life with Linn, his children, increasing numbers of grandchildren—and of course the Husky mascots. His youngest son, Kim, and his family shared the home with the senior Crosses, which was a comfort to them. When he was about 80, he required another heart bypass, but again recovered quite well. Within several years, Linn, who had been plagued with health problems for some time, began to fail and betook increasingly to her bed. After she died in 1999, Harry said, “I miss her terribly.” We saw him, at the groundbreaking for the new William H. Gates Law School building, in the summer of 2001, and he had lost much of his old stamina. But the end came rather unexpectedly. He slipped and fell, completely fracturing a knee cap; surgery was required, despite his age and history of heart problems. The surgery went well, and he began to be on his feet in the hospital. But complications set in. It was all too much, and his old heart gave out the evening of October 10, 2001.

Requiescat in pace, Harry M. Cross, my teacher, mentor, colleague, and friend—but always my teacher.

William B. Stoebuck, ’59
Judson Falknor Professor of Law Emeritus
PUNITIVE DAMAGES, EXPLANATORY VERDICTS, AND THE HARD LOOK

Richard W. Murphy*

Abstract: Juries in most American jurisdictions can inflict punitive damages awards against tortfeasors who have committed especially blameworthy torts. Sometimes their awards are startlingly large—multi-billion dollar awards have become increasingly frequent. Nonetheless, juries are generally under no obligation to explain their use of this vast power—a punitive damages verdict typically takes the form of an unexplained number.

Courts can and should change this practice. Under Federal Rule of Civil Procedure 49(b) and analogous state rules, courts could require juries to return “explanatory verdicts” that set forth the bases for their punitive damages awards. Several advantages would flow from adopting this simple reform. First, taking a cue from the “hard look” doctrine of administrative law, courts could review such verdicts to ensure that juries exercise their punitive discretion reasonably and legally. In some cases, where an explanatory verdict revealed error, a court could correct it efficiently by asking the jury to reconsider in light of supplemental instructions. Second, punitive damages awards are supposed to “send messages”; these messages would be clearer if juries used words as well as numbers to express them. Third, the power of civil juries to inflict punitive damages is controversial. Finding out how real juries in real cases justify their verdicts would shed light on whether they should possess this power.

Both the jury foreman and the judge . . . took the unusual step yesterday of making statements after the verdict [ordering Radovan Karadzic to pay $3.9 billion in punitive damages to victims of rape, torture, and genocide].

. . . Mr. Walters, a retired stage hand, said he and other members of the jury felt that they had a duty to express their outrage over the conduct of Bosnian Serb soldiers under the political leadership of Dr. Karadzic, who did not appear in court. “I hope the world gets the message,” he said. “What happened was reprehensible.”

* Associate Professor of Law, William Mitchell College of Law. My thanks to everyone who discussed the ideas in this Article with me or read drafts of its various incarnations—especially Wayne Logan, Mike Steenson, Neil Hamilton, Mike Paulsen, Eileen Scallen, and Evan Tager. Any mistakes or bad ideas contained herein are, of course, my own.

I. INTRODUCTION

On September 25, 2000, a civil jury in New York ordered Dr. Radovan Karadzic to pay $3.9 billion in punitive damages to victims of rape, torture, and genocide. The size of this award was remarkable but not unique—juries have granted four multi-billion dollar punitive damages awards over the last two years and eight over the last sixteen years. Certainly, this dollar figure sent a strong message that the jury despised Karadzic. The jury, however, wanted to clarify this message with words. With the permission of the judge, the jury foreman made a brief statement on the record explaining that the jury knew that “money cannot compensate” the plaintiffs for their suffering and that awarding damages was “a pitiful way to go about this procedure.” The jury’s attempt to put its verdict into context represented a departure from procedural norms. Verdict forms typically do not ask juries to explain the reasoning behind their punitive damages awards, which thus take the form of unexplained numbers.

Courts should abandon this practice and instead require juries to submit “explanatory” verdicts that describe the grounds for their punitive damages awards. Requiring such explanations would equip courts with information they need to ensure that these awards reflect legal and minimally reasonable deliberations. In addition, explanatory verdicts would help jurors to better express what they think of defendants and their conduct. The law has recognized since 1763 that one of the primary functions of punitive damages is to express jury outrage. Money talks, but sometimes not clearly; courts should encourage juries to speak their minds with words as well as dollar figures. Lastly, granting civil juries a discretionary power to punish is a controversial (and to some, deeply alarming) part of American tort law. Perhaps nothing could shed more

---

3. See infra notes 9–16 and accompanying text.
4. Karadzic Tr., supra note 2, at 29.
5. Cf. Paul Mogin, Why Judges, Not Juries, Should Set Punitive Damages, 65 U. Chi. L. Rev. 179, 212 (1998) (arguing that one reason judges should determine amounts of punitive damages awards is because they, unlike juries, can be required to explain the bases for their decisions).
light on whether juries are competent to exercise this power than asking them to explain their punitive damages verdicts and then collecting and studying their answers.

Commentators debate whether punitive damages awards are “rare” or not. Regardless of one’s view on this matter, it is hard to dispute that every now and then a really big verdict comes along. The multi-billion dollar punitive damages hit parade includes: a $3 billion verdict a California jury awarded in July 2001 against Philip Morris; a $3.42 billion verdict an Alabama jury awarded in December 2000 against ExxonMobil; the $3.9 billion verdict a New York jury awarded in September 2000 against Dr. Karadzic; a $145 billion verdict a Florida jury awarded in July 2000 against the “Big Five” tobacco companies; a $4.8 billion verdict a California jury awarded in August 1999 against General Motors; a $3.4 billion verdict a Louisiana jury awarded in 1997

---

8. A recent study reported that, in state courts of general jurisdiction in the 75 largest counties in the nation, roughly 3% of plaintiffs who won tort trials were awarded punitive damages. Perhaps surprisingly, plaintiffs were more likely to win punitive damages in bench trials (7.9% of plaintiff wins) than in jury trials (2.5% of plaintiff wins). Marika F. X. Litras et al., Civil Justice Survey of State Courts, 1996: Tort Trials and Verdicts in Large Counties, BUREAU OF JUSTICE STATISTICS BULLETIN, at 7, tbl.7 (Aug. 2000), available at http://www.ojp.usdoj.gov/-bjs/pub/pdf/tvvlc96.pdf [hereinafter BOJS BULLETIN]; see also Theodore Eisenberg et al., The Predictability of Punitive Damages, 26 J. LEGAL STUD. 623, 631–34 (1997) (reporting that, during fiscal year 1991–1992, in state courts of general jurisdiction in 45 of the 75 most populous counties in the nation, punitive damages were awarded in roughly 6% of tort cases with plaintiff wins (177 out of 2,849 cases)).

9. Marc Kaufman, Lung Cancer Victim Awarded $3 Billion, WASHINGTON POST, June 7, 2001, at A9 (reporting verdict holding Philip Morris liable on six counts of fraud, negligence, and making a defective product; awarding $3 billion in punitive damages and $5.54 million in compensatory damages to lung cancer victim); California: Lower Tobacco Award, N.Y. TIMES, Aug. 23, 2001, at A15 (reporting plaintiff’s agreement to court’s remittitur of $3 billion award to $100 million and defendant’s intent to appeal this lower amount).


13. See, e.g., GM Damages Cut by Over $3 Billion in Gas Tank Case, N.Y. TIMES, Aug. 27, 1999, at A18 (reporting trial court’s reduction of jury’s $4.8 billion punitive damages award for defective design of gasoline tanks to $1.09 billion and GM’s intention to appeal this remitted amount).
against various transportation companies;\textsuperscript{14} a $5 billion verdict an Alaska jury awarded in 1994 against Exxon for the Exxon Valdez spill;\textsuperscript{15} and a $3 billion verdict a Texas jury awarded in 1985 against Texaco.\textsuperscript{16} These mega-verdicts are extreme outliers.\textsuperscript{17} Still, they illustrate the potential scope of the punitive damages power and highlight the importance of understanding how and why juries choose to use it.

The first advantage of requiring juries to explain their punitive damages awards would be to help courts acquire just such an understanding of jury motivations. Courts face the problem of distinguishing legitimate from illegitimate punitive damages awards and reducing those they deem “excessive.” Under current practice, because punitive damages verdicts typically take the form of unexplained numbers, excessiveness review is necessarily a form of outcome review. Simplifying somewhat, a court must ask itself whether a minimally reasonable jury could have reached the actual jury’s verdict. Recent Supreme Court activity suggests how difficult it has proven to make this process coherent. Since 1989, the Court has heard at least six cases

\textsuperscript{14} More specifically, the jury found five defendants liable for $3.365 billion in punitive damages. All defendants but one, CSX Corp., eventually settled. The jury had found CSX liable for $2.5 billion in damages; the trial court remitted this amount to $850 million, which was affirmed on appeal. \textit{In re New Orleans Train Car Leakage Fire Litig.}, No. 2000-CA-0479, 2001 WL 737680, at *3, *19 (La. Ct. App. June 27, 2001).

\textsuperscript{15} See, e.g., \textit{Sea Hawk Seafoods, Inc. v. Alyeska Pipeline Serv.}, Co., 206 F.3d 900, 903 (9th Cir. 2000), \textit{cert. denied sub nom. Exxon Mobil Corp. v. Baker}, 531 U.S. 919 (2000). Multi-billion dollar verdicts make for lengthy post-trial and appellate proceedings. For instance, six years after the jury granted its $5 billion award and four years after entry of judgment, the Ninth Circuit affirmed the district court’s denial of Exxon’s Rule 60(b)(2) motion to set aside the verdict. Exxon had based its motion on a first juror’s testimony that she had received death threats during deliberations and a second juror’s testimony that a bailiff, referring to the first juror outside her presence, had pulled out his gun and remarked that it might help deliberations if “you put her out of her misery or something.” \textit{Id.} at 904. The Ninth Circuit agreed with the district court that Exxon had failed to show prejudice from the bailiff’s “tasteless joke” and that the other “threats” were figments of the first juror’s imagination. \textit{Id.} at 908, 910–14. The actual merits of the $5 billion award are the subject of a still-pending appeal before the Ninth Circuit.


\textsuperscript{17} See BOJS BULLETIN, \textit{supra} note 8, at 7, tbl.7 (reporting median jury-determined punitive damages award of $27,000 in state courts of general jurisdiction in 75 most populous counties in 1996; maximum jury-determined punitive damages award in sample was $138 million); \textit{Eisenberg, supra} note 8, at 632–34 & tbl.1 (reporting median jury-determined punitive damages award of $50,000 in study of jury verdicts in state courts of general jurisdiction in 45 of 75 most populous counties during fiscal year 1991–1992). Mean punitive damages awards tend to be far higher than medians due to a relatively small number of extremely large verdicts. \textit{See id.} at 634 (reporting a mean jury-determined punitive damages award of $534,000 for same sample).
Punitive Damages and the “Hard Look”

raising significant issues concerning control of punitive damages and, along the way, has created some mild due-process restrictions on them: (a) jury instructions must give jurors some vague idea what punitive damages are for; (b) courts must possess a measure of power to reduce them; and (c) “grossly excessive” punitive damages awards violate due process. 18

Absent from these efforts to control punitive damages has been a clear recognition that a court cannot ensure the propriety of a punitive damages award without knowing something about why a jury selected it—one must step beyond the outcome and review the deliberative process as well. The law limits jury discretion to punish by telling juries what kinds of “facts” they should consider as relevant to their task. For example, they should consider the reprehensibility of the defendant’s tort but not that the defendant happens to be from out-of-state. Having found the “facts” that fit the requisite categories, a jury must transmute them into a dollar figure. This step of spinning facts into dollars—of deciding how much the defendant ought to suffer for its tort—is a matter of policymaking, not fact-finding. In forming its “punishment policy,” a jury could make two prominent kinds of mistakes: (a) it might base its award on consideration of facts the law deems irrelevant to punishment; or (b) its punitive reaction to the facts might be outlandishly harsh (e.g., a billion dollar sanction for minor misconduct). To best determine whether a jury has made either kind of mistake, a court must first identify what facts the jury deemed material to punishment; with this information, the court can discern whether the jury considered the “right kind” of facts and whether these facts can reasonably support the jury’s chosen punishment.

The judicial need for information concerning deliberative processes is not unique to punitive damages review. In the context of administrative law, courts invoke “hard look” review of agency policy choices to ensure that they are based on consideration of “relevant factors” and are not tainted by “clear errors of judgment.” 19 To determine whether a choice

18. See infra notes 106–22 and accompanying text (discussing recent Supreme Court decisions wrestling with punitive damages issues).
19. See, e.g., Motor Vehicle Mfrs. Ass’n v. State Farm Mut. Auto. Ins. Co., 463 U.S. 29, 40–43 (1983) (summarizing standard of review applicable to agency policymaking). Originally, the “hard look” doctrine was understood as a way for courts to make sure that agencies themselves take “hard looks” at the problems confronting them. See, e.g., Greater Boston Tel. Corp. v. FCC, 444 F.2d 841, 851 (D.C. Cir. 1970). As the doctrine evolved, usage shifted, and it is now often understood to require courts to take “hard looks” at agency reasoning to ensure minimal rationality and legality.
passes this test, a reviewing court must know what the agency thought about in the course of making it.\textsuperscript{20} Courts therefore require agencies to explain the bases for their policy choices. If review indicates that an agency erred in its deliberations (it considered an “irrelevant factor,” for instance), then the court generally will remand to the agency for reconsideration in light of corrective instructions.\textsuperscript{21}

A form of this “hard look” approach could transfer neatly and usefully to judicial review of punitive damages. Juries should base their punishment choices upon the kinds of facts the law deems relevant, and these facts should “connect” in some minimally sensible way to the punitive damages verdicts they select. To police whether a punitive damages award satisfies these criteria, a court must obtain from the jury an explanatory verdict that summarizes the reasoning behind its punishment selection—it must learn the jury’s “punishment story.” Ideally, if a reviewing court determines that the jury made a deliberative error that it is capable of correcting (it considered an “irrelevant factor,” for instance), the court should “vacate” the jury’s preliminary verdict and “remand” for reconsideration in light of clarifying instructions. Such an approach would help ensure that we obtain from juries what we purport to want but do not always get: punitive damages awards that reflect legal, reasonable exercises of jury discretion to determine fair punishment.

In addition to rationalizing judicial review, a second advantage of explanatory verdicts would be to help juries to “send their messages” more clearly. No doubt there is nothing like a huge punitive damages verdict to concentrate the mind of a tortfeasor; surely the $5 billion award granted in the \textit{Exxon Valdez} matter got Exxon’s attention.

\textsuperscript{20} See Nat’l Lime Ass’n v. EPA, 627 F.2d 416, 451 n.126 (D.C. Cir. 1980) (describing this evolution). In any event, regardless of how one characterizes which entities do the “hard looking,” this form of review necessarily requires a court to immerse itself in the administrative record underlying an agency’s action to make sure that its decision-making process was legal and minimally rational. See Gary Lawson, \textit{Outcome, Procedure and Process: Agency Duties of Explanation for Legal Conclusions}, 48 RUTGERS L. REV. 313, 324 (1996) (discussing general contours of judicial scrutiny of administrative explanations by hard-look review).

\textsuperscript{21} See, e.g., ALFRED L. AMAN, JR. & WILLIAM T. MAYTON, ADMINISTRATIVE LAW 527–29 (2d ed. 2001) (discussing circumstances under which a court will remand due to agency failure to rely on reasoned decision making).
However, to infer the “message” behind a verdict, one must first infer the jury’s reasons for granting it. Trying to figure out these reasons from an unexplained number creates needless room for distortion—a point demonstrated by cases such as the infamous BMW of North America, Inc. v. Gore, in which an Alabama jury became a poster child for tort reform by inflicting a $4 million punitive damages award against BMW for failing to disclose that it had touched up the paint on an expensive sports sedan sold as new. Actually, as explored below, the jury likely had a sensible message to communicate that an explanatory verdict could have made clear.

A third advantage of explanatory verdicts would be that they could provide useful information for assessing whether juries should possess the power to inflict punitive damages in the first place. In recent years, scholars have injected an increasing amount of empirical research into the debate on this subject. Still largely missing is information concerning how actual juries would justify their awards given the chance. We could obtain this data by asking juries to include it in their verdicts. Over time, studying their answers could help demonstrate whether juries are up to the task of wielding this controversial power (or as competent as anyone else available, such as judges and bureaucrats).

Of course, concluding that it would make good policy sense for juries to “explain” their punitive damages awards raises a procedural question: Can courts require them to do so? At first glance, some might object that requiring such explanations would unduly intrude upon the privacy of deliberations in the jury room. Actually, the practice of asking juries to explain the grounds for their verdicts in limited procedural contexts dates back hundreds of years and was discussed with approval by the Supreme Court in the late nineteenth century. A modern-day authorization for a form of this practice exists in Federal Rule of Civil Procedure 49(b) (and parallel state rules), which permit a court to submit to a jury “written interrogatories upon one or more issues of fact the decision of which is

23. Id. at 564–65.
24. See infra notes 255–61 and accompanying text.
25. See Walker v. N. M. & S. Pac. R.R., 165 U.S. 593, 597 (1897) (quoting with approval, inter alia, Spurr v. Shelburne, 131 Mass. 429, 430 (Mass. 1881) (“It is within the discretion of the presiding justice to put inquiries to the jury as to the grounds upon which they found their verdict, and the answers of the foreman, assented to by his fellows, may be made a part of the record, and will have the effect of special findings of the facts stated by him.”)).
necessary to a verdict.” 26 With a just a touch of (legally permissible) creativity, courts could use this rule to require juries to submit explanatory verdicts that describe the factual findings that motivate their punitive damages awards.

The plan: Part II of this Article provides an overview of the law governing juries as they determine punitive damages awards and courts as they review them. A notable conclusion of this overview is that punitive damages law attempts to channel jury deliberations, but judicial review, notwithstanding rhetoric to the contrary, generally reviews punitive damages outcomes. Part III explores how courts could bridge this gap and obtain the information they need to review deliberative processes by using interrogatories to require juries to explain the grounds for their punitive damages awards. Part IV explores in more detail the benefits of doing so—better judicial review, more articulate verdicts, and the chance to study them systematically. Finally, Part V briefly responds to objections some might have to this proposal.

II. DETERMINING PUNITIVE DAMAGES AWARDS—CONTROLLING PROCESSES AND OUTCOMES

Under the traditional common-law model that has existed for nearly 240 years, juries have possessed vast discretion to determine the amount of punitive damages necessary to inflict proper retribution and deterrence upon defendants who have committed particularly awful torts. 27 In other words, juries have the power to determine how much tort should equal how much money. Judges, in general, should defer to jury-determined awards that do not strike them as outrageous.

This basic model has evolved in different ways in different jurisdictions, so one must be careful when making generalizations concerning the modern law of punitive damages. 28 Variance has

26. A substantial majority of the states have modeled their rules on jury interrogatories on this federal model. See, e.g., ALA. R. CIV. P. 49(c); ALASKA R. CIV. P. 49(c); ARIZ. R. CIV. P. 49(h); ARK. R. CIV. P. 49(a); COLO. R. CIV. P. 49(b).

27. See infra notes 36–49 and accompanying text.

28. For instance, most states, whether by common law or statute codifying the common law, permit juries to award discretionary punitive damages to punish and deter especially egregious torts. In a few states, however, courts have ruled that punitive damages are impermissible for state causes of action absent express legislative authorization. See, e.g., Int’l Harvester Credit Corp. v. Seale, 518 So. 2d 1039, 1041 (La. 1988) (“Under Louisiana law, punitive . . . damages are not allowable unless expressly authorized by statute.”); Crowley v. Global Realty, Inc., 474 A.2d 1056, 1058 (N.H. 1984) (recognizing similar rule for New Hampshire); Dailey v. N. Coast Life Ins. Co., 129 Wash. 2d 572,
increased in recent years as many states have taken aggressive steps to limit or modify this jury power. For instance, state legislatures have imposed caps on awards, increased in recent years as many states have taken aggressive steps to limit or modify this jury power. For instance, state legislatures have imposed caps on awards, raised burdens-of-proof, and passed allocation statutes requiring victorious plaintiffs to split their punitive damages winnings with the state.

Of most interest for present purposes, many states have tried to limit jury discretion by providing more specific guidance for their decision-making processes. This has been accomplished through various legislative measures, such as caps on awards, burdens of proof, and allocation schemes.

575. 919 P.2d 589, 590–91 (1996) (prohibiting "punitive damages without express legislative authorization"). The Nebraska Supreme Court has held that punitive damages for state causes of action violate the Nebraska constitution. See, e.g., State ex rel. Cherry v. Burns, 602 N.W.2d 477, 484 (Neb. 1999).

29. See ALA. CODE § 6-11-21 (1993) (setting default rule that punitive damages may not exceed greater of three times compensatory damages or $500,000); COLO. REV. STAT. ANN. § 13-21-102(1)(a) (West 1997) (providing that punitive damages generally may not exceed actual damages); CONN. GEN. STAT. ANN. § 52-240b (West 1991) (limiting punitive damages in product liability cases to double compensatory damages); FLA. STAT. ANN. § 768.73(1)(a)–(b) (West Supp. 2001) (establishing rebuttable presumption that punitive damages should not exceed greater of treble compensatory damages or $500,000); GA. CODE ANN. § 51-12-5.1(g) (2000) (limiting punitive damages to $250,000 for some categories of tort); KAN. STAT. ANN. § 60-3702(e) (1994) (limiting punitive damages generally to lesser of defendant’s recent gross annual income or $5 million); NEV. REV. STAT. ANN. § 42.005(1) (Michie 1996) (limiting punitive damages for certain categories of torts to $300,000 where compensatory damages are less than $100,000 and to three times compensatory damages where compensatory damages are $100,000 or greater); N.D. CENT. CODE § 32-03.2-11(4) (Supp. 1999) (limiting punitive damages to the greater of $250,000 or double compensatory damages); OKLA. STAT. ANN. tit. 23, § 9.1(B)–(D) (West Supp. 2001) (limiting punitive damages to greater of $100,000 or actual damages for reckless conduct and imposing other limits for malicious conduct); TEX. CIV. PRAC. & REM. CODE ANN. § 41.008(b) (Vernon 1997) (limiting punitive damages for certain categories of tort to double economic damages plus up to $750,000 in noneconomic damages); VA. CODE ANN. § 8.01-38.1 (Michie 2000) (limiting punitive damages to $350,000).

30. See, e.g., ALA. CODE § 6-11-20(a) (requiring clear and convincing evidence); ALASKA STAT. § 09.17.020(b) (Michie 2000) (same); CAL. CIV. CODE § 3294(a) (West 1997) (same); COLO. REV. STAT. ANN. § 13-25-127(2) (requiring proof beyond reasonable doubt); GA. CODE ANN. § 51-12-5.1(b) (requiring clear and convincing evidence); MINN. STAT. ANN. § 549.20.1(a) (West 2000) (same); MONT. CODE ANN. § 27-1-221(5) (1999) (same); UTAH CODE ANN. § 78-18-1(1)(a) (1996) (same).

31. See, e.g., ALASKA STAT. § 09.17.020(j) (2000) (requiring 50% of punitive damages award to be paid to state); GA. CODE ANN. § 51-12-5.1(e)(2) (requiring 75% of punitive damages award, less proportionate part of litigation costs, to be paid to state); 735 ILL. COMP. STAT. ANN. 5/2-1207 (West Supp. 2000) (granting court discretion to allocate punitive damages between plaintiff and state); IOWA CODE ANN. § 668A.1(2)(b) (West 1998) (requiring 75% of punitive damages award net of costs and fees to be paid to state); OR. REV. STAT. § 18.540(b) (1999) (requiring 60% of punitive damages award to be paid to state); UTAH CODE ANN. § 78-18-1(3) (1996) (requiring 50% of punitive damages award in excess of $20,000, net of costs and fees, to be paid to state).
deliberations.32 Jury instructions in many jurisdictions have been transformed from general admonitions to consider the need for retribution and deterrence into commands to base punishment on consideration of long lists of factors such as defendant reprehensibility, defendant “financial position,” and the ratio of compensatory to punitive damages.33

Judicial review of punitive damages awards has long used rhetoric that suggests that judges should examine jury deliberative processes for error—for example, they should look for evidence that jurors have been swayed by impermissible “passion and prejudice.”34 Actually, courts cannot review deliberations because they do not make a practice of asking juries for the information necessary to do so. More generally, although the law attempts to channel jury punishment discretion, courts have not taken the steps necessary to determine whether juries stay within those channels.

Instead, judicial review under the current regime generally boils down to checking whether awards are too big, or “excessive,” to use the legalese. The standards for excessiveness review have, like jury instructions, tended to become more complex over recent years—many courts now ostensibly determine whether an award is too big in light of long factor lists. The Supreme Court has recently constitutionalized this fray, holding that “grossly excessive” awards violate due process and providing vague guideposts for gauging when they do.35


The English common law first expressly recognized that civil juries have discretion to award plaintiffs extra-compensatory damages for the purpose of punishing defendants in the 1763 companion cases Wilkes v. Wood36 and the aptly named Huckle v. Money.37 These two cases created a model for determining and reviewing punitive damages awards that, in its essentials, remains in place to this day in many jurisdictions. Wilkes discusses when civil juries may award punitive damages and for what

32. See infra note 56 and accompanying text.  
33. See supra note 56 and accompanying text.  
34. See infra notes 76–79 and accompanying text.  
35. See infra notes 109–16 and accompanying text.  
reasons; *Huckle* addresses how courts are to review and control such awards.

Both cases arose out of the publication of the pamphlet *The North Briton*, No. 45, which executive authorities concluded defamed King George II. Lord Halifax, the Secretary of State, issued a general warrant authorizing a search of the house and papers of John Wilkes, a member of Parliament and the author and publisher of the pamphlet. Pursuant to the warrant, the defendant Wood and several other officials searched Wilkes’ house and seized his papers. 38

Wilkes sued Wood for trespass, seeking £5000 in damages. His counsel argued that the jury should award this princely sum of “large and exemplary damages” because “trifling damages would put no stop at all” to the government’s terrible misconduct. 39 For the defense, the Solicitor-General argued that Wilkes was only entitled to what we would now call compensatory damages, that the £5000 figure amounted to “tenfold damages,” and that, if there were any punishing (as opposed to compensating) to be done, the Crown should handle it through criminal prosecution. 40 In other words, the government should police itself. 41

The judge, Lord Chief Justice Pratt, thought little of this argument and denounced the general warrant before the jury, calling it “a point of the greatest consequence he had ever met with in his whole practice” and opining that “[i]f such a power is truly invested in a Secretary of State, and he can delegate this power, it certainly may affect the person and property of every man in this kingdom, and is totally subversive of the liberty of the subject.” 42 Turning to the power of the jury to bring the state to heel, he continued that

>a jury have it in their power to give damages for more than the injury received.

Damages are designed not only as a satisfaction to the injured person, but likewise as a punishment to the guilty, to deter, from any such proceeding for the future, and as proof of the detestation of the jury to the action itself. 43

---

39. *Id.* at 490.
40. *Id.* at 493.
41. Hmmm.
43. *Id.* at 498–99 (emphasis added).
Thus, the judge invited the jury to use punitive damages for retributive punishment, deterrence, and, in the parlance of our day, “to send a message” to the defendant. The jury accepted this invitation to the extent of awarding £1000 in damages.44

Inevitably, recognizing that juries possess discretionary power to inflict extra-compensatory damages raised the question of how courts should review and control the exercise of this power—an issue addressed by Huckle. As part of the North Briton crackdown, authorities arrested Huckle, a journeyman printer, and held him for six hours, feeding him “beef-steaks and beer.”45 This might sound like a nice evening out, but Huckle sued his captors for trespass, assault, and false imprisonment.46 The jury awarded him £300; the defense moved for a new trial on the ground that these damages were excessive.47

The Lord Chief Justice deferred to the jury’s judgment and denied the defendant’s motion. He admitted that Huckle could not have suffered more than £20 in compensatory damages.48 He nonetheless held “that it is very dangerous for the Judges to intermeddle in damages for torts; it must be a glaring case indeed of outrageous damages in tort, and which all mankind at first blush must think so, to induce a Court to grant a new trial for excessive damages.”49

Thus, combining Wilkes and Huckle provides the following principles: Juries have vast discretion to inflict punitive damages to several ends—retributive punishment, deterrence, and to send a message of jury “detestation.” They control the formula of how much malicious tort should equal how many dollars. Judges should show great deference to jury judgments on punitive damages but have discretion to reduce those they deem outrageous.

44. Id. at 499.
46. Sometimes it’s a matter of principle.
48. Id.
49. Id. at 769.
Punitive Damages and the “Hard Look”

B. The Wilkes Problem Today—Jury Determinations of Punitive Damages

Lengthy treatises discuss the evolving law of punitive damages in this country and how its particulars vary from jurisdiction to jurisdiction. Even so, the following essentials generally hold true. Before awarding punitive damages, a jury must make two preliminary determinations. First, because there is no such thing as a freestanding punitive damages claim, the jury must conclude that the plaintiff has proven all of the elements of an underlying tort. Second, the jury must determine that the defendant committed the tort with a sufficiently malicious intent to justify punishment. For example, an Alabama statute that is typical in this regard requires a showing that the defendant “consciously or deliberately engaged in oppression, fraud, wantonness, or malice with regard to the plaintiff.” Obviously, such abstract formulations of “badness” leave juries with a great deal of practical power to determine which defendants deserve punitive liability and which do not.

Once a jury has determined that a defendant’s conduct warrants punitive liability, the jury must then determine how big an award to inflict. Jury instructions for this purpose take different forms but ultimately boil down to instructing the jury that if it decides to grant punitive damages, jurors should determine the amount by considering the

---

50. See, e.g., GERALD W. BOSTON, PUNITIVE DAMAGES IN TORT LAW (1993); LINDA L. SCHLUETER & KENNETH R. REDDEN, PUNITIVE DAMAGES (3d ed. 1995).
51. See BOSTON, supra note 50, Ch. 31 (1993) (discussing general rule that “actual harm” or “actual damage” is a prerequisite for punitive damages because it is generally a prerequisite to underlying torts).
52. ALA. CODE § 6-11-20(a) (1993); see also e.g., MINN. STAT. ANN. § 549.20.1(a) (West 2000) (providing that defendant must show “deliberate disregard for the rights or safety of others”); RESTATEMENT (SECOND) OF TORTS § 908(2) (1979) (stating that punitive damages may be awarded “because of the defendant’s evil motive or his reckless indifference to the rights of others”). The most refreshingly direct statement of this standard came from the Supreme Court of West Virginia, which opined that punitive liability requires a finding that a defendant has committed an act that is either “really mean” or “really stupid.” TXO Prod. Corp. v. Alliance Res. Corp., 419 S.E.2d 870, 887–88 (W. Va. 1992), aff’d, 509 U.S. 443 (1993) (plurality).
53. Dan B. Dobbs, Ending Punishment in “Punitive” Damages: Deterrence-Measured Remedies, 40 ALA. L. REV. 831, 840 n.29 (1989) (noting that, with regard to the abstract standard for punitive liability, “[w]e probably cannot go beyond saying the conduct must be seriously wrong . . . and that it must be accompanied by a bad state of mind”).
defendant and its conduct in light of the purposes of punitive damages—retribution and deterrence.\(^{54}\)

Some jurisdictions offer virtually no further guidance. For instance, a New York pattern instruction advises:

There is no exact rule by which to decide the amount of punitive damages. The amount that you award as punitive damages need not have any particular ratio or relationship to the amount you award to compensate the plaintiff for (his, her) injuries. If you find that the defendant’s act was (wanton and reckless, malicious), you may award such amount as in your sound judgment and discretion you find will punish the defendant and discourage the defendant or other (people, companies) from acting in a similar way in the future.\(^{55}\)

Increasingly, however, jury instructions (and the statutes on which they are based) attempt to provide more detailed guidance by listing various factors that might bear on one’s sense of proper retribution and deterrence if one stopped to think of them. Minnesota’s pattern instruction is a good example, providing that:

If you decide to award punitive damages, consider, among other things, the following factors:

1. The seriousness of the hazard to the public that may have been or was caused by defendant’s misconduct
2. The profit defendant made as a result of the misconduct
3. The length of time of the misconduct and if the defendant hid it
4. The amount defendant knew about the hazard and of its danger
5. The attitude and conduct of defendant when the misconduct was discovered
6. The number and level of employees involved in causing or hiding the misconduct

\(^{54}\) Cf. Pac. Mut. Life Ins. Co. v. Haslip, 499 U.S. 1, 19–20 (1991) (holding that instruction that vaguely advised jury that punitive damages serve punishment and deterrence functions satisfied due process); see also infra note 107 (setting forth parts of Haslip instruction).

Punitive Damages and the “Hard Look”

7. The financial state of defendant

8. The total effect of other punishment likely to be imposed on defendant as a result of the misconduct. This includes compensatory and punitive damage awards to plaintiff and other persons.

9. The severity of any criminal penalty defendant may get.\(^56\)

Of course, regardless of whether a state takes a terse or wordy approach to such factor listing, neither offers any guidance to jurors concerning what to make of these factors. Suppose a jury reflects on the first factor listed in the Minnesota instruction above and determines that the defendant’s misconduct was really “serious.” Beyond the trivial observation that more serious torts should be punished more seriously than less serious torts—so what? Jurors cannot turn to the law for clear instructions on how harshly to punish.

The facts of a case offer little help in this regard either. For retribution, no matter how many factors about a case a state might insist that a jury consider, the facts cannot provide a logically sufficient basis for inferring how harshly the defendant “should” be punished—one cannot get ought from is that easily. One could learn all there is to know about Exxon; Prince William Sound; supertankers; the Exxon Valdez; alcoholism; Captain Hazelwood; his drinking habits; industry standards for staffing supertankers; the effects of crude oil spills on the environment; the number of seabirds, seals, and otters smothered by the spill; the amount in fines Exxon paid to Alaska; the cost of Exxon’s cleanup efforts; etc., and all these facts still would not provide a logically sufficient basis for inferring that a $5 billion sanction would cause Exxon the proper amount of “corporate pain” to right the scales of justice and

provide just retribution for the misconduct culminating in the *Exxon Valdez* oil spill.

The deterrence purpose of punitive damages has the potential to be more fact-driven depending on the model of deterrence one adopts. For instance, a law-and-economics approach would suggest that awards be scaled to ensure that profit-maximizing parties only act where the benefits of doing so exceed the harms.\(^57\) If a party thinks it can avoid paying for all the damages its actions cause, then it may act in ways that cause more harm than benefits. For instance, suppose \(A\) thinks that polluting a river will cause $100 in damages to \(B\) but that there is only a 50\% chance that it will be caught and held liable for compensatory damages in that amount. The expected monetary value of \(A\)'s compensatory liability is only $50 and, if \(A\) is risk-neutral, it will pollute if its expected profits from doing so would exceed that amount. To correct for this enforcement error and ensure efficient deterrence, a jury would have to divide the amount of harm caused by \(A\) by the chance \(A\) perceived of incurring liability. If, for example, the jury determined \(A\) caused $100 in damages and had perceived a 50\% chance of being caught, then it should divide $100 by ½ for a total damages award of $200. If \(A\) knows that juries compute damages in this manner, then, *ex ante*, \(A\) would expect liability of $100 for its action and will only pollute if the benefits it would accrue by doing so exceed that amount.\(^58\)

Punitive damages law does not, however, require juries to adopt this "efficiency approach" to deterrence.\(^59\) An efficiency-eschewing jury would know that, the bigger the award, the greater the deterrence on the defendant and others. Neither the law nor the facts of a case offer juries any clear stopping point on this scale; they do not help a jury determine how strong a deterrence message to send.

Thus, with due regard for the mushiness of words like "fact," "law," and "policy," to determine punitive damages awards, juries must tread


\(^{58}\) See *generally id.* at 954.

Punitive Damages and the “Hard Look”

beyond the law and the facts to create their own punishment policies.\textsuperscript{60} They must first, of course, figure out the who-did-what-to-whom facts of a case. Punitive damages law then instructs them which of these facts they should focus upon as they consider punishment. Connecting these facts with a given punishment dollar figure requires juries to turn to their personal values and emotional reactions to determine how harshly the defendant \textit{should} be punished. This “should” determination requires not fact-finding, but policymaking.

To make any sense at all, granting juries such power \textit{presupposes} that they possess special insights into the nature of fair punishment that ought to be respected.\textsuperscript{61} The argument must run something like this: someone needs to make the connection between the facts of a case and punishment dollars. That somebody should have some sense of how punishment ought to be handled. One place to find such “oughts” is community sentiment, i.e., punishment is fair if a cross-section of the community thinks it is fair. Juries provide that cross-section and therefore are suited to the art of determining fair punishment. This argument may be wrongheaded; there are lots of reasons not to put punishment up to a “community” vote.\textsuperscript{62} However, if juries lack any particular power to serve as a sort of “community conscience” for punishment purposes, then it is difficult to understand why the state should authorize them to determine punitive damages other than as an unfortunate accident of

\textsuperscript{60} See, e.g., Atlas Food Sys. & Servs., Inc. v. Crane Nat’l Vendors, Inc., 99 F.3d 587, 594 (4th Cir. 1996) ("The jury’s determination of the amount of punitive damages . . . is not a factual determination about the degree of injury but is . . . an almost unconstrained judgment or policy choice about the severity of the penalty to be imposed, given the jury’s underlying factual determinations about the defendant’s conduct."); \textit{see also} Cooper Indus., Inc., 121 S. Ct. at 1686 ("Unlike the measure of actual damages suffered, which presents a question of historical or predictive fact, the level of punitive damages is not really a ‘fact’ ‘tried’ by the jury.") (internal citations omitted); Mogin, \textit{supra} note 5, at 214 (noting that determination of amount of punitive damages generally requires “subjective judgment” regarding amount necessary to punish and deter).

\textsuperscript{61} See, e.g., Cass R. Sunstein et al., \textit{Assessing Punitive Damages (with Notes on Cognition and Valuation in Law)}, 107 YALE L.J. 2071, 2105 (1998) (noting that “a conventional understanding of such awards sees the jury as a sample from the community whose function is to provide an estimate of community sentiment”); \textit{see also} Angela P. Harris, \textit{Rereading Punitive Damages: Beyond the Public/Private Distinction}, 40 A.L.A. L. REV. 1079, 1113 (1989) (discussing the communitarian perspective that “the jury, as the repository of shared communal values, is already in possession of the necessary yardstick” for determining punitive damages awards). For an argument that juries are poorly equipped to serve as community consciences for determination of punitive damages awards, see Mogin, \textit{supra} note 5, at 215.

\textsuperscript{62} For instance, members of the community might be ignorant of past punishment practices or might feel animus toward minority groups.
Jury punishment expertise is an *axiom* of the current punitive damages regime.

Recent empirical work suggests that this axiom rests on shaky ground. To determine punitive damages awards, a jury must typically map bad conduct onto an unbounded dollar scale. Research indicates that people just are not good at this sort of thing. For some kinds of bad conduct, communal “outrage” standards exist. For instance, when researchers presented study participants with a series of descriptions of bad acts that led to personal injury and asked them to rank the acts in order of outrageousness, they tended to rank them the same way. Presumably, jurors, when they determine the amount of punitive damages awards, can and do access such communal outrage standards, at least in some contexts. However, when researchers asked the same study participants to determine appropriate financial sanctions to punish the same conduct, their responses were highly variable and unpredictable.

It appears to be the case that, although as human beings living in a given culture we develop rough standards for determining (or at least rank ordering) the outrageousness of some forms of bad conduct, we do not develop shared mental formulae for converting badness into dollars.

Recent research also suggests how extraneous, irrelevant information may unduly affect punitive damage awards. For instance, in the absence of firm badness-to-dollars formulae, juries may be subject to the “anchoring effects” of various irrelevant numbers they hear at trial. In other words, dollar figures that may have no genuine bearing on proper punishment, such as the plaintiff’s demand, may exercise a sort of gravitational pull on a jury’s ultimate punitive damages award.

---

63. “It is revolting to have no better reason for a rule of law than that it was laid down in the time of Henry IV.” Oliver Wendell Holmes, *The Path of the Law*, 10 HARV. L. REV. 457, 469 (1921).

64. See Sunstein et al., *supra* note 61, at 2098–2100.


66. Sunstein et al., *supra* note 61, at 2103, 2106–07. *But see* Eisenberg, *supra* note 8, at 637 (contending that punitive damages awards are predictable because they bear a statistically significant relationship to compensatory damages awards).


Of related interest, a recent study of a large sample of punitive damages awards concludes that they bear a statistically significant relationship to underlying compensatory damages awards and are therefore in some sense rational and predictable, contrary to popular belief. See Eisenberg et al.,
Thus, even if various juries were similarly outraged by similar bad acts, we could not expect these juries to transmute their outrage into dollars in a way that ensures similar punishment for similar defendants. To the degree proportionality of punishment presents a legitimate concern, this lack of a shared formula for transmuting badness into dollars suggests that perhaps it is not such a good idea to charge juries with the task of acting as *ad hoc* administrative bodies to determine and apply punishment policy in one case and then disband.68 Be that as it may, that is what the law of punitive damages requires juries to do.

Summarizing, the law grants juries tremendous discretion to determine the amount of punitive damages awards. One of the more notable patterns in tort reform over recent years is that many jurisdictions have attempted to channel this discretion by offering more detailed instructions to focus jury attention on certain categories of facts bearing on retribution and deterrence. Having focused jury deliberations in this manner, such instructions do not, however, help juries figure out how to

---

68. A number of commentators have suggested that, to help assuage proportionality concerns, judges rather than juries should determine the amount of punitive damages awards or at least play a more substantial role in that process. See, e.g., Atlas Food Sys. & Servs., Inc. v. Crane Nat’l Vendors, Inc., 99 F.3d 587, 594 (4th Cir. 1996) (arguing that judges, because they get more practice punishing people, have a comparative advantage over juries when it comes to determining punitive damages); Mogin, *supra* note 5, at 212 (“Because judges are in a better position to impose a punishment that is in line with the punishments imposed for similar misconduct, determination of the amount of punitive damages by judges would promote the interest in treating like cases alike.”); *cf.* Colleen P. Murphy, *Integrating the Constitutional Authority of Civil and Criminal Juries*, 61 GEO. WASH. L. REV. 723, 739–40 (1993) (stating that, given judicial professional experiences, “the judge presumably effectuates public policy goals with greater consistency than the one-time jury”). Of course, judicially-determined punishments also raise proportionality concerns—there are hanging judges and more merciful types. The perception that scattering criminal sentencing among judges in the federal system caused unfair disparities in punishment was one of the factors that motivated promulgation of the United States Sentencing Guidelines to control judicial sentencing discretion. U.S. Sentencing Guidelines Manual § 1A1.3 (“Congress sought reasonable uniformity in sentencing by narrowing the wide disparity in sentences imposed for similar criminal offenses committed by similar offenders.”).
weigh these factors. Neither the law nor the facts of a case provide a formula for converting jury-thoughts-on-punitive-damages-factors into dollars. Each jury must create its own formula for this purpose as a largely unconstrained policy choice based inevitably on juror values and intuitions concerning fair punishment. Entrusting juries with this power is troubling because, among other reasons: (a) people do not carry around stable formulae in their heads for this task, so different juries may be expected to punish similar misconduct dissimilarly; and (b) in the absence of clear standards to monetize punishment, juries (and perhaps other decision-makers, too, for that matter) may be subject to the “anchoring” effects of irrelevant numbers that they encounter during trial.

C. The Huckle Problem: Judicial Control of Jury Punitive Damages Verdicts

Today, jury punitive damages verdicts are subject to two very different kinds of control. First, a significant minority of states has taken the procrustean approach of capping punitive damages awards.69 Second, there is judicial review for “excessiveness,” which now comes in both common-law and constitutional flavors. On the common-law side, courts have reviewed punitive damages verdicts for excessiveness since 1763, when the Huckle court held that judges could interfere with damages awards that “all mankind at first blush” would find “outrageous.”70 On the constitutional side, the Supreme Court has, over the last decade or so, crafted due-process limitations on punitive damages, most notably holding in the 1996 BMW decision that “grossly excessive” awards are unconstitutional.71

What these approaches have in common is that, at the end of the day, they only permit a reviewing court to determine whether an award is too big for it to accept as permissible punishment—they focus on outcome review. This limitation is in tension with judicial rhetoric that suggests that courts should check for mistakes in jury deliberative processes, e.g., they should determine whether “passion,” “prejudice,” or other

69. See supra note 29 (citing examples).
illegitimate factors have tainted verdicts. Courts cannot put teeth into process review, however, without information concerning the reasoning supporting verdicts.

1. Common-Law Based Review

The traditional procedural device courts use to reduce punitive damages awards they deem too big is remittitur—which grants a defendant’s motion for a new trial on damages unless the plaintiff agrees to accept some lower figure determined by the judge.72 Recently, some legislatures and courts, stressing skepticism of juror abilities to determine appropriate punishment, have declared that judges should aggressively use this power to carefully police punitive damages awards with little deference.73 In keeping with a tradition dating back to Huckle, however, most courts claim to be far more deferential, frequently holding, for instance, that remittitur to cure excessiveness is only proper where a jury’s award “shock[s] the conscience.”74 Appellate courts have reviewed trial court remittitur decisions for abuse of discretion.75

72. See, e.g., Atlas Food Sys. & Servs., 99 F.3d at 593 (defining remittitur).
73. For instance, a Minnesota Court of Appeals has remarked:

A trial court . . . has broad discretion in determining whether to set aside a verdict as being excessive and should not hesitate to do so where it feels the evidence does not justify the amount, even if the verdict was not actuated by passion and prejudice. The “open-ended and volatile nature of punitive damages” requires a reviewing court to exercise close supervision over the award.

Mrozka v. Archdiocese of St. Paul and Minneapolis, 482 N.W.2d 806, 813 (Minn. Ct. App. 1992) (internal citation omitted); see also, e.g., Clifton v. Mass. Bay Transp. Auth., 2000 WL 218397, at *33 (Mass. Super. Ct., Feb. 3, 2000) (“The fact of the matter is that, while deference is owed to every jury verdict, courts have traditionally given far less deference to the award of punitive damages than to the award of compensatory damages for emotional distress.”). For an example of a legislative effort to tighten review, see FLA. STAT. ANN. § 768.74(3) (West 1997) (“It is the intention of the Legislature that awards of damages be subject to close scrutiny by the courts and that all such awards be adequate and not excessive.”).


Against this backdrop of claimed deference, courts have long described review of punitive damages awards as a search for improper “passion and prejudice” or “partiality” on the part of juries.\textsuperscript{76} Such language suggests that judges should check to make sure that juries do not get carried away during their deliberations by facts that should not affect punishment. For instance, a jury should not inflate its award against a large, faceless corporation just because it does not employ anyone locally; nor should a jury inflict extra punishment against a defendant because of her ethnicity. Many courts continue to use such rhetoric to this day. As the Supreme Court of Ohio recently observed:

The amount of punitive damages awarded may be excessive when it is determined to have been the product of \textit{passion and prejudice}. If the punitive damages award is not the result of passion and prejudice, and not the result of legal error, it is generally not within the province of a reviewing court to substitute its view for that of the jury.\textsuperscript{77}

And as the Supreme Court of Mississippi recently opined:

In considering the assertion of an excessive punitive damage award, this Court is not authorized to disturb a jury verdict regarding punitive damages because it seems too low or too high. It will only be altered or amended when it is so excessive that it evinces \textit{passion, bias, and prejudice} on the part of the jury so as to shock the conscience.\textsuperscript{78}

Passion-and-prejudice review raises at least two interesting issues regarding jury emotions and judicial mind-reading. First, it is not clear how, absent \textit{some} kind of passion, a jury is to determine how much retributive punishment a defendant deserves. Second, framing review as a search for “passion and prejudice” or “partiality” suggests that courts should find out what juries actually think about during their deliberations. Courts cannot effectively glean such information from jury verdicts that take the form of unexplained numbers, however. This led

\begin{footnotes}
76. \textit{Honda}, 512 U.S. at 425 (“In the 19th century, both before and after the ratification of the Fourteenth Amendment, many American courts reviewed damages for ‘partiality’ or ‘passion and prejudice.’”).


78. \textit{Paracelsus Health Care Corp. v. Willard}, 754 So. 2d 437, 444 (Miss. 1999) (emphasis added); \textit{see also Barnes}, 11 P.3d at 177 (declaring that punitive awards are to be reviewed for “passion and prejudice” or “gross excessiveness”).
\end{footnotes}
the Supreme Court to observe, sensibly, that courts applying this "standard" simply infer improper jury motive from the size of an award.\textsuperscript{79} In short, though couched in rhetoric that suggests jury deliberations should be reviewed for error, passion-and-prejudice review has really amounted to nothing more than a loose form of outcome review for excessiveness.

In many jurisdictions, judicial standards for determining the excessiveness of punitive damages awards have followed a roughly parallel track to jury instructions—transmogrifying into long lists of factors bearing on retribution and deterrence.\textsuperscript{80} The Supreme Court gave impetus to this movement in its 1991 opinion \textit{Pacific Mutual Life Insurance Co. v. Haslip},\textsuperscript{81} in which the Court held that traditional, common-law procedures for determining punitive damages awards do not amount to a \textit{per se} violation of due process.\textsuperscript{82} In the course of doing so, the Court cited with approval the review factors that Alabama courts have used to determine excessiveness.\textsuperscript{83} These \textit{Haslip} factors have served as an influential model for courts and legislatures across the country and include:

(a) whether there is a \textit{reasonable relationship} between the punitive damages award and the harm likely to result from the defendant’s conduct as well as the harm that actually has occurred; (b) the \textit{degree of reprehensibility} of the defendant’s conduct . . . ; (c) the \textit{profitability} to the defendant of the wrongful conduct and the desirability of removing that profit . . . ; (d) the ‘\textit{financial position}’ of the defendant; (e) all the \textit{costs of litigation}; (f) the \textit{imposition of criminal sanctions} on the defendant for its conduct, these to be taken in mitigation; and (g) the \textit{existence of other civil awards}

\begin{itemize}
\item \textsuperscript{79} \textit{Honda}, 512 U.S. at 425; see also, e.g., Sierra Club Found. v. Graham, 85 Cal. Rptr. 2d 726, 742 (Cal. Ct. App. 1999) ("Appellate review calls for setting aside such an award only when it appears excessive as a matter of law, ‘or where the recovery is so grossly disproportionate as to raise a presumption that it is the result of passion or prejudice.’"); Ingalls v. Paul Revere Life Ins. Group, 561 N.W.2d 273, 285 (N.D. 1997) ("Punitive damages are excessive when the amount of the award is so great that it indicates passion or prejudice on the part of the jury."); Grynberg v. Citation Oil & Gas Corp., 573 N.W.2d 493, 504 (S.D. 1997) ("Unless the verdict is so large as to clearly indicate that it must have been given under the influence of passion or prejudice, it should stand.").
\item \textsuperscript{80} Seeinfra note 84 and accompanying text.
\item \textsuperscript{81} 499 U.S. 1 (1991).
\item \textsuperscript{82} \textit{Id.} at 17.
\item \textsuperscript{83} \textit{Id.} at 15.
\end{itemize}
against the defendant for the same conduct, these also to be taken in mitigation.84

Courts, like juries, however, have little or no clear guidance vis-à-vis how they should weigh such factors; judges must therefore largely base excessiveness review on their own policy intuitions regarding fair punishment.85 Without commenting whether this result is a good or a bad thing, it seems that such review amounts to a judicial gut-check to make sure that awards are not too big.86 That said, this gut-check can have real


85. Justice Kennedy has described the problem this way:
To ask whether a particular award of punitive damages is grossly excessive begs the question: excessive in relation to what? The answer excessive in relation to the conduct of the tortfeasor may be correct, but it is unhelpful, for we are still bereft of any standard by which to compare the punishment to the malefaction that gave rise to it. A reviewing court employing this formulation comes close to relying upon nothing more than its own subjective reaction to a particular punitive damages award in deciding whether the award violates the Constitution. This type of review, far from imposing meaningful, law-like restraints on jury excess, could become as fickle as the process it is designed to superintend.

TXO Prod. Corp. v. Alliance Res. Corp., 509 U.S. 443, 466–67 (1993) (Kennedy, J., concurring). Examination of efforts by courts to develop clearer standards to guide review only dramatizes the force of Justice Kennedy’s point. For example, the Supreme Court has repeatedly stressed the importance of comparing the ratio of punitive to compensatory damages awards in assessing excessiveness. See Haslip, 499 U.S. at 18, 23 (observing that punitive damages award four times the size of compensatory award might be “close to the line” of violating due process, but eschewing “bright-line,” “mathematical” test); see also BMW of N. Am., Inc. v. Gore, 517 U.S. 559, 582 (1996) (declaring that ratio is a “guidepost” for excessiveness review). Lower court cases demonstrate how difficult it has been to turn this admonition into any kind of clear guidance. See, e.g., Cont’l Trend Res., Inc. v. Oxy USA Inc., 101 F.3d 634, 639 (10th Cir. 1996) (suggesting that in economic injury cases where “damages are significant and the injury not hard to detect, the ratio of punitive damages to harm [or potential harm] generally cannot exceed a ten to one ratio”); cf., e.g., Tronzo v. Biomet, Inc., 236 F.3d 1342, 1349–50 (Fed. Cir. 2001) (rejecting claim that 38,000:1 ratio between punitive and compensatory awards presented exceptional circumstances sufficient to justify review of punitive damages award where that issue was foreclosed by mandate of earlier appeal).

86. The potential of anchoring effects to affect punitive damages awards raises interesting questions concerning the psychology of their judicial review. Recall that studies suggest that jurors, faced with the problem of determining damages on an unbounded dollar scale, may be unduly influenced by “anchoring” numbers—i.e., their awards may be drawn toward dollar figures they hear at trial that may be of little or no relevance to fair punishment. See supra note 67 and accompanying text. If judges, too, are subject to this effect, then it would seem to follow that a judge reviewing an illegitimate punitive damages award would find her judgment concerning the “proper” amount of punitive damages subject to the “gravitational pull,” as it were, of the jury’s award. In other words, the higher a jury award, the higher the amount a judge might deem acceptable. More concretely, suppose a jury awards $1 million in punitive damages and a judge, on post-trial review, determines
consequences; a 1987 study suggested that, by the time post-trial and appellate review ends, plaintiffs on average only collect about fifty percent of the punitive damages that juries award.87

2. The Atlas Outlier—The Fourth Circuit Wrestles with Juries as Punishment Policymakers

In light of this Article’s focus on the policymaking role juries play in determining punitive damages awards, no discussion of judicial review would be complete without exploration of the Fourth Circuit’s innovative and relatively recent discussion of that same topic in *Atlas Food Systems & Services Inc. v. Crane National Vendors, Inc.*88 The key to the court’s analysis was to recognize, correctly, that juries engage in policymaking when they determine punishment.89 The court then reasoned that jury policymaking merits less deference than jury fact-finding.90

In the first *Atlas* trial, the plaintiff, who alleged that the defendant had sold it defective vending machines that allowed customers to steal food without paying, won $3 million dollars in punitive damages.91 The trial court found this amount to be excessive and ordered a new trial unless the plaintiff accepted a remitted award of $1 million.92 The plaintiff gambled on a new trial and won $4 million in punitive damages.

that this figure is unacceptably high and reduces the award 50% to $500,000. The existence of anchoring effects suggests that, if the jury had awarded $2 million instead, the judge would not have remitted to the same $500,000 figure. Instead, influenced by the higher initial jury award, she would have remitted to some larger amount; for example, she might have remitted 50% again down to $1 million. The possibility that such effects may exist (as common sense suggests they may) makes it all the more important, of course, that judges understand the grounds underlying jury-determined punitive damages awards so that they can avoid the “gravitational pull” of those tainted by error.

87. See generally *Mark Peterson et al., Rand Institute for Civil Justice, Punitive Damages: Empirical Findings* 27–29 (1987) (estimating that, after post-trial and appellate review and attendant settlement discussions, defendants actually pay approximately 50% of the punitive damages awarded by juries). *But see* Theodore Eisenberg & Martin T. Wells, *The Predictability of Punitive Damages Awards in Published Opinions, the Impact of BMW v. Gore on Punitive Damages Awards, and Forecasting Which Punitive Awards Will Be Reduced*, 7 *Sup. Ct. Econ. Rev.* 59, 64–65, 79 (1999) (reviewing published opinions issued in the year following the Supreme Court’s opinion in *BMW* that discuss the excessiveness of punitive damages awards; finding that courts reduced punitive awards in 13.1% of cases in which they affirmed punitive liability and plaintiffs had received substantial compensatory damages).

88. 99 F.3d 587 (4th Cir. 1996).

89. *Id.* at 594–95.

90. *Id.*

91. *Id.*

92. *Id.*
Undeterred by juries that kept disagreeing with him, the trial judge again ordered remittitur to $1 million. Rather than face a third trial on damages before the same judge, the plaintiff took an interlocutory appeal.93

The Fourth Circuit began its affirmance by observing that federal district courts, when reviewing punitive damages awards granted under the authority of state law, must apply "the state’s substantive law of punitive damages under standards imposed by federal procedural law."94 The award had been granted pursuant to South Carolina law, which provides that punitive awards should be reviewed in light of a typical laundry list of factors (i.e., defendant’s culpability, duration of misconduct, etc).95 The federal procedural law governing the propriety of ordering a new trial is Federal Rule of Civil Procedure 59; the specific question before the Fourth Circuit was whether the district court had abused its discretion by ordering a new trial pursuant to Rule 59.96 The court observed that

Rule 59 standards are well established in the Fourth Circuit: On such a motion it is the duty of the judge to set aside the verdict and grant a new trial, if he is of the opinion that (1) the verdict is against the clear weight of the evidence, or (2) is based upon evidence which is false, or (3) will result in a miscarriage of justice, even though there may be substantial evidence which would prevent the direction of a verdict.97

The court then noted that the first two prongs of this Rule 59 standard address “purely factual questions: whether the jury’s damages award is (1) ‘against the weight of the evidence’ or (2) ‘based upon evidence which is false.’”98 To review such “fact questions,” a court need only compare a jury’s verdict to the factual record created at trial. This approach cannot work for punitive damages awards, however, because the determination of the amount of such an award “is not a factual determination about the degree of injury but is, rather, an almost unconstrained judgment or policy choice about the severity of the penalty to be imposed, given the jury’s underlying factual determinations about

93. Id.
94. Id., at 593 (citing Browning-Ferris Indus. of Vt., Inc. v. Kelco Disposal, Inc., 492 U.S. 257, 279 (1989)).
95. Id. at 593 n.2.
96. Id. at 593.
97. Id. at 594 (emphasis added and citation omitted).
98. Id.
The court reasoned that, “[b]ecause the factual record provides no direct foundation for the amount of punitive damages” independent of jury policymaking, the fact prongs of Rule 59 analysis cannot apply. Instead, “review of the size of the jury’s award best utilizes the third prong of the Rule 59 review standard—whether the jury’s award would result in a miscarriage of justice.”

The Fourth Circuit then advised that jury policymaking does not merit the same extreme deference as jury factfinding. It supported this conclusion with the observation that judges, unlike juries, frequently impose criminal sentences and review punitive damages awards; courts therefore have a “comparative advantage” over juries when it comes to punishment policymaking in particular. It follows that, when invoking the “miscarriage of justice” standard to review a punitive damages award, “the district court has a participatory decisionmaking role that it does not have when reviewing a jury’s findings based solely on facts.”

The upshot: courts should review punitive damages awards “less deferentially than . . . factual findings which may be measured against the factual record.”

The importance of Atlas lies in its frank recognition that the policymaking role juries play when deciding punitive damages awards should have procedural consequences—a proposition this Article also explores, though to different effect. Atlas does little, of course, to clarify for judges how large punitive damages awards should be as a matter of good policy; it does, however, encourage them to overcome any scruples they might have to refrain from interfering with jury punishment policy decisions.

3. The Supreme Court, Due Process, and “Gross Excessiveness”

Since 1989, the Supreme Court has heard at least six cases raising significant punitive damages issues, which together have fashioned a set

---

99. Id. (emphasis added).
100. Id. at 594.
101. Id.
102. See id. at 594–95.
103. Id.
104. Id. at 594.
105. Id. at 595.
of loose due-process limits on awards.\textsuperscript{106} Some restrictions relate to required procedures. For instance, due process requires that trial court instructions give juries at least some idea concerning the purposes that punitive damages are supposed to serve,\textsuperscript{107} it also requires that courts enjoy some measure of real power to reduce those awards they deem excessive.\textsuperscript{108}

Most critically, in its 1996 decision \textit{BMW of North America, Inc. v. Gore},\textsuperscript{109} the Supreme Court held that punitive damages awards that are “grossly excessive” violate due process, and the Court provided “three

\begin{itemize}
\item \textsuperscript{106} Cooper Indus., Inc. v. Leatherman Tool Group, Inc., 532 U.S. 424, ___, 121 S. Ct. 1678, 1687–89 (2001) (holding that appellate courts should review \textit{de novo} district court determinations of whether punitive damages awards are so “grossly excessive” as to violate due process); BMW of N. Am., Inc. v. Gore, 517 U.S. 559, 574–75 (1996) (setting forth three “guideposts” to aid courts in determining whether punitive damages awards are “grossly excessive” and thus violate due process);

\item \textsuperscript{107} See Haslip, 499 U.S. at 19–22. This was the instruction the Court held provided sufficient guidance with regard to punitive damages:

\begin{quote}
This amount of money is awarded to the plaintiff but it is not to compensate the plaintiff for any injury. It is to punish the defendant. Punitive means to punish or it is also called exemplary damages, which means to make an example. So, if you feel or not feel, but if you are reasonably satisfied from the evidence that the plaintiff, whatever plaintiff you are talking about, has had a fraud perpetrated upon them and as a direct result they were injured and in addition to compensatory damages you may in your discretion award punitive damages.

Now, the purpose of awarding punitive or exemplary damages is to allow money recovery to the plaintiffs, it does to the plaintiff, by way of punishment to the defendant and for the added purpose of protecting the public by deterring [sic] the defendant and others from doing such wrong in the future. Imposition of punitive damages is entirely discretionary with the jury, that means you don’t have to award it unless this jury feels that you should do so.

Should you award punitive damages, in fixing the amount, you must take into consideration the character and the degree of the wrong as shown by the evidence and necessity of preventing similar wrong.

\textit{Id.} at 6 n.1 (internal quotation marks omitted).

\item \textsuperscript{108} Honda Motor, 512 U.S. at 420 (holding that Oregon constitutional provision that forbade courts from reducing punitive damages awards if any evidence at all supported punitive liability violated due process).

\item \textsuperscript{109} 517 U.S. 559 (1996).
\end{itemize}
guideposts” for gauging when an award transgresses this standard. The facts of BMW are irresistible: Dr. Gore sued BMW for failing to disclose that it had touched up the paint on an expensive sports sedan sold to him as new. The jury awarded Dr. Gore $4000 in compensatory and $4 million in punitive damages for his trouble, which the Alabama Supreme Court, for reasons that remain mysterious, remitted to $2 million. The Supreme Court ruled that this remitted award violated due process because Alabama had given BMW no notice that it could be walloped with a $2 million punishment for its conduct.

The “three guideposts” for determining “gross excessiveness” require courts to consider: (1) the reprehensibility of the defendant’s conduct; (2) the ratio between a punitive award and the harm the defendant’s conduct caused (or might likely have caused); and (3) the size of any civil or criminal sanctions the relevant legislature has created for punishing similar misconduct. Applying these guideposts to the facts before it, the Court observed that BMW’s nondisclosure was not really all that bad; that there was a 500:1 ratio between the $4000 in compensatory damages that Dr. Gore had won and the remitted punitive damages award of $2 million; and that Alabama’s maximum civil penalty for deceptive trade practices was a mere $2,000. In light of these circumstances, the Court concluded that Alabama had not given BMW reasonable notice that its undisclosed paint-job could lead to a multi-million dollar fine.

Although BMW expresses a hostile mood toward punitive damages, it is questionable whether its guidelines represent much of an advance on traditional, common-law-based excessiveness review. For one thing, the BMW “gross excessiveness” test suffers from the same problem as any other form of excessiveness review—it offers courts no clear formula for

---

110. Id. at 574–75.
111. Id. at 563.
112. Id. at 565–67.
113. Id. at 574–75. This “notice” dimension of BMW makes it seem a mix of substantive and procedural due process concepts. On the one hand, BMW instructs that awards must not be too big, which smacks of substantive due process. On the other, it hints that the problem was that Alabama gave BMW no warning that its conduct could be subject to such a massive penalty, which suggests a procedural due process issue. If the Court were serious about hanging BMW on a procedural hook, then it would seem to follow that the remitted award might have stood if Alabama’s deceptive trade practices act had provided for sanctions on the order of $2 million rather than $2,000.
114. Id.
115. Id. at 575–85.
116. Id. at 585.
weighing its “guideposts” and instead simply tells courts some factors they should think about when deciding how big awards should be.\textsuperscript{117} For another, some states’ highly malleable factors for excessiveness review already largely encompassed the guideposts.\textsuperscript{118} Finally, a recent study of published opinions reviewing punitive damages verdicts both before and after \textit{BMW} suggests that it did not immediately lead lower courts to review awards more aggressively.\textsuperscript{119}

Nonetheless, the Supreme Court has recently issued an opinion, \textit{Cooper Industries, Inc. v. Leatherman Tool Group, Inc.},\textsuperscript{120} which may give \textit{BMW} more bite in the future. In this case, the Court resolved a circuit split regarding whether appellate courts should review district court determinations of “gross excessiveness” under an abuse-of-discretion or \textit{de novo} standard.\textsuperscript{121} The Court chose the stricter \textit{de novo} standard in part on the ground that appellate courts need such control to develop case law that gives meaning to the abstract concept of “gross excessiveness.”\textsuperscript{122}

4. A Summary of Judicial Review

Courts have often used rhetoric suggesting that they review jury deliberations for “passion and prejudice.” In the absence of information from juries concerning their deliberations, however, passion-and-
Punitive Damages and the “Hard Look”

prejudice review must reduce to a form of outcome review for excessiveness. Judicial review of punitive damages awards for excessiveness (in both common-law and constitutional flavors) requires courts to determine whether an award is too big in light of various retribution and deterrence-related factors, e.g., the gravity of the defendant’s misconduct, its profitability, etc. Courts, like juries, have no settled formulae for weighing such factors and instead must largely rely on their own idiosyncratic notions of fairness to guide their review. In keeping with tradition, many courts—at least in their rhetoric—continue to stress deference to jury punitive damages awards. Some courts (sometimes at the bidding of their legislatures) have abandoned this rhetoric of deference in light of their suspicions that juries are not particularly competent to determine punishment.

III. WE HAVE THE PROCEDURAL POWER TO FIND OUT MORE: RULE 49(B), EXPLANATORY VERDICTS, AND ITERATIVE INSTRUCTIONS

The law places very few limits on jury discretion to determine punitive damages awards, but prominent among such loose constraints as do exist are: (a) the law tells juries what factors they should consider as relevant to punishment; and (b) judges can set aside awards they deem so “excessive” as to “shock the conscience.”123 Neither of these constraints can function well in the absence of information concerning the grounds for a jury’s punitive damages award. Without such an explanation, a judge cannot tell whether the jury based its award on consideration of irrelevant factors with no proper bearing on punishment. Moreover, whether a jury’s award is “excessive” should depend on what precisely it sought to punish. A judge reviewing an unexplained award can only guess as to jury motivations. Thus, for these and other reasons that will be discussed in more detail in Part IV, it would be good for reviewing courts to learn juries’ justifications for their punitive damages verdicts.

From this conclusion, two procedural questions follow: First, how much such information could courts obtain? The answer seems to be: a lot more than they generally seek. Courts could and should ask juries to answer open-ended interrogatories that ask them to describe the factual bases for their punitive damages awards. Second, where an

123. The third, procrustean constraint is the arbitrary caps that some states have placed on awards. See supra note 29.
“explanatory” verdict reveals that an award has been tainted by error, what steps might a court take to correct it? At least in those cases where the error does not suggest jury prejudice or incompetence, the court could and should ask the jury to reconsider its award in light of supplemental, corrective instructions. Adopting such procedures would enable courts to better ensure that punitive damages awards are both legal and minimally reasonable.

A. A Creative Approach to Rule 49(B)—Asking Juries for “Explanatory Verdicts”

Our litigation system places a high value on ensuring the privacy of jury deliberations and the finality of verdicts. For these reasons, jurors generally cannot offer admissible testimony to explain the reasoning (or lack of it) underlying their verdicts. In other words, juries enjoy a form of work-product protection for their deliberations. Not all jury thoughts are immune from scrutiny, however—those embodied in a verdict could not be more public, obviously, and courts have a great deal of relatively untapped power to shape the form of verdicts. The upshot: for the most part courts cannot force juries to explain their verdicts after the fact, but they can insist that juries issue explanatory verdicts, which may shed a great deal of light on jury reasoning.

1. The Logic of Asking

Stepping back, an observer from another procedural planet might conclude that one of the odder aspects of our civil litigation system is the way in which courts arm juries with the legal concepts necessary to reach a general verdict. In civil litigation, a general verdict merely requires a jury to declare whether the defendant is liable and, if so, how much the defendant owes in damages. To reach these conclusions, a jury must first determine the material facts of the case and then apply these facts to the law as set forth in the court’s instructions, which are often long and filled with jargon. Courts read these instructions to juries, which then retire to deliberate. A trial judge has discretion with regard to whether to

124. See infra notes 130–31 and accompanying text.

125. See, e.g., Walker v. N. M. & S. Pac. R.R., 165 U.S. 593, 596 (1897) (“[A] general verdict embodies both the law and the facts. The jury, taking the law as given by the court, apply that law to the facts as they find them to be and express their conclusions in the verdict.”).
Punitive Damages and the “Hard Look”

supply jurors with written copies of their instructions—traditionally, however, there has been some thought that this is a bad idea. Of course, a jury might find such legal training confusing (not to mention dispiriting) and ask for clarifying instructions. Courts have sometimes responded to such requests by repeating the instructions that confused the jury in the first place. To state the obvious, the law’s usual approach to teaching juries the law they ostensibly need to learn is—pedagogically speaking—a joke.

If the goal were to provide juries with a correct understanding of the law for them to apply to the facts of cases, it would be easy to imagine better ways to go about it. For instance, judges could sit in on jury deliberations. They would detect legal errors as juries made them and correct them through appropriate curative instructions. In addition, judges in the jury room could learn exactly what facts juries are finding and ensure that they are supported by substantial record evidence.

Less aggressively, rather than observe deliberations, courts might require juries to summarize the grounds for their verdicts. Such explanatory verdicts might similarly expose legal errors and isolate dubious fact-finding while at the same time preserving the privacy of

126. 9A CHARLES ALAN WRIGHT & ARTHUR MILLER, FEDERAL PRACTICE AND PROCEDURE, § 2555, at 435–36 (2d ed. 1995) (observing that “[s]ome believe that providing the jury a copy of the instructions enhances the danger that the jurors may pick passages out of context contrary to the rule that the jury must consider the charge as a whole”).

127. See Roger M. Young, Using Social Science to Assess the Need for Jury Reform in South Carolina, 52 S.C. L. REV. 135, 156–62 (2000) (collecting studies demonstrating that juries have grave problems comprehending their instructions).

128. See, e.g., Weeks v. Angelone, 528 U.S. 225, 229, 234 (2000) (holding that trial court in capital case did not err when it responded to jury’s query concerning its discretion to sentence defendant to life instead of death by referring jury to relevant portion of original instructions).

129. A recent study suggests the enormity of this joke in the context of punitive liability instructions. Researchers asked 726 mock jurors to determine punitive liability for four fact patterns closely based on cases in which appellate courts had ruled as a matter of law that the defendants could not be held liable for punitive damages. Reid Hastie et al., A Study of Juror and Jury Judgments in Civil Cases: Deciding Liability for Punitive Damages, 22 LAW & HUM. BEHAV. 287, 290–91 (1998). Study participants were read jury instructions (which were approximately 500 words long and excerpted from those used in the Exxon Valdez matter) and provided with copies to consult during deliberations. Id. at 291. After the mock juries returned their verdicts, the researchers asked the participants to fill out questionnaires designed to determine how well they had understood their instructions. Id. at 292. The median score on that test was 5% correct; the mean was 9% correct. Id. at 295. The authors of the study suspect that jurors may have had trouble distinguishing recklessness from negligence—a conclusion consistent with the fact that 67% of the mock juries that were able to reach a verdict found punitive liability even though appellate courts had ruled out such liability as a matter of law for the fact patterns at issue. Id. at 293, 303.
deliberations. Where an explanatory verdict brought such errors to light, a judge might, in an appropriate case, help the jury by offering clarifying instructions and recharging it for further deliberations. The process for educating juries could thus become a sort of stylized conversation rather than an exercise in issuing confusing legal commands once and then hoping for the best. No doubt such an approach would change the outcome of at least some cases.

2. Rule 606(b) Does Not Stop Courts from Asking

To some, however, the suggestion that judges require juries to summarize their deliberations in explanatory verdicts may run counter to a long-held habit of thinking of the jury room as something of a black box. Federal Rule of Evidence 606(b) exemplifies this view of the jury room as semi-sacrosanct, providing:

Inquiry into validity of verdict or indictment. Upon an inquiry into the validity of a verdict or indictment, a juror may not testify as to any matter or statement occurring during the course of the jury’s deliberations or to the effect of anything upon that or any other juror’s mind or emotions as influencing the juror to assent to or dissent from the verdict or indictment or concerning the juror’s mental processes in connection therewith, except that a juror may testify on the question whether extraneous prejudicial information was improperly brought to the jury’s attention or whether any outside influence was improperly brought to bear upon any juror . . . .

The thrust of this complicated provision is that, subject to various confusing exceptions, jurors cannot provide competent evidence explaining how they reasoned their way to a verdict.

At first glance, one might think that Rule 606(b) creates an obvious roadblock to courts requiring juries to explain the grounds for their punitive damages awards because seeking such information would amount to forbidden investigation into juror “mental processes.” This rule, however, forbids post-verdict inquiries into jury deliberations as a means of preserving the finality of verdicts—the thought being that, without such a rule, losing lawyers would harass and attempt to corrupt

130. FED. R. EVID. 606(b).
jurors to obtain testimony to undo trial outcomes. Rule 606(b), in short, protects verdicts from jurors and lawyers; it does not speak to what can go into a jury’s verdict in the first place. Requiring a jury to return a verdict of a certain form cannot violate this rule.

3. Rule 49(b) Authorizes Asking: The Mechanics of Explanatory Verdicts

Eliminating any Rule 606(b) objection clears the way for an analytically distinct procedural question: Do courts enjoy an affirmative power to require juries to explain their grounds for awarding punitive damages? The answer to this question seems to be yes—at least in federal court and in those states (a substantial majority) that have based their rules of civil procedure governing verdict forms on the federal model. Underlying any punitive damages award, there must be some sort of “punitive story”—some set of factual findings about the defendant and its conduct that motivated the jury to punish. Generally speaking, there are three different kinds of verdict forms in civil litigation—the general verdict, the special verdict, and the general verdict with interrogatories. Courts could use this last form to require juries to report their “punitive stories.”

As discussed above, in a civil case a general verdict may simply state whether the defendant was liable and the amount of damages owed. To reach a general verdict, a jury must apply the facts of a case to the law. This process leaves ample room for jury legal error.

Federal Rule of Civil Procedure 49 (and parallel state rules) authorizes courts to use either of two alternative verdict forms to diminish this problem. First, Rule 49(a) provides for “special verdicts,” which require a jury to issue a “special written finding upon each issue of fact” in response to a series of questions posed by the judge. The judge then applies the jury’s factual findings to the law to determine the case’s outcome. In theory, special verdicts can remove room for jury legal error; in practice, they tend not to do so because courts often pose

131. See Tanner v. United States, 483 U.S. 107, 119–20 (1987) (discussing policy considerations of verdict finality supporting Rule 606(b) and its common-law precursor); see also Fed. R. Evid. 606(b), advisory committee’s notes (observing that rule promotes “freedom of deliberation, stability and finality of verdicts, and protection of jurors against annoyance and embarrassment”).


questions at a level of abstraction that requires application of legal concepts (e.g., “Did the defendant materially breach the contract?” or “Was the defendant’s conduct negligent?”). Also, special verdicts can be treacherous to use because if a court omits an issue of fact from its verdict form without objection, the parties waive the right to trial by jury on that issue.\textsuperscript{135}

Rule 49(b) provides a second alternative, the “general verdict accompanied by answer to interrogatories,” which, as the name suggests, is an intermediate form designed to avoid both the opacity of general verdicts and the technical difficulties of special verdicts. Pursuant to Rule 49(b), a “court may submit to the jury, together with appropriate forms for a general verdict, written interrogatories upon one or more issues of fact the decision of which is necessary to a verdict.”\textsuperscript{136} Because the jury reaches a general verdict, this hybrid form avoids the technical pitfall of special verdicts that they must cover each material issue of fact on pain of waiver. At the same time, however, the use of select interrogatories permits a court to illuminate the jury’s fact-finding and minimize room for jury legal error.

Federal district courts have nearly blanket discretion to choose which of these verdict forms to use in a given case.\textsuperscript{137} Given the frequency of attacks on the competency of civil juries over recent decades, one might expect courts to use special verdicts or general verdicts with interrogatories frequently to exercise greater control over jury deliberations. Nonetheless, it seems that, in federal court at least, the general verdict is the form of choice.\textsuperscript{138}

\textsuperscript{135} Id.

\textsuperscript{136} FED. R. CIV. P. 49(b).

\textsuperscript{137} See, e.g., Bills v. Aseltine, 52 F.3d 596, 605 (6th Cir. 1995) (“Whether a court uses a special or general verdict rests in its discretion . . . .”), cert. denied, 516 U.S. 865 (1995); Jarrett v. Epperly, 896 F.2d 1013, 1020 (6th Cir. 1990) (stating that the court has discretion to determine form of jury verdict, and that the court’s exercise of that discretion is usually unreviewable); Barton’s Disposal Servs., Inc. v. Tiger Corp., 886 F.2d 1430, 1434 (5th Cir. 1989) (noting that the court has “great latitude in the framing and structure of the instructions and special interrogatories”).

\textsuperscript{138} Going beyond impressionistic surveys to determine the actual rates at which courts use these various forms would present an onerous research project. That said, see Elizabeth G. Thornburg, The Power and the Process: Instructions and the Civil Jury, 66 FORDHAM L. REV. 1837, 1840 (1998) (observing that “[c]ourts and commentators agree that the majority of federal jury-tried civil cases are submitted to the jury using a general charge”); see also WRIGHT & MILLER, supra note 126, § 2505, at 172 (2d ed. 1995) (observing that “the use of Rule 49(a) [special verdicts] never has been widespread in the federal courts”).
Punitive Damages and the “Hard Look”

A court that does choose to use one of the Rule 49 alternatives has similarly vast discretion to choose the form of the questions it will pose to the jury. Rule 49(a) speaks to this issue directly, providing,

the court [using a special verdict form] may submit to the jury written questions susceptible of categorical or other brief answer or may submit written forms of the several special findings which might properly be made under the pleadings and evidence; or it may use such other method of submitting the issues and requiring the written findings thereon as it deems most appropriate.

Although Rule 49(b) lacks parallel language, courts have interpreted it to incorporate Rule 49(a)’s broad grant of power over form.

Some courts have suggested that this discretion is limited by the principle that they should only pose questions of “ultimate fact,” which ask juries to categorize events in terms of legal concepts (e.g., “Did George drive negligently?”). On this view, courts should refrain from posing questions of “evidentiary fact,” which, roughly speaking, ask juries to describe events in “lay” terms without plugging them into legal pigeonholes (e.g., “Did George run the red light?”). This limitation is suspect for three reasons. First, the distinction between ultimate and evidentiary facts is famously unhelpful—there is no clear boundary between meaningful and meaningless questions of fact.

139. See, e.g., Romano v. Howarth, 998 F.2d 101, 104 (2d Cir. 1993) (noting that district courts have broad discretion to determine form of special interrogatories posed to jury); Lummus Indus., Inc. v. D.M. & E. Corp., 862 F.2d 267, 273 (Fed. Cir. 1988) (observing that trial judge has broad discretion to determine form of jury verdict and that exercise of that discretion is not ordinarily reviewable); Thornburg, supra note 138, at 1842 (observing that “[s]ix decades of case law” have failed to provide district courts with meaningful guidance vis-à-vis drafting special verdicts and interrogatories and collecting cases demonstrating that courts have approved use of extremely broad “omnibus questions” as well as extremely narrow questions directed at specific factual and legal contentions).

140. FED. R. CIV. P. 49(a) (emphasis added).

141. WRIGHT & MILLER, supra note 126, § 2512, at 221.

142. WRIGHT & MILLER, supra note 126, § 2512, at 221; see also, e.g., Act Up!/Portland v. Bagley, 988 F.2d 868, 876 (9th Cir. 1993) (Norris, J., dissenting from denial of petition for rehearing en banc) (noting that special verdicts generally pose questions of ultimate fact). For a discussion of the definition of “ultimate fact,” see Universal Minerals, Inc. v. C.A. Hughes & Co., 669 F.2d 98, 102 (3d Cir. 1981) (“An ultimate fact is usually expressed in the language of a standard enunciated by case-law rule or by statute, e.g., an actor’s conduct was negligent; the injury occurred in the course of employment; the rate is reasonable; the company has refused to bargain collectively.”) (internal citations omitted).

143. See Univ. Minerals, 669 F.2d at 102 (contrasting “basic facts” and “inferred factual conclusions” concerning “historical and narrative events” from “ultimate facts” which depend upon application of “legal precepts”).

1031
separating the two. Second, notwithstanding the haziness of this distinction, one can find examples of cases in which courts have asked juries to find extremely specific facts that surely fall on the “evidentiary” side of the line. Third, the plain text of Rule 49 contains no such limitation.

Despite the breadth of their discretion, in practice, courts seem to have adopted a crabbed approach to Rule 49 that tends to minimize the information that special verdict forms and interrogatories obtain from juries. For liability issues, courts will often pose questions in yes/no or check-a-box formats (e.g., “Do you find by a preponderance of the evidence that the defendant was negligent?”). For damages, courts seem generally to ask juries to fill in a blank (e.g., “The defendant is liable to the plaintiff for $_______.”).

On its face, however, Rule 49 authorizes courts to use whatever verdict forms they deem “most appropriate.” This broad language suggests that, where “appropriate,” courts could ask juries much more open-ended questions. For example, rather than ask a jury “Was the defendant negligent?,” a court might instead ask a jury to “Describe the defendant’s negligent acts.” By asking questions that invite descriptive

144. WRIGHT & MILLER, supra note 126, §1218, at 179:

Unfortunately, as was amply demonstrated by years of frustrating experience, it was difficult, if not impossible, to draw meaningful and consistent distinctions among ‘evidence,’ [ultimate] ‘facts,’ and ‘conclusions [of law].’ These concepts tended to merge to form a continuum and no readily apparent dividing markers developed to separate them.

145. See, e.g., Warlick v. Cross, 969 F.2d 303, 305 (7th Cir. 1992) (using interrogatory to determine if police officer planted evidence); Tillman v. Great Am. Indem. Co. of N.Y., 207 F.2d 588, 591 (7th Cir. 1953) (making detailed inquiries concerning defendant’s control of automobile); Gelfand v. Strohecker, Inc., 150 F. Supp. 655, 663 (N.D. Ohio 1956) (taking the extreme position that “[a]n interrogatory which sought merely to determine whether the defendant was negligent, without requiring the determination of the supporting facts would be improper”), aff’d 243 F.2d 797 (6th Cir. 1957).


147. See, e.g., id. For another typical example of jury interrogatories exemplifying this practice, see, for example, Dreiling v. General Electric Co., 511 F.2d 768, 774 (5th Cir. 1975). In that case, the jury was asked to answer the following questions:

1. Was General Electric Company negligent? Yes____ No____
2. Did the Pacemaker have a defect at the time General Electric sold it? Yes____ No____
3. What is the total amount of damages suffered by Mrs. McLelland? $_______

Id.

148. FED. R. CIV. P. 49(a).
Punitive Damages and the “Hard Look”

responses or “explanatory verdicts,” courts obviously could learn a great deal more information concerning jury reasoning than by spoonfeeding leading questions that must be answered by an unexplained “yes” or “no.”

Perhaps surprisingly, it is difficult to find published opinions discussing the propriety of asking open-ended interrogatories. One federal district court in a cursory opinion rejected the notion that they “must be put to the jury” in response to a defendant’s request. On the other hand, one can find cases in which courts have asked juries to give descriptive, explanatory answers in their verdicts without exploring the propriety of doing so. In short, the law on this point remains somewhat

149. Phillips v. N.W. Nat’l Ins. Co., 516 F. Supp. 762, 765 (W.D. Pa. 1981). In Phillips, the court rejected the defendant insurer’s request for an interrogatory asking the jury to identify specific language in its policy providing coverage to the plaintiff. The court noted that the insurer had “cited no cases that suggest such an open-ended interrogatory must be put to the jury.” Id. It then ruled that its decision “to submit only those interrogatories capable of simple and direct answers seems in accord with the preferred view.” Id. (citing Bertinelli v. Galoni, 200 A. 58, 60 (Pa. 1938); 89 C.J.S. TRIAL § 531 (1981)). The court’s discussion is interesting for two reasons. First, it does not explain why the jury’s answer to an interrogatory asking for identification of material contractual language would be anything other than “simple and direct.” Second, the insurer’s failure to cite supporting cases in tandem with the court’s reliance on a forty-three-year-old state court case and unclear secondary authority suggests the paucity of law on point.

150. The Supreme Court itself has supplied an example of the use of “open-ended” jury interrogatories—albeit in a case that did not discuss their propriety. In Walker v. New Mexico & South Pacific Railroad, the appellant challenged the constitutionality of a statute of the territory of New Mexico that authorized courts to require juries to return “special findings of fact” and to enter judgment on such findings even where they conflicted with a jury’s general verdict. Walker v. N.M. & S. Pac. R.R., 165 U.S. 593, 594–95 (1897). In holding that this precursor to Rule 49(b) did not violate the Seventh Amendment, the Court noted that the use of special verdicts had been recognized at common law and that “[i]t was also a common practice when no special verdict was demanded and when only a general verdict was returned to interrogate the jury upon special matters of fact.” Id. at 597. Later in the opinion, the Court quoted some of the jury interrogatories and their answers, including the following:

Q. 9. If you state in answer to the last question that there was such an arroyo, state where it is, its length, breadth, and the height of its banks. A. West of the city of Socorro and east of the Catholic graveyard; its banks are about two feet, its width about sixty feet, and about a mile in length, more or less.

Q. 14. How far from the main line of the railroad, in a westerly direction, are the mouths of the arroyos testified to by the witnesses? A. Three quarters mile to main arroyo, and one quarter of a mile to lower arroyo.

Q. 15. What is the character of the land lying between the mouths of the arroyos and the main line of the railroad is it level or sloping, and for what purposes was it used in 1886? A. It is level now, and in 1886 it was an arroyo, and there is no ditch now excepting the company drain.

Id. at 601. For an especially interesting example of the use of open-ended interrogatories, see Rowland v. Mad River Local School District, 730 F.2d 444, 456–60 (6th Cir. 1984). Appendix A to
unsettled. That said, given the expansive language of Rule 49 and the many cases stressing the breadth of trial court discretion over verdict forms, a court that deemed it “appropriate” to ask a jury to answer an open-ended interrogatory should have ample authority to do so.

B. Fixing Errors by Recharging Juries with Corrective Instructions

Moving past issues of form, the point of asking juries to answer interrogatories is that it sometimes exposes errors, which raises the question of how courts should act to correct them. For instance, in a negligence case, a jury might find that the defendant exercised due care yet at the same time find it liable for damages. Rule 49(b) offers courts the following options for dealing with such inconsistencies:

When the answers [to interrogatories] are consistent with each other but one or more is inconsistent with the general verdict, judgment may be entered pursuant to Rule 58 in accordance with the answers, notwithstanding the general verdict, or the court may return the jury for further consideration of its answers and verdict or may order a new trial. When the answers are inconsistent with each other or one or more is likewise inconsistent with the general verdict, judgment shall not be entered, but the court shall return the jury for further consideration of its answers and verdict or shall order a new trial.151

To avoid undue interference with jury power, a court should make every reasonable effort to reconcile apparent inconsistencies among the dissent in this case set forth in full the special verdict form that the magistrate judge had submitted to the jury. Id. Several of its questions asked the jury to explain its responses to other yes/no questions. For instance, the form queried:

1. Did Mrs. Rowland’s statement . . . regarding Mrs. Rowland’s love for another woman in any way interfere with the proper performance of either Mrs. Rowland’s or Elaine Monell’s duties . . . ?

2. If your answer to question 1 was “YES,” state in the notebook provided how the statement in question interfered with the performance of duties . . . .

Id. at 456; see also Mark S. Brodin, Accuracy, Efficiency, and Accountability in the Litigation Process—The Case for the Fact Verdict, 59 U. CIN L. REV. 15, 93 (1990) (praising use of combination of “Yes/No” and “Explain How” questions in Rowland for enabling “jurors to expand upon their responses in narrative fashion while at the same time confining them to the pivotal fact disputes”). For a third example of use of open-ended interrogatories, see infra note 175 (setting forth portion of verdict form in death-penalty case that invited jurors to describe mitigating factors).

151. FED. R. CIV. P. 49(b) (emphasis added).
interrogatory answers and a general verdict.\textsuperscript{152} But, where a court’s reconciliation efforts fail, the court may order a jury to deliberate further in the hope that it will resolve the inconsistency.\textsuperscript{153} Where inconsistent answers indicate that the jury found the court’s initial instructions confusing, the court should offer clarification.\textsuperscript{154} In other words, rather than simply ignore a jury’s errors or order a new trial, a court may take appropriate steps to help the jury fix its mistakes. Alternatively, if the jury’s responses suggest an unwillingness or inability to follow the law, the court should order a new trial.\textsuperscript{155}

The court’s post-verdict power to help the jury is not limited to correcting the jury’s mistakes but extends to correcting its own mistakes. If a jury goes astray because the court gave incorrect or incomplete instructions on the law, it may, in an appropriate case, recharge the jury for further deliberations in light of corrective instructions. For instance, in the sexual harassment case \textit{Bonner v. Guccione},\textsuperscript{156} the district court determined that the jury’s special verdict answers, though not plainly inconsistent, had been affected by a mistaken instruction concerning the applicable statute of limitations under the New York State Human Rights Law (NYSHRL).\textsuperscript{157} The plaintiff’s attorney brought this problem to the

\begin{footnotesize}
\begin{itemize}
\item 152. \textit{See, e.g.}, Atl. & Gulf Stevedores, Inc. v. Ellerman Lines, Ltd., 369 U.S. 355, 364 (1962) (observing that “[w]here there is a view of the case that makes the jury’s answers to special interrogatories consistent, they must be resolved that way”); Wilks v. Reyes, 5 F.3d 412, 415–16 (9th Cir. 1993) (observing that courts must make every reasonable effort to harmonize components of jury verdicts and holding that general verdict that defendant was liable for violating plaintiff’s civil rights was not inconsistent with jury’s determination to award no damages).
\item 153. \textit{See, e.g.}, McLaughlin v. Fellows Gear Shaper Co., 786 F.2d 592, 596–97 (3d Cir. 1986) (upholding resubmission of case to jury for further deliberations to resolve inconsistencies in light of supplemental interrogatories).
\item 154. \textit{See, e.g.}, Hafner v. Brown, 983 F.2d 570, 573–75 (4th Cir. 1992) (upholding district court’s supplemental, post-verdict instructions to jurors clarifying rules for awarding compensatory and punitive damages where jury had assessed punitive damages against some defendants without finding them liable for compensatory damages); \textit{McLaughlin}, 786 F.2d at 596–97 (upholding, in products liability case, trial court’s submission of supplemental interrogatories on the issue of foreseeability intended to resolve inconsistencies in jury’s answers to special interrogatories). \textit{But cf.} Jacobs Mfg. Co. v. Sam Brown Co., 19 F.3d 1259, 1267 (8th Cir. 1994) (finding that trial court may not invite inconsistency by submitting supplemental, post-verdict interrogatories to a jury that has delivered a clear, unambiguous verdict), \textit{cert. denied}, 513 U.S. 1190 (1995).
\item 155. \textit{Wright & Miller, supra} note 126, § 2513, at 233 (observing that a court’s decision whether to order a new trial “should reflect the degree of confidence it has in the jury’s ability to straighten the inconsistency out satisfactorily without compromising the fairness of the process or the integrity of the result”).
\item 156. 178 F.3d 581 (2d Cir. 1999).
\item 157. \textit{Id.} at 584–85, 587.
\end{itemize}
\end{footnotesize}
court’s attention after the faulty instructions had been read to the jury but before it had retired to deliberate.\textsuperscript{158} Rather than seek immediate correction, which, in the plaintiff’s attorney’s strategic view, would have “emphasize[d] something that the jury hasn’t indicated they are in a quandary over,” she instead requested that the charge be corrected, if necessary, by a post-verdict interrogatory.\textsuperscript{159} The jury returned special verdicts finding the defendant liable for creating a hostile work environment but awarding no damages, a result that suggested to the court that the jury’s damages analysis had been affected by the incorrect statute-of-limitations instruction, which permitted the plaintiff to recover damages for only a very short period of time.\textsuperscript{160} The district court responded by giving the jury a supplemental instruction on the correct statute of limitations under the NYSHRL, and the jury, given the chance to reconsider, awarded $90,000 in damages.\textsuperscript{161}

The Second Circuit upheld the trial court’s decision to issue the supplemental instruction.\textsuperscript{162} The defendants had urged on appeal that it was improper for the trial court to correct its mistake because, among other reasons, the plaintiff had waived her objection to the original instruction by failing to seek immediate correction.\textsuperscript{163} The defendants based this argument on Federal Rule of Civil Procedure 51, which governs jury instructions and provides, in pertinent part, that “[n]o party may assign as error the giving or the failure to give an instruction unless that party objects thereto before the jury retires to consider its verdict . . . .”\textsuperscript{164} The appellate court rejected this argument. As one ground for doing so, it noted that the purpose of Rule 51 is “to prevent unnecessary new trials because of errors the judge might have corrected if they had been brought to his attention at the proper time.”\textsuperscript{165} Because

\footnotesize{\begin{itemize}
\item \textsuperscript{158} Id. at 584.
\item \textsuperscript{159} Id.
\item \textsuperscript{160} See id. at 585.
\item \textsuperscript{161} Id.
\item \textsuperscript{162} Id. at 586–88.
\item \textsuperscript{163} See id. at 586.
\item \textsuperscript{164} \textsc{Fed. R. Civ. P.} 51. In most Circuits, appellate courts will review objections that are untimely under Rule 51 (where it applies) only for “plain error”; the Second and Ninth Circuits are even more draconian and generally treat failure to satisfy Rule 51 as a complete block to appellate review of instructional error. 9 \textsc{James William Moore et al., Moore’s Federal Practice}, § 51.21[2], at 51-49-50 (Matthew Bender 3d ed.).
\item \textsuperscript{165} \textit{Bonner}, 178 F.3d at 586 (quoting \textsc{Cohen v. Franchard Corp.} 478 F.2d 115, 122 (2d Cir. 1973)).
\end{itemize}}
the trial court had corrected its instructional error before discharging the jury, there was no need for a retrial, and Rule 51 was not implicated.\textsuperscript{166} As further support for this analysis, the court cited the Supreme Court’s decision in \textit{City of Newport v. Fact Concerts, Inc.}\textsuperscript{167} In that civil rights action, a jury found the defendant city liable for punitive damages and set forth the amount in a special verdict form that distinguished punitive from compensatory damages (i.e., it was clear from the verdict form how much of the total award was compensatory and how much was punitive—this is not always the case).\textsuperscript{168} The city sought to overturn the punitive damages award on the ground that municipalities cannot be held liable for such damages under Section 1983.\textsuperscript{169} The plaintiff argued that, under Rule 51, the city had waived this argument by failing to timely object to the district court’s instructions, which had authorized the jury to award punitive damages.\textsuperscript{170} The district court first rejected the plaintiff’s waiver argument and then rejected the city’s belated defense on its merits.\textsuperscript{171}

On appeal, the plaintiff continued to press its waiver argument. The Supreme Court also rejected it, taking a functional view of Rule 51 and noting that, “[b]ecause the District Court reached and fully adjudicated the merits, and the Court of Appeals did not disagree with that adjudication, no interests in fair and effective trial administration advanced by Rule 51 would be served if we refused now to reach the merits ourselves.”\textsuperscript{172} In this pragmatic vein, the Court observed that the district court’s waiver analysis

may have been influenced by the unusual nature of the instant situation. Ordinarily, an error in the charge is difficult, if not impossible, to correct without retrial, in light of the jury’s general

\begin{itemize}
\item \textsuperscript{166} Id. at 586–88; \textit{see also} Barrett v. Orange County Human Rights Comm’n, 194 F.3d 341, 349 (2d Cir. 1999) (plaintiff could object to instructions issued midway through deliberations even though plaintiff could have but failed to make same objection at time of original instructions). \textit{But cf.} Parker v. City of Nashua, 76 F.3d 9, 13 (1st Cir. 1996) (noting that few cases address whether waiver principles block a party that has failed to properly object to instructions before a jury first retires to deliberate from later objecting to same instructions when they are reissued to jury midway through deliberations).
\item \textsuperscript{167} 453 U.S. 247 (1981).
\item \textsuperscript{168} Id. at 253.
\item \textsuperscript{169} Id.
\item \textsuperscript{170} Id. at 255.
\item \textsuperscript{171} Id. at 253–55.
\item \textsuperscript{172} Id. at 256.
\end{itemize}
verdict. In this case, however, we deal with a wholly separable issue of law, on which the jury rendered a special verdict susceptible of rectification without further jury proceedings.173

In other words, the district court could have granted the defendant city’s post-trial motion to strike the punitive damages portion of the verdict without any need for a new trial. Therefore, Rule 51 did not block the city’s belated objection because the rule’s efficiency rationale was largely inapplicable.

In short, regardless of whether the source of legal error is a court’s mistaken instruction or a jury’s misunderstanding of a correct instruction, Rule 49 and cases like Bonner teach that trial courts can, with due care, ask juries to reconsider their verdicts in light of corrective instructions. Sometimes, instructing the jury can be an iterative process.

C. Explanatory Verdicts, Iterative Instructions, and Punitive Damages

It could prove fairly simple and highly informative for courts to use open-ended jury interrogatories and iterative instructions to improve procedures for determining and reviewing punitive damages verdicts. Again, the law typically requires juries to determine the amount of punitive damages awards in light of their consideration of a list of factors bearing on retribution and deterrence. For example, juries may be required to consider, among other factors, the seriousness of the defendant’s tort, its awareness of the dangers it created, the profitability of the misconduct, and the defendant’s financial state.174 In other words, the law instructs jurors to mull over various categories of facts concerning the defendant and its conduct. Having thought about the “right” facts, juries are then to choose how harshly to punish. Knowing which facts a jury deemed material to punishment would enable a reviewing court both: (a) to make sure that the jury thought about the “right stuff”; and (b) to make sure that these facts can “reasonably” support the jury’s award (i.e., that the jury has not made a crazy policy choice).

Courts could obtain such information by using open-ended interrogatories to ask juries to identify the facts that motivate their

173. Id. at 256 n.12.
174. See supra note 56 and accompanying text.
Punitive Damages and the “Hard Look”

punishment decisions. A court adopting such an approach might submit an interrogatory looking something like this to a jury:

If you decide that the defendant committed [fill in name of tort] with reckless disregard for the rights of others, the law permits but does not require you to award punitive damages against the defendant to punish the defendant and discourage it or others from committing such acts in the future. As you decide how much, if any, in punitive damages to award, you should think about: [fill in the jurisdiction’s punitive damages factors].

If you award punitive damages, explain what facts about the defendant and its conduct influenced your decisions to award punitive damages and how much to award.

The jury’s response to such an interrogatory would form part of the verdict read in court. With such an explanatory verdict in hand, the court—with the assistance and insistence of counsel—could, to use administrative law terms, take a “hard look” at the jury’s punitive

175 There are, of course, many different ways one could pose interrogatories to try to learn the “punitive story” behind an award. One might, for instance, pose a separate interrogatory for each punitive damages factor the jury was instructed to consider (e.g., “Identify any facts about the defendant’s financial condition that influenced your decision concerning how much to award,” etc.). Such an approach could have the advantage of focusing jury consideration on each of the factors the law declares relevant. It might, however, also encourage juries to confabulate and make up reactions that they did not have. My perhaps naïve psychological speculation is that one would obtain the most insight into jury deliberations by using open-ended interrogatories that are as general as possible and thus do not force juries to describe their findings in preconceived pigeonholes.

A recent death-penalty case, United States v. Bin Laden, provides an example of a hybrid approach. The verdict form listed a specific set of mitigating factors on which the defense had based its argument for mercy. The form also, however, informed jurors:

The law does not limit your consideration of mitigating factors to those that can be articulated in advance. Therefore, you may consider during your deliberations any other factor or factors in [the defendant’s] background, record, character, or any other circumstances of the offense that mitigate against imposition of a death sentence.

The following extra spaces are provided to write in additional mitigating factors, if any, found by any one or more jurors. If more space is needed, write ‘continued’ and use the reverse side of this page.

United States v. Bin Laden, No. S(7) 98 Cr. 1023, Transcript of Trial at 7254 (S.D.N.Y., June 5, 2001), available at http://www.cryptome.org/usa-v-ubl-59.htm. In response, the jury described five “unlisted” mitigating factors: killing the defendant would make him a martyr; killing him would not necessarily alleviate the suffering of the victims; lethal injection would not make him suffer; life imprisonment was worse than death; and the defendant was “raised in a completely different culture and belief system.” Id. at 7335 (S.D.N.Y., June 12, 2001), available at http://cryptome.kaiizo.org/usa-v-ubl-63.htm.
damages verdict to ensure that its punishment policy decision was minimally reasonable and based on consideration of facts that the law deems relevant to punishment.

A court that discovered that a jury erred by basing punishment on the wrong kinds of considerations could take steps to correct such error in an efficient manner. As discussed in more detail below, it is clear that, in BMW of North America, Inc. v. Gore, an Alabama jury improperly punished BMW for conduct that occurred outside Alabama. There is also strong evidence that, in Texaco, Inc. v. Pennzoil Co.—the first case to culminate in a multi-billion dollar punitive damages verdict—the jury improperly based its punishment of Texaco on the conduct of Getty Oil representatives. It is impossible to know how often juries have made similar errors in other cases because we do not make a general practice of asking them to explain the grounds for their awards, but no doubt there have been other such incidents.

Confronted with such a mistake in an explanatory verdict, a judge would not be helpless. Punitive damages instructions should not aspire to be complete guides for jury deliberations, e.g., a court cannot, as a practical matter, list for a jury every “irrelevant factor” that should not affect punishment. Nonetheless, where a jury bases punishment on an “irrelevant factor,” one could “blame” the instructions—i.e., if only the court had clearly told the jury that it should not base punishment on factor X, the jury would not have done so. As Bonner demonstrates, where an instructional error taints deliberations, a judge may recharge the jury with clarifying instructions to help it correct its verdict. Thus, a court could ask a jury to reconsider its punitive damages award in light of an instruction that factor X has no bearing on punishment.

177. See infra text accompanying notes 211–16.
179. See infra note 232 and accompanying text.
180. See supra notes 156–66 and accompanying text.
181. In some cases, a court might determine that a jury is incapable of discounting from punishment an irrelevant factor it has considered—it cannot psychologically “unring the bell.” In such situations, a court could order remittitur or a new trial. On the other hand, there is no reason to think that the BMW jury, for example, could not have cured its verdict of legal error given the benefit of corrective instructions and thus saved the parties years of post-trial litigation. See infra text accompanying notes 211–24.
The chief obstacle to adopting such an approach could prove to be judicial attitudes toward waiver. Courts have frequently rejected appeals of punitive damages awards on the ground that defendants have waived the right to claim instructional error by failing to make timely objections. Similarly, one could blame a court’s failure to instruct a jury that factor X is irrelevant to punishment on the failure of defense counsel to seek such an instruction in a timely manner. As discussed above, however, it is important to understand Rule 51 in light of its purpose, which is to give trial courts a chance to fix instructional errors before their correction requires the time and expense of a retrial. By the time a case reaches an appellate court, the only way to fix a faulty instruction is generally by ordering a new trial. In such situations, one should expect courts to aggressively apply waiver rules. As cases such as City of Newport and Bonner demonstrate, however, where a court can correct an error without retrying a case, there is little reason to take a draconian approach to waiver. The instant suggestion that courts should issue corrective instructions to help juries fix errors in punitive damages determinations fits within the logic of these latter cases—the whole point of this iterative approach to instructions is to fix errors without new trials.

It should also be noted that a strict waiver approach is especially unsuited to punitive damages determinations in light of their open-ended nature. Again, punitive damages instructions cannot list for juries every factor they should not consider as a basis for punishment, and it would be foolish and counterproductive to try. It therefore makes no sense to insist on dogmatic application of waiver rules premised on the idea that the instructions should be absolutely complete and correct the first time.

182. See, e.g., Inter Med. Supplies, Ltd. v. EBI Med. Sys., Inc., 181 F.3d 446, 463 (3d Cir. 1999) (ruling that defendant waived objection to punitive damages instructions by failing to timely object), cert. denied, 528 U.S. 1076 (2000); Grynberg v. Citation Oil & Gas Corp., 573 N.W.2d 493, 503–04 (S.D. 1997) (same).

183. See, e.g., Reynolds v. Green, 184 F.3d 589, 595 (6th Cir. 1999) (noting that Rule 51 “was not intended to require pointless formalities” but was instead “designed to prevent unnecessary new trials”); Bonner v. Guccione, 178 F.3d 581, 586 (2d Cir. 1999); Greaser v. Mo. Dept. of Corrs., 145 F.3d 979, 984 (8th Cir. 1998) (noting that Rule 51 prevents litigants from strategically using defective instructions as grounds for new trial where they could have sought correction in time for court to cure error without new trial).

184. See supra notes 156–73 and accompanying text.

185. Listing everything that a jury should not think about is impractical for at least two reasons: (1) there is an infinite number of irrelevant factors; and (2) listing a factor tends to draw attention to it, compare, an apocryphal cure for the hiccups—run around the house three times without thinking of the word “wolf.”
Instead, a court that discovers that its instructions need clarification or correction—rather than cry waiver and ignore the error—should help the jury to follow the law by asking it to reconsider its award in light of supplemental instructions.

Requiring juries to submit explanatory verdicts that tell the “stories” behind their punitive damages awards is a legitimate procedural option. Moreover, in appropriate cases, where an explanatory verdict indicates that such an award has been tainted by correctable error, a court may ask the jury to reconsider its award in light of corrective instructions.

IV. COURTS SHOULD ASK JURIES TO EXPLAIN THEIR PUNITIVE DAMAGES VERDICTS AND TAKE “HARD LOOKS” AT THE RESULTS

Requiring juries to describe the factual bases for their punitive damages awards would offer at least three important advantages over current practice. First, although this mild reform would not fundamentally alter the power of juries to make punishment policy by intuiting formulae for transmuting malicious torts into dollars, it would enable courts to take “hard looks” at jury verdicts to ensure that they are at least based on minimally sensible consideration of legally relevant factors. Second, explanatory verdicts would enable juries to express themselves more clearly. The law has recognized since 1763 that punitive damages awards are supposed to send messages—let these messages be as articulate as possible. Third, explanatory verdicts would provide useful information for determining whether juries should possess the power to inflict punitive damages in the first place.

A. A Closer Look at the “Hard Look” and Its Suitability for Punitive Damages Review

It is a fundamental principle of modern administrative law that regulatory agencies must explain the bases for their significant discretionary policy choices that are subject to judicial review. If such a choice is subjected to legal challenge, a reviewing court will examine

186. See supra note 43 and accompanying text.
187. See Lawson, supra note 19, at 319 (“Where [agency] discretion involves an issue of policy significance, well-settled principles of administrative law typically impose a substantial duty of explanation on the agency.”).
Punitive Damages and the “Hard Look”

the agency’s explanation to ensure that the agency based its decision on legal and minimally sensible deliberations. 188 Exploring why courts have taken this approach to review of agency policymaking will shed light on why they should adopt a similar approach for reviewing punitive damages verdicts.

Our legal system once took the view that a reviewing court should affirm those administrative policy decisions for which it could dream up any minimally rational, marginally sane, supporting rationale. 189 Under such a system, it would not matter if the EPA promulgated rules to control air pollution based on its reading of goat entrails so long as, at the end of the day, a reviewing court concluded that the regulations so promulgated happened to amount to minimally reasonable ways to implement the Clean Air Act. 190 This is a form of outcome review and in this respect akin to review of punitive damages awards for excessiveness under current practice.

Courts no longer cede this much power to agencies. For starters, they now refuse to make up hypothetical rationales to support agency choices. Instead, under the venerable precedent of SEC v. Chenery Corp., 191 “[t]he grounds upon which an administrative order must be judged are those upon which the record discloses that its action was based.” 192 A primary reason for this refusal is that relying on judge-made rationales to support administrative action obviates much of the point of setting up “expert” agencies in the first place. To stick with the Clean Air Act example, we want the EPA’s considered judgment on how much air pollution to permit, not the views of generalist courts. 193 Of course, to judge the

188. See infra notes 195–98 and accompanying text.
189. See Pac. States Box & Basket Co. v. White, 296 U.S. 176, 185–86 (1935) (turning back challenge to Oregon regulation requiring uniform sizes for berry containers and holding that administrative action, like legislative action, must be upheld “if any state of facts reasonably can be conceived that would sustain it”).
190. The EPA might get lucky, or animal divination might work. “There are more things in heaven and earth, Horatio, than are dreamt of in your philosophy.” WILLIAM SHAKESPEARE, THE TRAGEDY OF HAMLET act I, sc. 5; see also Lawson, supra note 19, at 318–19 (providing an illuminating discussion of judicial review of agency decisionmaking processes and noting that under current doctrine, an agency decision based on astrological divination would fail the Administrative Procedure Act’s “arbitrary or capricious” test).
191. 318 U.S. 80 (1943).
192. Id. at 87.
193. Cf. id. at 88 (“If an [administrative] order is valid only as a determination of policy or judgment which the agency alone is authorized to make and which it has not made, a judicial judgment cannot be made to do service for an administrative judgment.”).
sufficiency of an agency’s rationale supporting its policy choice, a
reviewing court must require the agency to explain that rationale. 194

Technically, under the Administrative Procedure Act, courts will
vacate those agency policy choices they deem “arbitrary and
capricious.” 195 Courts developed the “hard look” doctrine as a standard
for determining whether an agency explanation satisfies this ostensibly
deferential standard in certain contexts. 196 The Supreme Court offered the
standard summary of this type of review in Motor Vehicle Manufacturers
Ass’n of the United States v. State Farm Mutual Automobile Insurance
Co., 197 stating,

[t]he scope of review under the “arbitrary and capricious” standard
is narrow and a court is not to substitute its judgment for that of the
agency. Nevertheless, the agency must examine the relevant data
and articulate a satisfactory explanation for its action including a
rational connection between the facts found and the choice made.
In reviewing that explanation, we must consider whether the
decision was based on consideration of the relevant factors and
whether there has been a clear error of judgment. Normally, an
agency rule would be arbitrary and capricious if the agency has
relied on factors which Congress has not intended it to consider,
entirely failed to consider an important aspect of the problem,
offered an explanation for its decision that runs counter to the
evidence before the agency, or is so implausible that it could not be
ascribed to a difference in view or the product of agency
expertise. 198

Court could not determine whether Secretary of Transportation’s informal adjudication authorizing
expenditure of federal funds to construct highway through park was arbitrary and capricious without
access to the “administrative record” before the agency; remanding to district court for review
“based on the full administrative record that was before the Secretary at the time he made his
decision”).
195 5 U.S.C. § 706(2)(A) (1995) (providing that courts should set aside agency action that is
“arbitrary, capricious, an abuse of discretion, or otherwise not in accordance with law”).
196 See, e.g., Nat’l Lime Ass’n v. EPA, 627 F.2d 416, 451 n.126 (D.C. Cir. 1980) (describing
evolution of the “hard look” doctrine); Sandra B. Zellmer, The Devil, the Details, and the Dawn of
the 21st Century Administrative State: Beyond the New Deal, 32 Ariz. St. L.J. 941, 999–1003
198 Id. at 43 (quotation marks and citations omitted; emphasis added).
The “hard look” requires agencies to base their policy choices on reasoned decision-making that can be documented to a court’s satisfaction. One can crudely divide agency violations of this requirement into two kinds of error. First, the *legal* dimension of hard-look review ensures that, during its deliberations, an agency has thought about the kinds of things the law has told it to think about, i.e., that the agency has based its action on consideration of the “relevant factors.” Second, the *rationality* dimension of hard-look review ensures that an agency’s explanation makes some minimal amount of sense, i.e., the agency’s policy choice and the reasoning underlying it are not tainted by “clear error[s] of judgment.”

A federal agency must always base its actions on consideration of the “relevant factors” that Congress has told it to think about in the agency’s organic statute. For instance, under the Clean Air Act (CAA), the EPA has a duty to set national ambient air quality standards (NAAQs) for pollutants at levels that “are requisite to protect the public health” with “an adequate margin of safety.” Thus, the “relevant factors” to consider when promulgating a NAAQ are health and safety. Stipulate that it would make vastly more policy sense for the EPA also to consider industry implementation costs when it sets permissible pollutant levels—affected industries think so anyway. No doubt the EPA could promulgate fabulously “reasonable” NAAQs based on careful cost-benefit analyses that would surely survive substantive review for “clear errors of judgment.” The Supreme Court has nonetheless observed that such fabulous NAAQs would be illegal for the simple reason that Congress did not tell EPA to consider cost—it is an irrelevant factor so far as this portion of the CAA is concerned. The best reasoned, most sensible agency policy choice in the world must fail if the agency based that choice on illegal consideration of categories of facts and goals that Congress did not instruct it to consider.

In addition to satisfying legal requirements, an agency must also satisfy a reviewing court that its deliberations were minimally reasonable. Inevitably, the strictness of this rationality review varies from court to court and from agency to agency both because rationality is

---

200. See Whitman v. Am. Trucking Ass’n, 531 U.S. 457, ___, 121 S. Ct. 903, 911 n.4 (2001) (observing that if respondents challenging EPA’s rules on particulate matter and ozone standards could prove that EPA was “secretly considering the costs of attainment without telling anyone,” then there would be “grounds for vacating [these rules] because the Administrator had not followed the law”).
often in the eye of the beholder and because courts must review many different kinds of actions taken by many different agencies. Still, the Supreme Court has stressed that this standard is supposed to be highly deferential—a court is “not to substitute its judgment for that of the agency.”

Having determined that an agency’s action failed a “hard look,” a court should not order implementation of its own favorite policy choice in place of the agency’s flawed one. To do so would usurp the role that the legislature assigned to the agency, not the court. Instead, the court should issue an opinion explaining where the agency went wrong and remand—leaving the agency free to attempt its policymaking process again.

There are clear parallels between the structure of administrative policymaking and jury punishment selection that make the “hard look” especially suitable for judicial review of punitive damages awards. In both contexts, the state has granted decision-makers a measure of discretion to make policy choices in light of various legally constraining “relevant factors.” Congress has told the EPA to clean up the air; it has not told the agency exactly how to accomplish this task in part because Congress does not have this information. The EPA is supposed to exercise its technical expertise to figure this out—that is its role. Congress has, however, limited this discretion by telling the EPA to be sure and focus on public health and safety concerns as it makes its policy choices.

Similarly, various jurisdictions have commanded juries to inflict punitive damages as necessary to punish malicious torts properly. They have not told juries exactly how much to award; instead, punitive damages law operates on the assumption that juries are our policy “experts” in charge of choosing just punishment. Jury discretion is


203. See BREYER ET AL., supra note 201, at 347 (noting that “normal remedy” for agency’s failure to satisfy a “hard look” is to “remand for further proceedings” in which “agency remains free to try again”).

204. See supra note 199 and accompanying text.

205. Cf. supra note 60 (citing case law discussing the “policymaking” nature of punitive damages determinations).
constrained, however, by law instructing jurors to base punishment on consideration of various “relevant factors” bearing on retribution and deterrence.

Judicial review for rationality is supposed to be quite deferential in both of these contexts. In the punitive damages context, “rationality” review takes the form of common-law “excessiveness” and constitutional “gross excessiveness” review. A court engaging in “excessiveness” review does not ask itself whether the award is greater than the amount the judge would order given the chance. Rather, in most jurisdictions, the court asks a more deferential question akin to “Does this award shock my conscience given the factors I am to consider?” The “gross” part of the Supreme Court’s “gross excessiveness” test likewise requires deference—a court that deems an award merely “excessive” should not strike it for violating due process; only those awards that go so far beyond the pale as to be “grossly excessive” merit this treatment. This deference to juries parallels deference to agency judgment in the administrative context—a policy choice does not amount to a “clear error of judgment” merely because a court disagrees with it.

In neither context is deference *carte blanche*, of course. A policy choice might strike a court as so totally senseless as to indicate a “clear error of judgment” (e.g., suppose that the EPA banned gasoline to reduce air pollution or that a jury inflicted a punitive damages award of $1 trillion against a defendant for a trivial tort). Or a decision-maker’s explanation for an action might demonstrate that its choice was tainted by consideration of an “irrelevant factor” (e.g., suppose a jury based a punitive damages award on a defendant’s nationality, or that the EPA considered industry costs in promulgating a NAAQ). In either case, the court should set aside the policy choice at issue.

The court should not, however, impose its own judgment to replace one that it has thrown out—to do so would usurp another body’s policymaking role. Thus, where an agency policy action fails review, a court typically will remand to give the agency a chance to revisit its

206. *See generally supra* notes 69–79 and accompanying text (discussing judicial review of punitive damages awards). *But cf.* supra notes 73, 88–105 and accompanying text (discussing cases suggesting more aggressive review of punitive damages awards).

207. The Supreme Court’s recent opinion in *Cooper Industries, Inc. v. Leatherman Tool Group, Inc.*, is not to the contrary. In *Cooper*, the court held appellate courts should conduct *de novo* review of district court evaluations of “gross excessiveness.” 532 U.S. 424, ___, 121 S. Ct. 1678, 1687–89 (2001). This holding did not alter the deferential relation between courts and jury that the “gross excessiveness” test entails. *See id.* at 1682 n.4 (noting this limitation of *Cooper*).
choice in light of the court’s correction. In the punitive damages context, juries do not get this same chance, but the logic of their policymaking role suggests they should. For instance, suppose a court determines that a jury’s explanatory punitive damages verdict fails “hard look” scrutiny because the jury based punishment on consideration of an irrelevant factor. Suppose also that nothing about the error suggests that the jury could not correct its mistake if given the benefit of appropriate clarifying instructions. Rather than impose its own punitive damages award through remittitur, the court should instead “remand” to the jury for further deliberations in light of corrective instructions. Alternatively, if the nature of the jury’s error suggests that it is incompetent or prejudiced, the court should order remittitur and/or a new trial on damages.208

Thus, it would make sense to apply some form of “hard look” review to punitive damages verdicts given the structure of punitive damages determinations and the premise that we want awards to reflect jury policy judgments. The only barrier to such application turns out to be lack of information concerning the basis of jury awards, and the only barrier to obtaining such information seems to be inertia. As discussed above, a strong argument can be made that courts have authority to ask juries for explanatory verdicts that describe the factual bases for their punitive damages awards.209 Moreover, a strong argument can be made that where an explanatory verdict indicates that a jury has made a correctable mistake in its deliberations, a court has authority to ask the jury to reconsider its award in light of corrective instructions.210 Thus, courts both could and should subject punitive damages verdicts to the “hard look.”


Two especially famous punitive damages cases—BMW and Texaco—provide perfect, concrete examples of how a “hard look” approach could improve procedures for ensuring the legality and rationality of punitive damages awards. In BMW, a “hard look” would have revealed that a

208. A wildly excessive award, e.g., a trillion dollar award for a trivial tort, might suggest that the jury is irremediably incompetent or prejudiced and could not fix its error with the benefit of corrective instructions. In this sort of situation, a court should order remittitur and/or a new trial on damages.
209. See supra notes 148–50 and accompanying text.
210. See supra notes 151–73 and accompanying text.
seemingly absurd verdict was the product of a reasonable jury’s legal mistake that it could have corrected if given the chance to learn the law. In *Texaco*, the “hard look” would have revealed that the world’s first multi-billion dollar punitive damages award was an illegal attempt to punish someone other than the defendant.

Recall that Dr. Gore sued BMW for selling him an expensive sports sedan as new without disclosing that it had been repainted to touch up acid rain damage and that the jury awarded him $4 million in punitive damages. Critics of punitive damages sometimes portray this award as an archetypal example of jury lunacy. This characterization is unfair—to all appearances, the BMW jury actually took an extremely rational approach to punitive damages that might make a law-and-economics scholar proud. Testimony at trial showed that the car sold to Dr. Gore was worth $4000 less than it would have been had it never been damaged and repainted; there was also evidence that BMW had sold 983 cars nationwide with undisclosed minor repairs. The jury determined that selling these cars without disclosing their repairs amounted to “gross, oppressive or malicious” fraud for which Alabama law authorizes punitive damages. From the plain math of the case, it seems apparent that the jury reached its $4 million punitive award by multiplying $4000 (Dr. Gore’s compensatory damages, which the jury used as a proxy for the profit BMW made from each nondisclosure) times 1000 (the approximate number of nondisclosures nationwide).

On appeal, the Alabama Supreme Court agreed that taking away BMW’s nationwide profits from nondisclosure was probably what the jury had in mind. But there was a problem—the court also ruled that the jury could only punish conduct that had occurred in Alabama, where BMW had sold only fourteen of the cars at issue.

---


213. *BMW*, 517 U.S. at 564.

214. *Id.* at 565.


216. *Id.* More specifically, the court held that, although the jury could consider extra-Alabama conduct to determine whether BMW had engaged in a “pattern or practice” or nondisclosure, the jury could not, as a matter of law, “use the number of similar acts that a defendant has committed in other jurisdictions as a multiplier when determining the dollar amount of a punitive damages
Suppose that, confronted with this problem, the court had tried to remit the award to a level that matched what we will call the jury’s punitive reaction—the policymaking step by which the jury transmuted the “bad” facts it had found into the dollar figure of $4 million. The jury’s punitive reaction was simply to “take away the profits.” To honor this punitive reaction, the court should have remitted the award to $56,000—the product of $4000 (estimated profit per nondisclosure) times fourteen (the number of nondisclosures in Alabama).\(^{217}\)

The court instead set itself the task of determining a “constitutionally reasonable” amount of punitive damages to punish the fourteen offending sales in Alabama.\(^{218}\) To this end, it purported to use a “comparative analysis” that considered verdicts in other cases “involving the sale of an automobile where the seller misrepresented the condition of the vehicle and the jury awarded punitive damages to the purchaser.”\(^{219}\) On the basis of this “analysis,” the court remitted the jury’s sensible (if legally mistaken) $4 million award to $2 million—a number which seems to have absolutely nothing to do with anything about the case except that it marked a 50% reduction in the jury’s punitive damages verdict.\(^{220}\)

BMW appealed to the Supreme Court, which later ruled that this $2 million remitted award violated due process because it was “grossly excessive” punishment for the fourteen Alabama nondisclosures under the three-guidepost analysis the Court devised for the occasion.\(^{221}\) Regardless of this due process point, however, it is clear that the real scandal of the case was not the jury’s original $4 million punitive

\(^{217}\) Id.

\(^{218}\) Id. at 627. Nonetheless, at the risk of naïve cognitive speculation, it seems brutally obvious that, absent encountering the $4 million verdict, the Alabama Supreme Court would never have dreamed of awarding $2 million for failing to disclose minor repairs to 14 cars. One hopes not, anyway.

\(^{219}\) Id.

\(^{220}\) Id. at 627. Nonetheless, at the risk of naïve cognitive speculation, it seems brutally obvious that, absent encountering the $4 million verdict, the Alabama Supreme Court would never have dreamed of awarding $2 million for failing to disclose minor repairs to 14 cars. One hopes not, anyway.

\(^{221}\) Id. at 627. Nonetheless, at the risk of naïve cognitive speculation, it seems brutally obvious that, absent encountering the $4 million verdict, the Alabama Supreme Court would never have dreamed of awarding $2 million for failing to disclose minor repairs to 14 cars. One hopes not, anyway.

\(^{217}\) Id. at 627.
Punitive Damages and the “Hard Look”

...damages verdict but rather that the Alabama Supreme Court substituted its own (rather bizarre) policymaking punitive reaction for that of the jury in the name of excessiveness review. Again, the jury’s punitive reaction was merely to “take away the profits”—to connect roughly 1000 nondisclosures to a $4 million award. The Alabama Supreme Court’s substitute punitive reaction connected fourteen nondisclosures with a $2 million award. If one estimates, as the jury seems to have done, that BMW fraudulently profited $4000 per car, then the $2 million remitted award equaled roughly thirty-six times BMW’s improper profits from the Alabama sales. There is no reason to think that a jury that had decided on a take-away-the-profits “punishment policy” would have fined BMW this amount.

Instead of subjecting the jury’s $4 million award to standard, outcome-based review, suppose that the Alabama courts had applied the proposed hard-look approach. To start, the court would have asked the jury to identify the facts upon which it had based punishment. The jury might well have responded with a punitive story something like:

Touching up the paint of a $40,000 BMW 535i sports sedan that some doctor bought is not the worst thing ever. But the law says its gross fraud, and BMW should not make money by violating the law. It profited about $4000 per car from nondisclosure. It failed to disclose repairs to “new car” buyers about 1000 times, give or take. So we’ve awarded $4 million in punitive damages to take away BMW’s illegal profits.

Surely such an explanation, which accords perfectly with the jury’s verdict, would show a highly rational chain of reasoning connecting the facts with the verdict and could not be said to constitute a “clear error of judgment.”222

By way of contrast, imagine that the jury’s explanation had made clear that it had awarded the $4 million solely to punish BMW for the sale of just one car to Dr. Gore. Adoption of a “punishment policy” connecting...

---

222. Indeed, in fairness, this hypothetical BMW explanation is perhaps too “reasonable” to serve as a model of how hard-look review of punitive damages awards for minimal deliberative rationality would generally work. The jury’s economic approach to punishment lends itself to an instrumental, means-end analysis—it wanted to deter wrongdoing by a profit-seeking entity, so it took the profit out of fraud. It is more difficult to say what it means for an award based on retributive concerns to be “reasonable.” How much financial pain does it take to right the scales of justice? The answer is not clear, which, indeed, is part of the reason that jury punishment selection is best regarded as policymaking rather than fact-finding. A court crediting the jury’s role in punishment policymaking must certainly give due regard to the difficult, values-infused nature of such choices.
one undisclosed paint touch-up with a $4 million award would presumably strike most people as evidence that the BMW jury does not live in the same moral universe as most of the rest of us. A judge applying the “hard look” to such an explanation would instantly detect the jury’s irrational “clear error of judgment” and throw out its verdict.

By obscuring the facts that motivate punishment, current practice impedes detection of such errors. By way of illustration, now suppose that the BMW jury had enjoyed legal authority to punish extra-Alabama sales but that it nonetheless awarded the full $4 million just to punish a single failure to disclose to Dr. Gore. Under standard excessiveness review principles, a reviewing judge would have to determine whether she could make up any rationale minimally sufficient to support this $4 million award. She might conclude that the jury’s punishment was reasonable because one could reach this figure by multiplying a plausible estimate of BMW’s profits per nondisclosure ($4000) times an approximation of the number of nondisclosures (1000). Given such an eminently reasonable hypothetical rationale, the judge should uphold the award, but by doing so she would unintentionally affirm the jury’s absurd punishment policy choice. Of course, were the judge to obtain an explanatory verdict, she would immediately perceive the jury’s irrational punitive reaction.

Returning to the facts and law of the real BMW, the actual problem with the jury’s verdict was not irrationality or bad judgment, but rather that it was illegal because the jury had considered an irrelevant factor—it had sought to punish sales that took place outside Alabama and that had not affected Alabamans. A “hard look” at an explanatory verdict would have revealed this error and memorialized it in the record in short order. Unlike the $4-million-punishment-for-one-nondisclosure hypothetical just discussed, this error would not suggest that the jury was in any way incompetent, crazy, or “out-of-control.” Ideally, therefore, the court should have given the jury a chance to fix its verdict by asking it to reconsider its punitive damages award in light of clarifying instructions that it could not punish BMW for its conduct outside Alabama. It is, of course, impossible to know how the jury would have responded to this request; most likely, however, it would have returned an unassailably reasonable and legal punitive damages verdict of $56,000.

223. See supra note 216 and accompanying text.
224. Mildly ironic postscript: the legal system eventually reached something like this result. After the Supreme Court threw out the $2 million remitted award, the Alabama Supreme Court ordered
Punitive Damages and the “Hard Look”

The BMW case is unusual in the sense that the mathematics behind the punitive damages award provided more information than is typically available regarding the jury’s reasoning. Most of the time, a court will not be able to ignore such information because it will not have it in the first place. A particularly bad side effect of this ignorance is that it can cause courts to affirm awards that do not reflect jury punishment policy. When reviewing an unexplained punitive damages award for excessiveness, a court must indulge the (frequently incorrect) assumption that the jury properly understood its instructions.225 Simplifying somewhat, the reviewing judge must determine whether, in her view, a jury that properly understood the law reasonably could have reacted to the evidence in the case by awarding punitive damages in the amount selected. The hypothetical punitive reaction a judge must imagine for purposes of review may have no relation at all to the jury’s actual punitive reaction. Suppose, as in BMW, a jury makes a mistake of law that causes it to inflate its punitive damages award, i.e., if the jury had properly understood the law, it would have granted a smaller amount of damages given its level of outrage at the defendant’s conduct. The judge assumes the jury understood her instructions, however, and therefore will tend to assume that the jury’s punitive reaction was more severe than it really was. A court that affirms in such a situation effectively gives its stamp of approval to an award that does not reflect accurately the jury’s punitive reaction. Such a result makes hash of the notion that juries have a special “expert” role to play in determining punishment.

Texaco provides another excellent example of a case in which a hard-look approach to punitive damages might have radically changed the outcome. Recall that a Texas jury imposed a $3 billion punitive damages verdict against Texaco for tortiously interfering with Pennzoil’s contract to buy Getty Oil.226 During deliberations, jurors sought clarification whether Texaco could be held liable for the conduct of certain Getty Oil

225. See supra note 129 (discussing results of study testing mock juror comprehension of instructions on punitive liability).

representatives.\textsuperscript{227} Texaco, in that particular case, was liable only for its own conduct. Unfortunately, its counsel, incorrectly believing that the instructions already informed the jury of this legal proposition, did not immediately press for a supplemental instruction.\textsuperscript{228} That evening, Texaco’s counsel realized that the instruction at issue had been in Texaco’s \textit{proposed} instructions, but that the judge had not included it in the final charge.\textsuperscript{229} The next morning, he requested that the court issue a supplemental instruction that “a party is only responsible for the actions of its own employees, agents or representatives acting within the scope of their employment.”\textsuperscript{230} Although the jury had not yet returned a verdict, the judge denied this request.\textsuperscript{231} Post-verdict juror interviews and statements revealed that the desire to punish Getty Oil’s representatives strongly influenced the jury’s decision to inflict a \$3 billion punitive damages award against Texaco.\textsuperscript{232}

The jury considered an “irrelevant factor” when it determined Texaco’s punishment. That punishment was illegal. Had the court asked the jury to explain the basis for its punitive damages award and then invoked hard-look review, it could have easily detected the jury’s error and then given it a chance to reconsider its decision in light of clarifying instructions. That this approach to the business of determining punishment would be preferable to playing “gotcha” with jury instructions should be self-evident.

The Supreme Court’s recent decision in \textit{Cooper Industries, Inc. v. Leatherman Tool Group, Inc.},\textsuperscript{233} provides yet another example of how a jury might go astray in determining punitive damages.\textsuperscript{234} In this case, the plaintiff had sued the defendant for trademark infringement and false advertising in connection with its manufacture and promotion of a multi-
Punitive Damages and the “Hard Look”

function pocket tool.235 On the infringement claim, the district court incorrectly instructed the jury that the defendant’s deliberate copying of the plaintiff’s product had been wrongful.236 The jury found that the defendant had infringed, but awarded no damages on that claim. The jury also found that the defendant had engaged in false advertising and, for this claim, awarded $50,000 in compensatory and $4.5 million in punitive damages.237 The Supreme Court noted that, in light of the district court’s incorrect instruction that copying was “wrongful,” the jury’s punitive damages award for the false advertising claim might “have been influenced by an intent to deter Cooper from engaging in . . . copying in the future.”238 In other words, the Supreme Court speculated that the incorrect instruction on one claim might have “leaked” over into the jury’s punitive damages determination for another claim. To repeat a now familiar refrain—a “hard look” approach would remove the need for such speculation.

Another potential mistake that springs to mind: Some states now instruct juries to factor into their punitive damages determinations the effect of other sanctions the defendant has already incurred for the same course of conduct.239 The rough idea is that juries should subtract from their awards the value of punishment dollars the defendant has already paid elsewhere. A defendant might reasonably fear, however, that a jury would instead use information concerning other punishments as evidence of the blameworthiness of the defendant’s conduct and as a benchmark for further punishments (i.e., a jury might conclude, “that other jury determined the defendant deserved $10 million in punishment, that sounds good to us—fine the defendant another $10 million”).240

235. Id. at 1679–81.
236. Id. at 1688.
237. Id. at 1680.
238. Id. at 1688.
239. See, e.g., MINN. STAT. ANN. § 549.20, Subd. 3 (West 2000) (instructing juries to consider “the total effect of other punishment likely to be imposed upon the defendant as a result of the misconduct, including compensatory and punitive damage awards to the plaintiff and other similarly situated persons, and the severity of any criminal penalty to which the defendant may be subject”).
240. See Juzwin v. Amtorg Trading Corp., 705 F. Supp. 1053, 1056 (D.N.J. 1989), judgment vacated on reh’g, 718 F. Supp. 1233 (D.N.J. 1989) (noting that a defendant might well be reluctant to present “such prejudicial evidence” to a jury); Mogin, supra note 5, at 214–15 (“Evidence that another jury, court, or agency has sanctioned the defendant is at least as likely to lead a jury to increase its award as it is to lead it to reduce the award.”); cf. supra note 67 and accompanying text (discussing “anchoring” effects).
Of course, these last two examples are speculative, but, along with BMW and Texaco, they illustrate a general point: jury “punishment policy” determinations are multi-faceted, fact-sensitive, value-laden, and intuitive. There are many wrong turns a jury might take in performing the difficult task of turning malicious torts into dollars. Moreover, jury instructions are often confusing, and even the most verbose cannot address every twist and turn deliberations might take. It therefore seems inevitable that some punitive damages awards juries grant will be tainted by consideration of improper factors or will amount to “clear errors” in judgment. If these sanctions take the form of unexplained numbers, then these errors are more likely to go undetected. “Hard look” review of explanatory punitive damages verdicts could catch and correct at least some such errors.

C. Show Juries the Respect of Asking Them To “Speak” Clearly

As the Lord Chief Justice remarked back in 1763 in Wilkes, one of the purposes of a punitive damages award is to serve as “proof of the detestation of the jury to the [defendant’s] action itself.”241 In more modern vernacular, as frequently observed by courts and perhaps even more frequently by plaintiffs’ attorneys, punitive damages are supposed to “send a message” to malicious tortfeasors that society will not tolerate their misbehavior. Rendering punitive damages verdicts as unexplained dollar figures makes such condemnatory messages needlessly inarticulate.242

The “message” behind a given award is a function of the purpose behind it—misread the rationale and misread the message. To draw yet another lesson from BMW, one might infer from the $4 million punitive damages award that the jury was absolutely infuriated by BMW’s nondisclosure of one minor paint repair and had concluded that just retribution for this one transgression required a punishment 1000 times greater than Dr. Gore’s $4000 compensatory award. The attendant message, then, was to shout to the world the jury’s contempt for and fury at BMW’s “fraud.” Were this description accurate, one might conclude

with some justification that the jurors were, in a word, crazy, which seems to be the view of some critics.\textsuperscript{243}

Of course, in all likelihood, the BMW jury’s primary motivation was to deter fraud. Understood in this light, the jury’s analysis and “message” were eminently rational. It determined that BMW had made about $4 million in illegal profits from its nondisclosure practice and decided to take this profit away. The jury’s attendant message was simply to tell BMW and other firms that they cannot keep ill-gotten gains. That the award was tainted by a legal error was not the jury’s fault, but rather that of a judicial system that fails to provide adequate help.

On a closely related point, in addition to clarifying condemnatory messages, asking jurors to explain their punishment choices would show them a measure of respect by treating them as thoughtful persons who might have something to say worth memorializing in official judicial records. The infamous “hot coffee” case, \textit{Liebeck v. McDonald’s Restaurants, P.T.S., Inc.},\textsuperscript{244} like BMW, provides an example of a case in which a jury might have benefited from such respect. In that case, Liebeck, an octogenarian, sued McDonald’s for selling her scalding coffee at a drive-through window, which she spilled on herself, causing third-degree burns.\textsuperscript{245} A New Mexico jury awarded her $160,000 in compensatory damages and $2.7 million in punitive damages,\textsuperscript{246} which the trial judge later remitted to $480,000.\textsuperscript{247}

The \textit{Liebeck} jury became another prime exhibit of the “out-of-control” jury.\textsuperscript{248} To determine whether that jury was “out-of-control,” however, one ought to know how it justified its award to itself. Post-trial

\begin{footnotesize}
\begin{itemize}
  \item \textsuperscript{243} See Thornburgh, supra note 212, at A47.
  \item \textsuperscript{244} No. CV-93-02419, 1995 WL 360309 (D.N.M. Aug. 18, 1994).
  \item \textsuperscript{246} \textit{Id.}
  \item \textsuperscript{248} See, e.g., Gerlin, supra note 245, at A1:
\end{itemize}
\end{footnotesize}
interviews indicate that, if asked to do so, the jury might have issued an explanatory verdict something like:

Before hearing the evidence, we would never have dreamed of awarding millions of dollars for serving coffee too hot. But at trial we learned that because of the temperature of McDonald’s coffee, Ms. Liebeck suffered third degree burns when she spilled McDonald’s coffee on herself that she bought at a drive-through window. The burns were so severe she had to spend seven days in the hospital and have skin grafts done.

We also learned that McDonald’s keeps its coffee much hotter than most other restaurants. It has gotten at least 700 complaints about coffee burns over the last ten years; it has paid some people more than $500,000 because of their burns. A McDonald’s employee testified that even though they knew of the risks of such burns, they hadn’t consulted burn experts on the danger and they had no plans to turn down the heat. McDonald’s just doesn’t care that it burns people with its coffee.

We also heard that McDonald’s sells about $1.35 million worth of coffee per day. To punish McDonald’s and make it pay attention to this burn problem, we decided to take away two days’ worth of these sales, $2.7 million.249

Whatever else might be said about such an explanation, it is not patently crazy. The jury’s “message” was not that one spill of a drink that is supposed to be hot in the first place should equal millions of dollars in punishment. Rather, the message was something like: Where a company engages in a pattern of conduct that harms hundreds of people, it can expect a punishment that will get its attention. On this view, the only arguably “arbitrary” part of the jury’s decision was its policymaking linkage between the facts it found and the decision to fine McDonald’s two days worth of coffee sales—why not one day or three days? But, given the intuitive nature of the problem of converting tort into dollars, a certain amount of this kind of “arbitrariness” in this policymaking step is inevitable absent wholesale reform of punitive damages (e.g., by providing that punitive damages will be determined by simply trebling compensatory damages, as in antitrust law).

249. See generally id. (discussing evidence that the jury heard at the McDonald’s hot-coffee trial).
Punitive Damages and the “Hard Look”

This is not to suggest that the jury’s punitive damages award was unassailably “correct.” It is, however, to acknowledge that the jury might have had something sensible to say regarding McDonald’s conduct. Had the jury been asked to state that message clearly as part of its verdict, it might not have been subjected to so much ridicule. Or, more realistically, perhaps an explanatory verdict would not have changed public reaction at all; people with an agenda to follow likely would have ignored the jury’s explanation in their eagerness to condemn it. In either event, however, the judicial system that asked this jury to perform the difficult task of determining punishment ought to have accorded it the respect of asking for its “side of the story” in some official way that would have been memorialized in court records and easily found by those interested in learning it.

Some juries might actually want this chance. Recall that in the Karadzic case, the jury foreman made a brief post-verdict statement for the record explaining why the jury had awarded $3.9 billion in punitive damages against a man whom it knew would likely never pay a cent to anyone in response to an American judgment.250 It was apparently at least somewhat important to the group of ordinary citizens asked to “punish” Dr. Karadzic that the court record some of its thoughts as well as its dollar figure.251

Punitive damages verdicts are supposed to express jury reactions to defendant conduct. They will serve this function better if juries use words as well as dollars to speak their minds. Asking for explanatory verdicts would also show juries a measure of respect by indicating that our civil litigation system regards them as capable of providing explanations worth reading and preserving.

D. Explanatory Verdicts and the Fearless Search for Information

A big understatement: Granting juries the discretionary power to impose punitive damages is one of the more controversial aspects of our civil litigation system. Critics of the current regime often argue that juries, given their lack of experience and training, are incapable of

250. See supra note 4 and accompanying text.
251. Cf. Mark Curridan, Power of Twelve, A.B.A.J., Aug. 2001, at 36 (reporting anecdotal evidence that frustrated juries are more frequently volunteering explanations for their verdicts); see also Brodin, supra note 150, at 69 (observing that “jurors themselves have on occasion spontaneously provided explanation and elaboration of their decision[s] in obvious frustration over the confines of the general verdict”; collecting and discussing cases).
making reasonable punishment determinations and often inflict absurd, “out-of-control” verdicts that create a terrible drag on the economy and stifle innovation while offering little or no offsetting social benefit.252 Proponents counter that punitive damages are a rarely invoked but vital means for juries to punish and deter terrible corporate misconduct.253 Moreover, civil juries are especially appropriate bodies for determining awards because they can speak to community standards of decency and justice.254

This debate is ideologically charged and carries high economic stakes; it may therefore be somewhat immune from factual resolution. That said, researchers have over recent years unearthed increasing amounts of interesting information regarding, inter alia: the frequency with which juries and judges award punitive damages,255 the amounts of median and mean punitive awards,256 and the types of torts most likely to trigger them.257 Scholars have studied the ability of jurors to comprehend punitive liability instructions;258 whether communal outrage standards exist,259 whether jurors can “map” their outrage onto an unbounded dollar

252. See, e.g., Browning-Ferris Indus. v. Kelco Disposal, Inc., 492 U.S. 257, 282 (1989) (opining that “[a]wards of punitive damages are skyrocketing” and are preventing the marketing of new prescription drugs, airplanes, and motor vehicles) (O’Connor, J., concurring in part and dissenting in part); Theodore B. Olson, The Parasitic Destruction of America’s Civil Justice System, 47 SMU L. Rev. 359, 359 (1994) (suggesting that runaway civil damages awards have helped transform the civil justice system in this country into an insatiable, almost-invisible “giant underground fungus”).


254. See, e.g., Angela P. Harris, Rereading Punitive Damages: Beyond the Public/Private Distinction, 40 ALA. L. REV. 1079, 1113 (1989) (discussing the communitarian perspective that “the jury, as the repository of shared communal values, is already in possession of the necessary yardstick” for determining punitive damages awards).

255. See, e.g., supra note 8 (citing BOJS BULLETIN and Eisenberg studies).

256. See e.g., supra note 17 (citing BOJS BULLETIN and Eisenberg studies).

257. See, e.g., Eisenberg, supra note 8, at 645 (reporting frequency of punitive damages awards by tort category).

258. See supra note 129 (discussing Hastie study reporting abysmal levels of mock jury comprehension of punitive liability instructions excerpted from those used in Exxon Valdez matter).

259. See supra note 65 and accompanying text (discussing Sunstein study reporting that, when asked to rank outrageousness of various injury-causing acts, study participants tended to respond similarly).
scale in a consistent manner; and the statistical relationships between punitive damages awards and their underlying compensatory damages awards.

What is still missing from these efforts is systematic reporting of how real juries would justify real punitive damages verdicts. It might prove to be the case that juries are perfectly capable of providing rough summaries of their punitive damages deliberations that show that they consider relevant facts in a reasonable manner to determine punishment (or that they do at least as good a job as judges in this regard). Such a result would tend to diminish concerns that juries are poorly suited for the job of determining punishment in civil litigation. Or, explanatory verdicts might suggest that juries are woefully inept at this task. Either result would add useful data to the punitive damages debate.

V. DISSOLVING SOME OBJECTIONS: FRAMING EXPECTATIONS AND RESPECTING JURIES

This Article proposes that courts take a “hard look” approach to “explanatory” punitive damages verdicts. Of potential objections to this proposal, three spring to mind as apt for brief preemptive response: (a) asking juries, untrained in the law, to explain their verdicts asks for the impossible and will not clarify much of anything anyway; (b) “hard look” review would show insufficient deference to juries; and (c) asking for explanatory verdicts is really just a ploy to make life easier for defendants. None of these objections justifies refusing to try the proposal.

260. See supra note 66 and accompanying text (reporting that same study participants who had ranked injury-causing acts in similar order of outrageousness, when asked to select appropriate financial punishment for these acts, gave highly variable and unpredictable responses).

261. Eisenberg, supra note 8, at 647–49 (reporting robust statistical relationship between logarithms of punitive damages awards and logarithms of underlying compensatory damages awards).

262. Cf. Sunstein et al., supra note 61, at 2111–12 (suggesting that it is unlikely that real juries actually base punitive damages awards on deterrence and instead probably focus on retribution).

263. Requiring juries to meet the “judicial” standard for punitive damages explanations would not necessarily hold them to all that high a standard. Remittitur decisions are often extremely perfunctory and conclusory. See, e.g., Dunn v. HOVIC, 1 F.3d 1371, 1391 (3d Cir. 1993) (en banc) (providing perfunctory two-paragraph explanation for remittitur to $1 million of $2 million punitive damages award that district court had previously remitted down from jury’s award of $25 million); BMW of N. Am., Inc. v. Gore, 646 So. 2d 619, 628–29 (Ala. 1994) (engaging in conclusory “comparative analysis” of jury’s punitive damages award of $4 million; ordering remittitur to $2 million with no real explanation of why court chose that figure), rev’d, 517 U.S. 559 (1996).
to see how it works. The first two, however, should affect how courts apply the “hard look” approach and what they can expect of it.

A. Explanatory Verdicts Need Not Ask for the Impossible

Whether juries are, in general, up to the task of explaining the grounds for their punitive damages verdicts depends on how we frame our expectations concerning what constitutes an adequate explanation. In this regard, it is useful to bear in mind the steps a jury must follow to determine the amount of an award. It must: (1) find the who-did-what-to-whom facts of the case; (2) determine which of these facts it should focus upon given its instructions on the law; and then (3) take the policymaking step of connecting these facts to a specific dollar amount.264 Presumably, most juries should be more than capable of roughly describing which facts they deemed material to punishment; doing so would “explain” the first two steps of the jury’s deliberative process.

The problem, if any, comes in “explaining” the final policymaking step. In certain respects, this step may be irreducibly intuitive, particularly as it relates to retribution. One of the axioms of the current punitive damages regime is that it is a good thing for jurors to consult their personal values concerning fair punishment; juries should rely on their “punitive reactions” to connect the facts of a case to a given dollar amount. This punitive reaction may not be something easy to put into words; it may just seem to be the case to a jury that, given Exxon’s conduct in connection with the Exxon Valdez oil spill and its effects, $5 billion is the right-sized sanction. Judicial review of explanatory verdicts would need to respect the limitations of language and the intuitive dimensions of punishment selection—a verdict should not be considered defective for failure to put into words what cannot usefully be said.

So stipulate the explanatory punitive damages verdicts would only amount to very rough and partial reports of jury reasoning. Recognition of this fact would not rob the “hard look” approach of meaning. Sometimes, notwithstanding the relatively unconstraining nature of punitive damages law, the jury just gets it really wrong. As demonstrated by analysis of BMW and Texaco above, rough explanatory verdicts could likely reveal (and memorialize for purposes of post-trial motions and appeal) instances in which juries base punishment on improper factors—

264. See supra notes 54–60 and accompanying text.
Punitive Damages and the “Hard Look”

e.g., the BMW jury’s punishment of extra-Alabama conduct and the Texaco jury’s punishment of Texaco for the conduct of Getty Oil representatives.265

Also, even the roughest explanatory verdicts would better equip courts to determine the general nature of the jury punitive reactions that connect the facts of cases to verdicts. To state the obvious, the reasonability of the BMW jury’s punitive reaction depended in part on whether it intended its $4 million punitive damages award to punish the one “fraudulent” sale to Dr. Gore or the (roughly) 1000 sales that happened nationwide. If the jury intended the former, then its award amounted to a “clear error of judgment” that should have been thrown out on that ground. If the latter, then the jury’s verdict arguably displayed excellent judgment that was unfortunately tainted by legal error that was not the jury’s fault. Explanatory verdicts would take some of the guesswork out of searching for “clear errors of judgment.”

B. The “Hard Look” Need Not Improperly Diminish Deference to Juries

A closely related objection is that it would be inappropriate to import administrative law concepts into judicial review of jury verdicts because doing so would tend to lessen deference to juries. The essence of the proposed “hard look” approach, however, is simply to provide better information to courts. Nothing inherent in this proposal requires courts to show juries less than their “usual” level of judicial deference as they scrutinize such information for legal error or irrationality.

Again, to determine a punitive damages award, a jury should take three steps: it must find the “facts” of the case; determine which of these facts are legally relevant given its instructions; and then make the policy choice of selecting punishment. With regard to the first step, nothing in the instant proposal requires or suggests that courts show anything other than their usual level of deference to jury fact-finding.

At the second step, a “hard look” would check whether a jury has based punishment on legally relevant factors. As a practical matter, this sort of check would tighten control on juries, but it would do so in a perfectly legitimate way. Juries currently have no legal right to base punishment on the “wrong” kinds of facts—that they sometimes do so is

265. See supra notes 223–24, 232 and accompanying text.
a function not of legitimate deference but rather of courts’ lack of information.

Courts review the rationality of how juries take the third step of making punishment policy whenever they review for excessiveness. The problem is that, under current practice, courts conduct their rationality review without the benefit of jury explanations for their decisions. The proposed approach would supply this information and allow courts to engage in informed excessiveness review that determines whether a jury’s award was minimally rational in light of the facts it actually deemed material to punishment. Of course, “minimal rationality” is hardly a bright-line test; scholars have noted that the aggressiveness with which courts conduct rationality review of agency action varies from court to court and agency to agency.266 The key point for present purposes once again, however, is that nothing inherent in the instant proposal would require courts to change the “quantum” of rationality they expect from juries. Instead, the proposal would simply give courts information they have always needed to properly check that verdicts survive a rationality test that they were always supposed to pass.267 In short, asking juries to explain their punitive damages verdicts need not and should not cause courts to improperly interfere with jury decisionmaking prerogatives.

C. Explanatory Verdicts Would Not Unfairly Tip the Balance in Favor of Defendants

The great virtue of the punitive damages verdict as an unexplained number is that it provides the clarity of ignorance—we cannot challenge what we do not know. By providing information concerning deliberations, explanatory verdicts would create avenues for attacking some punitive damages awards, which might lead some to conclude that the hard-look proposal has a pro-defendant edge.

266. See supra note 201 and accompanying text.

267. Moreover, those concerned that the instant proposal would lessen deference to juries should bear in mind that remittitur rates (and scattered case law) suggest that review of punitive damages awards is frequently quite aggressive under the current regime. See supra note 87 (discussing remittitur rates); cf. supra notes 73, 88–90 and accompanying text (citing cases calling for aggressive review of jury punitive damages awards). By contrast, at least as a matter of judicial rhetoric, court review of agency policy decisions is supposed to be highly deferential. Motor Vehicle Mfrs. Ass’n v. State Farm Mut. Auto. Ins. Co., 463 U.S. 29, 43 (1983) (“The scope of review under the ‘arbitrary and capricious’ standard is narrow and a court is not to substitute its judgment for that of the agency.”).
On one level, the answer to this objection is: good. It is supposed to be hard to use the machinery of the state to punish. Contrast the procedures for criminal punishment with those used in civil litigation seeking punitive damages. Generally speaking, for criminal punishment, a legislature must proscribe conduct; a prosecutor must decide that the defendant’s conduct merits investigation and prosecution; a judge or jury must determine guilt beyond reasonable doubt; and a judge must then determine sentence, perhaps subject to punishment guidelines promulgated by an administrative body.\(^{268}\) In many jurisdictions, to win punitive damages, a private plaintiff need only persuade a jury that a preponderance of the evidence shows that the defendant committed a “malicious” tort worthy of some sort of financial sanction in light of various punishment-related factors.\(^{269}\) Prima facie, to add another level of control to this process and make it more transparent would seem a good thing.

Moreover, it is not at all clear that explanatory verdicts would only benefit defendants. Judges might find some jury explanations sufficiently persuasive to affect remittitur decisions in plaintiffs’ favor. For instance, in the McDonald’s hot-coffee case, the judge remitted the jury’s $2.7 million punitive damages award to $480,000.\(^{270}\) Post-trial interviews indicated that the jury might have had a pretty good explanation for its award, and its apparent punitive reaction—that McDonald’s should pay two days’ worth of its coffee sales to ensure that the verdict got its attention—had a certain plausibility.\(^{271}\) Perhaps an explanatory verdict in that case would have persuaded the judge to deny the defendant’s remittitur request.\(^{272}\)

\(^{268}\) See, e.g., Richard W. Murphy, Superbifurcation: Making Room for State Prosecution in the Punitive Damages Process, 76 N.C. L. REV. 463, 534–36 (1998) (observing that punishment of crime provides an archetypal example of the liberty-protecting function of separation-of-powers as it requires cooperative action by legislatures, prosecutors, courts, and juries); see also Dorsey D. Ellis, Jr., Punitive Damages, Due Process, and the Jury, 40 ALA. L. REV. 975, 991–95 (1989) (noting that criminal conviction requires proof beyond a reasonable doubt whereas most jurisdictions require only a preponderance of the evidence for punitive damages liability).

\(^{269}\) See generally supra note 52 and accompanying text.


\(^{271}\) See supra text accompanying note 249.

\(^{272}\) See Galanter & Luban, supra note 242, at 1439–40 (arguing that requiring juries to explain their punitive damages awards would not constitute an “antiplaintiff rule” because “the price of unexplained exercises of discretion in jury awards is often an unaccountable decision by the judge to remit”).
VI. CONCLUSION

Civil juries in most American jurisdictions possess vast discretion to inflict punitive damages awards to punish defendants who have committed “malicious” torts. This power has been controversial since the common law first recognized it over 200 years ago, and courts continue to struggle to find coherent procedures and standards to control it. They have frequently stated that they review punitive damages verdicts for improper jury “passion and prejudice,” but a court cannot review an unexplained number for deliberative error.

The only thing standing between courts and a better, more transparent process is inertia. Given a slightly creative but legally permissible approach to Federal Rule of Civil Procedure 49 (and its state analogues), courts could require juries to answer open-ended interrogatories that ask them to explain the bases of their punitive damages awards. Taking a page from administrative law, courts could subject the “explanatory” verdicts they thus obtain to a form of hard-look review. In appropriate cases, where an explanatory verdict indicated that an award was tainted by correctable error, a court might ask the jury to reconsider in light of clarifying instructions.

Adopting these procedures would help rationalize judicial review by giving courts information they need to ensure that juries base punishment on the kinds of facts the law deems relevant to the task. It would also enable courts to ensure that the factual findings upon which juries actually base punishment can reasonably support the punitive damages awards they inflict, i.e., that juries make “reasonable” punishment policy choices. In addition to improving judicial review, asking juries to speak with words as well as dollars would help them send clearer condemnatory “messages” and, more than incidentally, would show them respect. Lastly, explanatory verdicts could shed light on the controversial question of whether juries should exercise the power to inflict punitive damages.
THE ASYMMETRY OF STATE SOVEREIGN IMMUNITY

Richard H. Seamon*

Abstract: This Article discusses whether a State has sovereign immunity from claims for just compensation. The Article concludes that the States are indeed immune from just-compensation suits brought against them in federal court; States are not necessarily immune, however, from just-compensation suits brought against them in their own courts of general jurisdiction. Thus, the States’ immunity in federal court is not symmetrical to the States’ immunity in their own courts. This asymmetry, the Article explains, is the result of the Due Process Clause of the Fourteenth Amendment. The Due Process Clause obligates a State to provide a means of paying just compensation every time the State takes private property for public use. A State may be able to meet this obligation by establishing a non-judicial compensation system. If a State fails to establish an adequate non-judicial compensation system, however, the State’s remedial obligation under the Due Process Clause falls upon the State’s courts. This resolution respects both a State’s constitutional right to avoid private lawsuits and an individual’s constitutional right to just compensation.

The principles of sovereign immunity and just compensation are on a collision course. The principle of sovereign immunity bars you from suing an unconsenting State for money.1 The principle of just compensation requires the State to pay you money if it takes your property for public use.2 The question is: Can you sue the State for

---

1. See, e.g., Alden v. Maine, 527 U.S. 706, 745–49 (1999). A State can consent to suits against it, thereby waiving its sovereign immunity, and some States have done so for claims seeking just compensation. See id. at 755 (stating that “sovereign immunity bars suits only in the absence of consent”). As discussed infra notes 249–59 and accompanying text (Part V.A), however, in many States the waiver of sovereign immunity from just-compensation claims is unclear or incomplete. See, e.g., DANIEL R. MANDELKER ET AL., FEDERAL LAND USE LAW § 4A.02[5][d] (1998) (observing that “in many states the availability of a compensation remedy in land use cases is not clear”). The issue discussed in this Article is therefore one that can arise in many States. Moreover, the analysis proposed in this Article has implications outside the context of just-compensation claims against unconsenting States. See infra notes 260–394 and accompanying text (Parts V.B through V.D).

2. The Fifth Amendment to the U.S. Constitution says in relevant part, “[N]or shall private property be taken for public use, without just compensation.” U.S. CONST. amend. V. This
money if it takes your property for public use without paying you? Surprisingly, and contrary to the belief of many commentators, the United States Supreme Court has never answered that question.3

Two principles of symmetry suggest that you cannot sue an unconsenting State for just compensation. First, the Court’s case law establishes a symmetry between the States’ immunity from lawsuits brought in federal court and their immunity from lawsuits brought in their own courts.4 Because the Court’s case law also strongly suggests that States are immune from just-compensation suits brought in federal court,5 this symmetry suggests that they are likewise immune from such suits in their own courts. In addition, the Court’s case law suggests that the immunity of the States is symmetrical to the immunity of the United States, and that the United States would be immune from just-compensation claims (had it not waived its immunity from such claims in the Tucker Act).6 The symmetry between the States’ sovereign immunity and that of the United States is thus a second symmetry suggesting that States are immune from just-compensation suits. Significantly, both of these symmetries—(1) between the States’ immunity in federal court and their immunity in their own courts and (2) between the States’ immunity and that of the United States—underlay the Court’s recent decision in Alden v. Maine,8 which held that States are immune in their own courts, as well as in federal court, from private actions based on Article I statutes.9 The Court could plausibly rely on the same symmetries to hold that States are immune from just-compensation claims in both their own courts and federal court.

This Article argues for a partly contrary conclusion: Unconsenting States may sometimes be sued for just compensation in their own courts, though not in federal court. That conclusion rests on the thesis that the prohibition against certain takings of property without just compensation applies to the States under the Fourteenth Amendment. See, e.g., Dolan v. City of Tigard, 512 U.S. 374, 383–84 & n.5 (1994).

3. See infra note 19 (citing commentary reading Supreme Court precedent to hold that Just Compensation Clause overrides sovereign immunity).
4. See infra notes 112–14 and accompanying text (Part III.B).
5. See infra notes 73–91 and accompanying text (Part II.B and II.C).
7. See infra notes 115–39 and accompanying text (Part III.C).
9. See id. at 712 (describing Seminole Tribe v. Florida, 517 U.S. 44 (1996), as making clear that “Congress lacks power under Article I to abrogate the States’ sovereign immunity from suits commenced or prosecuted in the federal courts,” and holding in the case before it that “the powers delegated to Congress under Article I . . . do not include the power to subject nonconsenting States to private suits for damages in state courts”).
States’ immunity in federal court and their immunity in their own courts are not symmetrical. The States’ federal-court and state-court immunities are not symmetrical because of the Due Process Clause of the Fourteenth Amendment. The procedural component of the Due Process Clause imposes on the States remedial obligations that sometimes require the involvement of the States’ own courts. In particular, the Due Process Clause requires that, when a State takes private property for public use, the State must have a “reasonable, certain and adequate” procedure for paying just compensation. A State may be able to meet this remedial obligation without involving its courts. For example, a State may be able to create a procedure under which property owners seek just compensation from the State’s executive or legislative branch. If a State does not provide an adequate, alternative compensation procedure, however, the Due Process Clause would require the State to let itself be sued for just compensation in its own courts of general jurisdiction. In short, the Due Process Clause could subject an unconsenting State in its own courts to suits from which it would be immune in federal court.

This asymmetry reflects the differing roles of the state and federal courts with respect to a State’s due process obligations. When due process obligates a State to provide a remedy, the State’s courts can enable the State to meet that obligation by providing the remedy. The federal courts cannot serve that function for the States. Although a federal court can often remedy a State’s violation of the Due Process Clause, in doing so the federal court is not discharging the State’s obligation to provide a remedy through its own tribunals. Thus, even if a federal-court remedy is available for a State’s violation of the Due Process Clause, that does not excuse the State’s failure to make its own remedy available. By the same token, the Due Process Clause can require that a remedy be available in state court, even though that remedy would not be available in federal court because of sovereign immunity. In sum, because the federal courts play a different role from that of a State’s own courts in meeting the State’s due process obligations, the States’ immersion in federal court is not symmetrical.

10. See infra notes 241–48 and accompanying text (Part IV.D).
11. U.S. CONST. amend. XIV, § 1 (“[N]or shall any State deprive any person of life, liberty, or property, without due process of law.”).
13. See, e.g., Williamson County Reg’l Planning Comm’n v. Hamilton Bank, 473 U.S. 172, 194 (1985); see also supra note 181 and accompanying text.
sovereign immunity in federal court and in their own courts can be asymmetrical.

Although the analysis offered here concerns just-compensation claims brought against unconsenting States under the Due Process Clause of the Fourteenth Amendment, the analysis has broader implications. Those implications flow from this Article’s thesis that it is the procedural component of the Due Process Clause in combination with the Just Compensation Clause—and not, as some commentators have assumed, the Just Compensation Clause alone—that overrides sovereign immunity in some circumstances. The implications of this Article’s thesis are threefold.

First, a State has remedial obligations under the Due Process Clause not only when it takes private property for public use but also when it causes other deprivations of people’s life, liberty, or property. Those remedial obligations sometimes entail awards of retroactive monetary relief from the state treasury, such as for the State’s erroneous collection of taxes. Thus, the Due Process Clause can create asymmetries between a State’s immunity in its own courts and its immunity in federal court besides the asymmetry related to just-compensation claims.

Second, the federal government has its own remedial obligations under the Due Process Clause of the Fifth Amendment. Under the analysis proposed here, for example, the federal government has an obligation, notwithstanding its sovereign immunity, to provide a procedure to pay just compensation when it takes private property for public use. Thus, Congress could not repeal the existing Tucker Act remedy for governmental takings unless an adequate alternative remedy were available.

Finally, the analysis proposed here clarifies two aspects of Congress’s power to enforce the Due Process Clause against the States under Section 5 of the Fourteenth Amendment. First, it clarifies Congress’s power to enforce the remedial obligations imposed on States by the doctrine of procedural due process. The analysis shows, specifically, that Congress cannot necessarily enforce those remedial obligations by regulating state conduct that gives rise to those remedial obligations but that does not itself violate the Fourteenth Amendment. This is why the Court struck down a federal statute in Florida Prepaid Postsecondary Education Expense Board v. College Savings Bank. The statute at issue there

15. See U.S. Const. amend. XIV, § 5 (“The Congress shall have power to enforce, by appropriate legislation, the provisions of this article [i.e., this Amendment].”).

regulated the States’ infringement of patents, conduct that could give rise to remedial obligations under the Due Process Clause but that did not “by itself violate the Constitution.” Second, Congress’s power to enforce the Due Process Clause of the Fourteenth Amendment does not depend on whether or not remedies are available in federal court for the States’ violations of the Fourteenth Amendment. As this Article shows, the availability of federal-court remedies does not excuse a State’s failure to meet its own remedial obligations under the procedural component of the Due Process Clause. Nor can the availability of federal-court remedies undo state conduct that violates the substantive components of the Due Process Clause.

This Article proceeds in five parts. Part I shows that, contrary to the view of many commentators, it is an open question whether the just-compensation principle overrides the principle of state sovereign immunity. Part II shows that, although that question is open, Supreme Court precedent virtually compels the conclusion that States are indeed immune from just-compensation suits brought in federal court. Part III discusses precedent suggesting that States are likewise immune from just-compensation suits brought in their own courts. Part IV argues that, despite the precedent discussed in Part III, the Due Process Clause sometimes subjects unconsenting States to just-compensation suits in their own courts. Finally, Part V discusses the implications of Part IV’s due-process analysis.

17. Id. at 643.

18. I am aware of only one other commentator who has suggested that the Due Process Clause overrides state sovereign immunity from just-compensation claims, and she did so only in passing and based on the assumption, rejected in Alden v. Maine, 527 U.S. 706, 754 (1999), that the States’ immunity in their own courts is a common-law, rather than a constitutional, doctrine. See Roberta Rosenthal Kwall, Governmental Use of Copyrighted Property: The Sovereign’s Prerogative, 67 TEX. L. REV. 685, 763 (1989) (stating that “[a] due process violation . . . occurs if plaintiffs are precluded from litigating whether a given state activity constitutes a taking”); id. at 763–64 (referring to “common-law doctrine of sovereign immunity” applicable in state courts). Other commentators have discussed, in varying levels of detail, the broader issue of the relationship between sovereign immunity and the Due Process Clause. See, e.g., Nicole A. Gordon & Douglas Gross, Justiciability of Federal Claims in State Court, 59 NOTRE DAME L. REV. 1145, 1151 n.22, 1171–74 (1984) (arguing primarily that Supremacy Clause requires state courts to adjudicate federal claims without regard to sovereign immunity, but observing that due process may impose similar requirement); Ellen D. Katz, State Judges, State Officers, and Federal Commands After Seminole Tribe and Printz, 1998 WIS. L. REV. 1465, 1493 (asserting that, given restrictions on federal-court jurisdiction on claims against States, principles of due process may create a special obligation on part of state courts to adjudicate federal claims against the State); see generally Carlos Manuel Vázquez, Sovereign Immunity, Due Process, and the Alden Trilogy, 109 Yale L.J. 1927 (2000) [hereinafter Vázquez, Alden Trilogy] (discussing Court’s federalism decisions from October 1998 Term); Louise Weinberg, The Power of Congress over Courts in Nonfederal Cases, 1995 B.Y.U. L.
I. THE UNITED STATES SUPREME COURT HAS NOT DECIDED WHETHER SOVEREIGN IMMUNITY BARS CLAIMS AGAINST UNCONSENTING STATES FOR JUST COMPENSATION

You would think that the U.S. Supreme Court would have decided by now whether a State can be sued for just compensation without its consent. Indeed, many commentators think that the Court did decide that issue in *First English Evangelical Lutheran Church v. County of Los Angeles*.19 I respectfully submit that those commentators are mistaken.20
As a plurality of the Court recently recognized, First English does not resolve whether “the sovereign immunity rationale retains its vitality” for claims seeking just compensation under the Fourteenth Amendment.\(^{21}\)

\(\text{A. The Holdings of First English}\)

In First English, the Court decided two things. First, the Court held that, when a land-use regulation has “taken” private property, it is not enough for a state court to invalidate the regulation prospectively.\(^{22}\) The Constitution also entitles the landowner to just compensation for the

\(^{20}\) See Boise Cascade Corp. v. Or. State Bd. of Forestry, 991 P.2d 563, 568 (Or. Ct. App. 1999) (observing that First English “did not squarely present” this issue and admitting that Oregon Court of Appeals’ resolution of the issue in that case “is not beyond dispute”), cert. denied, 121 S. Ct. 1363 (2001); Robert Brauneis, The First Constitutional Tort: The Remedial Revolution in Nineteenth-Century State Just Compensation Law, 52 Vand. L. Rev. 57, 137–38 (1999) (observing that Supreme Court “has never held that the [Just Compensation] Clause abrogates either federal or state sovereign immunity”).


\(^{22}\) See First English, 482 U.S. at 319 (holding that “[i]nvalidation” of ordinance that caused a regulatory taking, “though converting the taking into a ‘temporary’ one, is not a sufficient remedy to meet the demands of the Just Compensation Clause”).
period during which the regulation was in effect and caused a taking. This holding recognized so-called “temporary regulatory takings.”

Second, the Court held that, when a temporary regulatory taking occurs, the Just Compensation Clause, of its own force, gives rise to a monetary cause of action in “inverse condemnation” that is enforceable in state court. As the Court put it, the Just Compensation Clause is “self-executing.” Thus, the plaintiff in First English could sue for just compensation in state court even though there was no state statute or state case law recognizing a cause of action for temporary regulatory takings. It was “not necessary” for state law to create the cause of action, because it was created by the United States Constitution.

Neither of the holdings in First English concerned sovereign immunity. The first holding resolved a remedial question: What remedy does the Constitution require for a temporary regulatory taking? That question can arise whether or not the governmental defendant has sovereign immunity. Indeed, First English illustrates the point, as the defendants in that case were units of local government, which the Court has long held do not share the State’s immunity.

---

23. See id. at 321 (“We . . . hold that where the government’s activities have already worked a taking of all use of property, no subsequent action by the government can relieve it of the duty to provide compensation for the period during which the taking was effective.”).

24. See id. at 313 (“We now turn to the question whether the Just Compensation Clause requires the government to pay for ‘temporary’ regulatory takings.”).

25. See id. at 315–16.

26. Id. at 315 (quoting United States v. Clarke, 445 U.S. 253, 257 (1980)); see also Kirby Forest Indus., Inc. v. United States, 467 U.S. 1, 5 n.6 (1984) (noting that, when federal government takes possession of private property without bringing a condemnation proceeding, property owner’s right to bring inverse-condemnation suit “derives from the self-executing character of the constitutional provision with respect to condemnation”) (internal quotation marks omitted).

27. The state court in First English had dismissed the claim for a regulatory taking on the ground that California law did not create a cause of action to support the claim. See First English, 482 U.S. at 308–09; see also Agins v. Tiburon, 598 P.2d 25, 28–31 (Cal. 1979) (holding, apparently as a matter of state law, that property owner had no cause of action in inverse condemnation for unduly burdensome land-use restriction), aff’d on other grounds, 447 U.S. 255 (1980).

28. First English, 482 U.S. at 315 (quoting Jacobs v. United States, 290 U.S. 13, 16 (1933)).

29. See id. at 311, 313 (describing this as a “remedial question”).

30. See id. at 308 (stating that complaint named county and county flood control district as defendants); see also Bd. of Trs. v. Garrett, 531 U.S. 356, 356, 121 S. Ct. 955, 965 (2001) (“[T]he Eleventh Amendment does not extend its immunity to units of local government.”); Alden v. Maine, 527 U.S. 706, 756 (1999) (treating as well-settled the principle that cities, counties, and other units of local government are not entitled to sovereign immunity, and implying that this principle applies to actions brought in state court as well as in federal court).
from the Constitution when a temporary regulatory taking occurs. The question whether the plaintiff has a cause of action, however, differs from the question whether the defendant has sovereign immunity. The failure to distinguish between these two questions is what seems to have led to the mistaken view among some commentators that First English decided the immunity issue.

Nonetheless, the Court has repeatedly distinguished the issue whether the plaintiff has stated a cause of action from the issue whether the defendant has sovereign immunity from that cause of action. Perhaps the most famous case differentiating the issues is Larson v. Domestic & Foreign Commerce Corp. In Larson, the Court held that sovereign immunity barred a tort suit against a federal officer. The Court rejected the plaintiff’s contention that the officer’s commission of a tort precluded the defense of sovereign immunity. That contention, the Court explained, “confuses the doctrine of sovereign immunity with the requirement that a plaintiff state a cause of action.” Larson is only one of many cases in which the Court has said that the issue whether the plaintiff has stated a cause of action and the issue whether the defendant has sovereign immunity from that cause of action are “analytically distinct.”

---

31. See First English, 482 U.S. at 315 (“We have recognized that a landowner is entitled to bring an action in inverse condemnation as a result of the self-executing character of the constitutional provision with respect to compensation.”) (internal quotation marks and citations omitted).
32. 337 U.S. 682 (1949).
33. Id. at 703.
34. Id. at 691–93.
35. Id. at 692–93.
gives rise to a cause of action for monetary relief, but First English did not address whether States are immune from that cause of action.37

B. Sovereign Immunity in First English

The Court in First English did mention sovereign immunity. The Court did so, however, only to paraphrase an argument by the U.S. Solicitor General as amicus curiae.38 The relevant passage is in footnote nine of the opinion and, including the Court’s response, is as follows:

The Solicitor General urges that the prohibitory nature of the Fifth Amendment, combined with principles of sovereign immunity, establishes that the Amendment itself is only a limitation on the power of the Government to act, not a remedial provision. The cases cited in the text, we think, refute the argument of the United States that “the Constitution does not, of its own force, furnish a basis for a court to award money damages against the government.” Brief for United States as Amicus Curiae 14. Though arising in various factual and jurisdictional settings, these cases make clear that it is the Constitution that dictates the remedy for interference with property rights amounting to a taking.39 This passage makes clear that the Solicitor General was not directly arguing that sovereign immunity barred just-compensation claims.

officials and issue of whether they were immune from that cause of action under doctrine of official immunity).

37. Professor Fallon has pointed out to me that the distinction between the existence of a monetary cause of action and the existence of sovereign immunity is pretty thin in the context of the Fifth Amendment, where the only defendant against which a cause of action for just compensation originally could have run was the United States, a defendant otherwise capable of asserting sovereign immunity. See E-mail from Richard H. Fallon, Professor of Law, Harvard Law School, to Richard H. Seamon, Assistant Professor of Law, Univ. of S.C. (Jan. 11, 2001) (on file with author). I agree with that observation but would add that the Court in First English did not seem to rely on the original understanding of the Fifth Amendment in concluding that it creates a monetary cause of action. See William Michael Treanor, The Original Understanding of the Takings Clause and the Political Process, 95 COLUM. L. REV. 782, 782 (1995) (asserting that, under “original understanding” of Just Compensation Clause, compensation was not required “when government regulations limited the ways in which property could be used”). This circumstance, coupled with the fact that the defendants in First English were units of local government that could not claim sovereign immunity, convinces me that the Court in First English did not intend to address sovereign immunity. See First English Evangelical Lutheran Church v. County of Los Angeles, 482 U.S. 304, 308 (1987).

38. See First English, 482 U.S. at 316 n.9.

39. Id. (internal citations omitted).
Instead, he was arguing that, in light of the prohibitory wording of the Just Compensation Clause and general principles of sovereign immunity, the Clause should be interpreted only to prospectively nullify government action that has caused an uncompensated taking of private property for public use, and not to create a cause of action for retroactive monetary relief. Thus, the Solicitor General’s argument was that the Just Compensation Clause functioned as a limit on governmental power rather than as the source of a monetary cause of action. Accordingly, the Court did not mention sovereign immunity in response to the Solicitor General’s argument. Rather, the Court reiterated the point made elsewhere in its opinion that the Clause itself creates a private cause of action for a monetary remedy but does not otherwise “limit the governmental interference with property rights.”

As Professor Brauneis has accordingly observed, footnote nine of First English hints—but it only hints, without firmly deciding—that the just-compensation principle overrides the sovereign-immunity principle.

C. Sovereign Immunity in Del Monte Dunes

In City of Monterey v. Del Monte Dunes at Monterey, Ltd., a plurality of the Court again hinted that the just compensation requirement may override sovereign immunity. At the same time, the plurality confirmed that the issue is open.

In Del Monte Dunes, the Court held that the plaintiff had a right to a jury trial in a federal-court action asserting an inverse-condemnation claim against a city under 42 U.S.C. § 1983. To justify that conclusion, the majority had to explain why a right to a jury trial exists in an inverse-condemnation action under Section 1983 even though there is no right to a jury trial when the government brings an action to condemn property.

40. See id.; see also Brief for the United States as Amicus Curiae Supporting Appellee at 16–17, First English (No. 85-1199) (discussing sovereign immunity to support interpreting Just Compensation Clause as not creating “self-effectuating damage remedy”).

41. First English, 482 U.S. at 315.

42. See Brauneis, supra note 20, at 138 (stating that “the reference to sovereign immunity in First English that some have taken to be an oblique hint about abrogation may be explicable on other grounds”).


44. Id. at 720–21.

45. See id. at 711–18. But see Grant, supra note 19, at 146, 149 (arguing that the Court’s precedent holding that there is no right to jury trial in condemnation proceeding is “manifestly wrong” and that the same right applies in inverse condemnation).
A plurality of the Court (the majority minus Justice Scalia) offered as one difference between inverse-condemnation actions and direct-condemnation actions that, in inverse-condemnation actions, the government often disputes that any “taking” has occurred for which it is liable, whereas in direct-condemnation actions the government virtually never disputes its liability for a taking. The dissent replied that this “absence-of-liability” rationale had not actually been used in the cases denying the right to a jury in direct-condemnation actions. Instead, the dissent observed, courts in those cases often relied on a sovereign-immunity rationale; they reasoned that the government’s power to claim total immunity from suits for just compensation included the lesser power to allow such suits without providing a jury. In responding to the dissent, the plurality cited footnote nine of First English:

> Even if the sovereign immunity rationale retains its vitality in cases where the [Fourteenth] Amendment is applicable, cf. First English, 482 U.S. at 316 n.9, it is neither limited to nor coextensive with takings claims. Rather, it would apply to all constitutional suits against the Federal Government or the States, but not to constitutional suits such as this one against municipalities like the city of Monterey.

The plurality seems to be saying that the old, “sovereign immunity” rationale does not provide a workable standard for determining when a right to a jury trial exists. Of greater importance for our purposes, the plurality suggests doubt about whether sovereign immunity protects States from just-compensation claims based on the Fourteenth Amendment. At the same time, the plurality clearly believed the issue

46. See id. at 712 (plurality opinion) (finding “[m]ost important” difference between direct-condemnation proceeding and property owner’s action to redress uncompensated taking under § 1983 to be that, “when the government initiates condemnation proceedings, it concedes the landowner’s right to receive just compensation . . . . Liability simply is not an issue”). Justice Scalia did not join this part of the plurality opinion because he believed that “all § 1983 actions must be treated alike insofar as the Seventh Amendment right to jury trial is concerned.” Id. at 723 (Scalia, J., concurring in part and concurring in the judgment). He therefore did not think that it was appropriate to focus on the particular features of inverse-condemnation actions brought under Section 1983 to determine whether there was a right to jury trial in those actions.

47. Id. at 742 (Souter, J., dissenting).

48. Id.

49. Id. at 714 (plurality opinion).

50. This expression of doubt will be all the more intriguing to those who like to count votes on the Court because the doubt is voiced by Justice Kennedy, often a swing vote on issues of federalism. See Ernest A. Young, Aiden v. Maine and the Jurisprudence of Structure, 41 WM. & MARY L. REV. 1601, 1652 (2000). In Del Monte Dunes, Kennedy speaks for himself and three other Justices (Chief
to be open—a belief that it made clear when it asserted something to be so “[e]ven if . . . sovereign immunity . . . retains its vitality.” 51

D. Immunity vs. Just Compensation: A Conflict Not Yet Resolved

First English did not decide whether the just-compensation principle overrides the sovereign-immunity principle, and the plurality opinion in Del Monte Dunes confirms that the issue is unresolved. 52 In misreading First English as resolving the issue, commentators have ignored the difference between the issue whether a cause of action exists and the issue whether a particular defendant has sovereign immunity from that cause of action. The next Part of this Article demonstrates that, although the issue is indeed open, the Court’s case law strongly suggests that the Just Compensation Clause does not override the States’ immunity in federal court. The more difficult question of whether unconsenting States can be sued for just compensation in their own courts is addressed in

51. Del Monte Dunes, 526 U.S. at 714 (plurality opinion) (emphasis added). In a personal communication with the author, Professor Fallon characterized this passage from the plurality opinion in Del Monte Dunes as dictum. See Fallon E-mail, supra note 37. I am not entirely convinced of this characterization. In my view, the passage may be more accurately characterized as an expression of doubt made in the course of describing the rationale for the plurality’s conclusion. Cf. Seminole Tribe v. Florida, 517 U.S. 44, 67 (1996) (“When an opinion issues for the Court, it is not only the result but also those portions of the opinion necessary to that result by which we are bound.”). Although the plurality’s expression of doubt about sovereign immunity’s vitality in the face of the Fourteenth Amendment is fairly offhand, it seems significant that the remark is made in response to a remark from Justices who are usually hostile to sovereign immunity. This circumstance, in my view, prevents treating that expression of doubt as utterly casual.

52. Besides the suggestion of the plurality in Del Monte Dunes, the only other decision of the Court suggesting a position on the issue is Lake Country Estates, Inc. v. Tahoe Regional Planning Agency, 440 U.S. 391 (1979). That case involved a private action against an interstate compact in which the plaintiffs alleged, among other things, a taking of their property without just compensation. Id. at 394. The question before the Court was whether the interstate compact was “entitled to the immunity that the Eleventh Amendment provides to the compacting States themselves.” Id. at 393. It might be inferred from the Court’s decision that, if the compact did have sovereign immunity, the just-compensation claim would be barred by that immunity. That inference, of course, is contrary to the plurality’s suggestion in Del Monte that the Fourteenth Amendment overrides sovereign immunity. Lake Country is no more than suggestive, however, because the Court held that the compact did not have sovereign immunity from the just-compensation claim asserted there. See id. at 400–02.
Parts III and IV.

II. Although the issue is open, the Court’s case law virtually compels the conclusion that unconsenting states are immune from just-compensation suits brought in federal court.

As shown in Part I, the Court has never squarely decided whether States are immune from private actions for just compensation. As explained in this Part and as the lower federal courts have held, however, the Court’s decisions virtually compel the conclusion that States are immune from just-compensation actions when those actions are filed in federal court.53

Admittedly, the Court’s decisions stop short of producing an airtight conclusion, precisely because none involves a suit for just compensation directly against the State. As discussed below in Section A, the absence of such cases reflects that it took some time for the Court to decide that the Just Compensation Clause even applied to the States (via the Fourteenth Amendment). By the time the Court did so, the Court had also decided that neither a State nor its officials can be sued in federal court by a private plaintiff seeking money from the state treasury, even for a violation of the Fourteenth Amendment.

Although none of these decisions involved a federal-court action brought directly against a State for just compensation, the inescapable implication of those decisions is that such actions are barred by sovereign immunity. As discussed below in Section B, the Court has uniformly held that federal courts cannot entertain suits directly against a State or suits against state officers that seek money from the state treasury. None of these holdings suggest any exception for suits seeking just compensation.

A. The Court Has Never Squarely Addressed Whether Federal Courts Can Hear Suits for Just Compensation from the States

The Court has not had much chance to decide whether States are immune from private suits for just compensation. In 1833, the Court held that the Just Compensation Clause of the Fifth Amendment did not apply of its own force to the States; instead, the Clause applied only to the federal government. In 1877, the Court rejected the argument that the Just Compensation Clause applied to the States by virtue of its incorporation into the Due Process Clause of the Fourteenth Amendment. It was only in 1897 that the Court reversed its position on incorporation. In 1897, the Court held that the Just Compensation Clause was incorporated into the Due Process Clause of the Fourteenth Amendment. Before 1897, though, a suit against a State alleging a violation of the Just Compensation Clause failed to state a cause of action and was thus defective quite apart from any immunity problem.

By the time that the Court decided in 1897 that the Just Compensation Clause applies to the States, the Court had developed an “officer suit” doctrine that enabled takings claimants to get remedies against individual

---

55. See Davidson v. New Orleans, 96 U.S. 97, 105 (1877) (“If private property be taken for public uses without just compensation, it must be remembered that, when the fourteenth amendment was adopted, the provision on that subject, in immediate juxtaposition in the fifth amendment, with the one we are construing, was left out.”).
57. Id. at 241 (“[A] judgment of a state court . . . whereby private property is taken for the State or under its direction for public use, without compensation made or secured to the owner, is . . . wanting in the due process of law required by the Fourteenth Amendment . . . .”). The Court in Chicago, Burlington & Quincy Railroad did not expressly rely on an incorporation theory. See Dolan v. City of Tigard, 512 U.S. 374, 383–84 & n.5, 405–07 (1994) (Steven, J., dissenting) (arguing that Chicago, Burlington & Quincy Railroad is better understood as resting on doctrine of substantive due process); see also Richard C. Cortner, The Supreme Court and the Second Bill of Rights 24–29, 215 (1981) (noting that case did not rest on incorporation theory). Nonetheless, the Court has consistently understood its decision in that case as resting on an incorporation theory. See Dolan, 512 U.S. at 384 n.5; see also, e.g., Palazzolo v. Rhode Island, 121 S. Ct. 2448, 2457 (2001).
58. See Davidson, 96 U.S. at 105 (stating that State’s taking of private property for public use would not violate Due Process Clause); Barron, 32 U.S. at 251 (holding that Court lacked jurisdiction to review state-court decision rejecting taking claim against city because claim did not state a violation of Fifth Amendment). But cf. Ann Woolhandler & Michael G. Collins, Judicial Federalism and the Administrative States, 87 Cal. L. Rev. 613, 626–32 (1999) (discussing early cases in which Court applied “general constitutional principles” in diversity cases involving power of state and local governments to take and condemn property).
officers and thereby avoid sovereign immunity. A famous case illustrating the officer-suit doctrine is United States v. Lee, which was decided in 1882 in favor of the descendants of Robert E. Lee. In Lee, the Court upheld the eviction of federal officers from Arlington Cemetery because the land used for the cemetery had been taken from Lee without just compensation. In 1897, the Court relied on Lee to award similar injunctive relief against state officers who had occupied the plaintiff’s land without paying just compensation. At about the same time, the Court upheld an award of personal damages against an official who had caused an uncompensated taking. Thus, by 1897, a property owner whose property had been taken by a State without just compensation could sue the responsible state officials for an injunction preventing their continued “taking” of the property and for damages out of their own pockets. Because such an “officer suit” was not considered to be a suit against the sovereign, it was not barred by sovereign immunity.

Given the remedies available in an officer suit, a takings claimant would have been crazy to sue a State directly for just compensation in federal court. By 1897, the Court had repeatedly stated that a person could not sue a State directly (as distinguished from suing its officers) without the State’s consent. The Court had so held even for claims

60. 106 U.S. 196 (1882).
61. Id. at 223.
62. See Tindal v. Wesley, 167 U.S. 204, 212–23 (1897) (relying primarily on Lee to hold that sovereign immunity did not bar ejectment action against state officers who took plaintiff’s land without just compensation).
64. See supra notes 60–63 (citing cases); see also Scranton v. Wheeler, 179 U.S. 141, 151–53 (1900) (holding that sovereign immunity did not bar ejectment claim against federal officer in which claimant alleged a taking without just compensation).
65. See, e.g., United States v. Texas, 143 U.S. 621, 644 (1892) (stating that “the judicial power of the United States does not extend to suits of individuals against States”); Cunningham v. Macon & Brunswick R.R., 109 U.S. 446, 451 (1883) (“It may be accepted as a point of departure unquestioned, that neither a State nor the United States can be sued as defendant in any court in this country without their consent . . . .”); Railroad Co. v. Tennessee, 101 U.S. 337, 339 (1879) (“The principle is elementary that a State cannot be sued in its own courts without its consent.”); Beers v. Arkansas, 61 U.S. (20 How.) 527, 529 (1857) (“It is an established principle of jurisprudence in all civilized nations that the sovereign cannot be sued in its own courts, or in any other, without its consent and permission . . . .”); cf. Curran v. Arkansas, 56 U.S. (15 How.) 304 (1853) (holding that individuals could sue State for money in its own courts with State’s consent).
State Sovereign Immunity

alleging violations of the Fourteenth Amendment. In the 1890 case of
North Carolina v. Temple, for example, the Court held that sovereign
immunity barred a federal-court action against a State alleging that the
State had violated the Fourteenth Amendment by failing to make
payments on state bonds. Moreover, Temple was only one of many
cases in which sovereign immunity defeated constitutional claims against
States that had reneged on their bond obligations. This case law,
combined with the availability of relief against state officers, goes far to
explain why the Court never had a case that required it squarely to decide
whether the Just Compensation Clause overrode the States’ sovereign
immunity.

Nonetheless, the officer suit had a major shortcoming. You could sue
an officer for damages out of the officer’s own pocket and for an
injunction preventing the officer from violating federal law. You could
not sue an officer, however, to get money from the state treasury. That
was one type of relief as to which the Court would not indulge the fiction
that the suit was against the officer, rather than against the State. As the
Court said,

No suit . . . can be maintained against a public officer which
seeks to compel him to exercise the State’s power of taxation; or
to pay out its money in his possession on the State’s obligations,

66. 134 U.S. 22 (1890)
67. Id. at 25, 30 (holding that Eleventh Amendment barred federal-court action against State
alleging violations of Contract Clause and Fourteenth Amendment).
68. See, e.g., Hans v. Louisiana, 134 U.S. 1, 16 (1890) (holding that sovereign immunity barred
federal-court action by a citizen against his own State alleging that State’s failure to pay interest on
its bonds violated Contract Clause of U.S. Constitution). Many of these suits rested on the
suits were uniformly rejected to the extent that the relief sought would “expend itself on the public
treasury.” Land v. Dollar, 330 U.S. 731, 738 (1947); see also Seamon, Separation of Powers, supra
note 59, at 168–74.
69. In addition, it appears that just compensation became increasingly available in the state courts
during the 19th century. See Akhil Reed Amar, THE BILL OF RIGHTS: CREATION AND
RECONSTRUCTION 149–51 (1998) (discussing nineteenth-century case law recognizing right to just
compensation as natural right); Brauneis, supra note 20, at 85 n.114 (discussing gradual adoption of
state constitutional provisions guaranteeing just compensation); J.A.C. Grant, The “Higher Law”—
Background of the Law of Eminent Domain, 6 WIS. L. REV. 67, 70–71 (1931) (discussing evolution
of just-compensation concept in United States as judicial development based on theory of natural
rights).
70. See, e.g., Louisiana v. Jumel, 107 U.S. 711, 727–28 (1883). The Court in Jumel and other
cases after the Civil War repudiated its earlier “party of record” rule, under which sovereign
immunity never barred suits against government officers. See generally Seamon, Separation of
or to execute a contract, or to do any affirmative act which affects the State’s political or property rights. 71

Officer suits for that sort of relief were deemed to be “really” suits against the sovereign itself. 72

This limitation on officer suits clearly would have prevented someone from getting an injunction requiring an officer to pay just compensation from the state treasury. Perhaps because this was so clear, there appear to be no reported decisions by the United States Supreme Court in which a plaintiff sought that sort of relief.

B. The Court’s Sovereign Immunity Decisions Strongly Suggest that Federal Courts Cannot Hear Suits for Just Compensation from the States

Although none of the Court’s cases involved federal-court suits seeking just compensation from the government treasury, there are cases in which plaintiffs alleging violations of the Just Compensation Clause sought remedies different from those traditionally available in officer suits (i.e., damages out of the officer’s pocket and injunctions prohibiting the officer from continuing to take the plaintiff’s property). The Court’s rejection of nontraditional remedies in these cases demonstrates that sovereign immunity would have barred the use of an officer suit to get just compensation from a State.

One such case, Belknap v. Schild, 73 involved a patent-infringement claim against federal officers. The plaintiff held a patent for an improvement to caisson gates. 74 He sued the officers who had been using a supposedly infringing caisson gate at a federal navy yard. 75 The Court

---


72. See, e.g., Int’l Postal Supply Co. v. Bruce, 194 U.S. 601, 605 (1904) (patent-infringement action to enjoin official’s use of infringing government property “really was against the United States”); see also Ford Motor Co. v. Ind. Dep’t of Treasury, 323 U.S. 459, 464 (1945) (“And when the action is in essence one for the recovery of money from the state, the state is the real, substantial party in interest and is entitled to invoke its sovereign immunity from suit even though individual officials are nominal defendants.”).

73. 161 U.S. 10 (1896).

74. Id. at 11. A caisson is a watertight chamber used for underwater construction and repair of, e.g., ships. See MERRIAM WEBSTER’S COLLEGIATE DICTIONARY 160 (10th ed. 1993) (“caisson,” definition 2(a)).

75. See Belknap, 161 U.S. at 23 (observing that, according to government’s pleading, the allegedly infringing caisson gate “was made and used by the United States in a dry dock at a navy yard, and the defendants only operated and used it as officers, servants, and employe’s [sic] of the United States”).
recognized that the officers’ conduct violated the Just Compensation Clause, remarking that “the United States have no more right than any private person to use a patented invention without license of the patentee or making compensation to him.” Because of that violation, the Court held that the plaintiff was entitled to damages from the officers’ own pockets. The Court also held, however, that the plaintiff was not entitled to an injunction preventing federal officers from continuing to use the navy-yard caisson gate that allegedly infringed his patent. That injunctive relief was barred, the Court explained, because it sought to control government-owned property. The Court’s explanation made it clear that sovereign immunity would likewise bar an injunction tapping the government treasury: “[N]o injunction can be issued against officers of a State, to restrain or control the use of property already in the possession of the State, or money in its treasury when the suit is commenced.” This reference to the inaccessibility of “money in [the] treasury” shows that, just as the plaintiff in Belknap could not enjoin use of the infringing caisson gate, he could not have gotten compensation from the U.S. Treasury.

The Court reached a similar conclusion in a case involving a State. In Hopkins v. Clemson Agricultural College, the plaintiff claimed that a state college had taken his land without just compensation by building a dike that caused his land to flood. The Court held that the plaintiff was entitled to an award of damages that could be executed against property that the College owned in its own right, separate and apart from the State. The Court appeared to assume, however, that the award of damages could not be executed against state-owned property or funds.

76. Id. at 16.
77. Id. at 23 (holding that defendant officials could be “held liable to the patentee for their own infringement of his patent”); see also id. at 18 (“[O]fficers or agents, although acting under order of the United States, are . . . personally liable to be sued for their own infringement of a patent.”).
78. Id. at 23 (holding that lower court “erred in awarding an injunction against the defendants”).
79. Id. at 18–25.
80. Id. at 18 (emphasis added).
81. Id.; see also Int’l Postal Supply Co. v. Bruce, 194 U.S. 601, 605 (1904) (relying on Belknap to reject claim for injunctive relief in patent-infringement case against federal government).
82. 221 U.S. 636 (1911).
83. Id. at 638–40.
84. Id. at 648.
85. See id. (observing, in response to defendant’s argument that State held title to land underlying the college, that money judgment could be satisfied out of land that college itself owned in fee simple and out of its various sources of income).
Moreover, the Court held that the plaintiff was not entitled to an injunction requiring the College to destroy the dike.\textsuperscript{86} That relief was not warranted because the title to the dike, as well as the title to the land on which the dike was located, might be held by the State, not the College:

The state, therefore, may be a necessary party to any proceeding which seeks to affect the land itself, or to remove any structure thereon which has become a part of the land. If so, and unless it consents to be sued, the court cannot decree the removal of the embankment which forms a part of the State’s property.\textsuperscript{87}

The Court in \textit{Hopkins}, as in \textit{Belknap}, said that a court could not compel an officer to “pay out [the State’s] money . . . or to do any affirmative act which affects the State’s political or property rights.”\textsuperscript{88} Thus, the decision in \textit{Hopkins}, like that in \textit{Belknap}, makes clear that sovereign immunity would have barred an award of just compensation from the sovereign’s treasury.

\textbf{C. The Bottom Line: Hints but No Holdings that Permit Federal Courts To Require States To Pay Just Compensation}

\textit{Belknap} and \textit{Hopkins} date from about the turn of the century but are still good law. Throughout the twentieth century, the Court has held that sovereign immunity bars federal suits seeking money from the state treasury, including for violations of federal law.\textsuperscript{89} Except for footnote 9 of \textit{First English}, a majority of the Court has never suggested that an unconsenting State can be sued for just compensation in federal court.

Nonetheless, a plurality of the Court in \textit{Del Monte Dunes} doubted whether “the sovereign immunity rationale retains its vitality in cases where th[e] [Fourteenth] Amendment is applicable.”\textsuperscript{90} This could be read as a hint that the sovereign-immunity principle does not apply when a

\begin{itemize}
  \item \textsuperscript{86} Id. at 648–49.
  \item \textsuperscript{87} Id. at 649.
  \item \textsuperscript{88} Id. at 642 (emphasis added).
  \item \textsuperscript{89} See, e.g., Edelman v. Jordan, 415 U.S. 651, 665–78 (1974) (holding that relief in federal-court officer suit that would require payment of retroactive monetary relief from state treasury was barred by sovereign immunity); Ford Motor Co. v. Ind. Dep’t of Treasury, 323 U.S. 459, 462–70 (1945) (holding that sovereign immunity barred federal-court action against state agency and officials for refund of taxes allegedly collected in violation of Commerce Clause and Fourteenth Amendment); see also Dugan v. Rank, 372 U.S. 609, 620 (1963) (stating that suit is considered to be one against the sovereign if the judgment sought would expend itself on the public treasury).
  \item \textsuperscript{90} City of Monterey v. Del Monte Dunes at Monterey, Ltd., 526 U.S. 687, 714 (1999) (plurality opinion) (citing, with a “cf.” signal, First English Evangelical Lutheran Church v. County of Los Angeles, 482 U.S. 304, 316 n.9 (1987), discussed at text accompanying notes 19–42, supra).
\end{itemize}
claim is based on the Just Compensation Clause or some other provision made applicable to the States under the Fourteenth Amendment. To carry through on this hint while remaining faithful to the precedent discussed in this Part, the Court would need to conclude that a private action for just compensation against an unconsenting State has to be brought somewhere other than in a federal court, such as in the State’s own courts. That is precisely the result for which this Article contends. To reach that result, however, the Court would also have to grapple with case law discussed in the next Part, which suggests that the States’ immunity in their own courts is symmetrical to their immunity in federal court and to the immunity of the federal government.

III. SYMMETRY SUGGESTS THAT STATES ARE ALSO IMMUNE FROM JUST-COMPENSATION SUITS BROUGHT IN THEIR OWN COURTS

The thesis of this Article is that unconsenting States can sometimes be sued in their own courts for just compensation. Nonetheless, there is Supreme Court case law supporting a contrary conclusion, and that case law is discussed in this Part. The case law includes a case discussed above in Part II, Hopkins v. Clemson Agricultural College. Hopkins could be read as directly supporting state sovereign immunity from just-compensation claims in state courts. Other case law supplies indirect support by establishing two dimensions of symmetry in the law of sovereign immunity. The first symmetry is between the States’ sovereign immunity in federal court and the States’ sovereign immunity in their own courts. The second symmetry is between the sovereign immunity of the States and the sovereign immunity of the federal government. These two dimensions of symmetry are not only deeply rooted in the Court’s precedent; they were reaffirmed by the Court in Alden v. Maine. Nonetheless, the case law discussed in this Part does not foreclose this Article’s thesis that the Fourteenth Amendment sometimes overrides the States’ sovereign immunity from just-compensation claims.

91. See infra notes 178–248 and accompanying text (Part IV).
92. See infra notes 178–248 and accompanying text (Part IV).
93. 221 U.S. 636 (1911).
94. See infra notes 98–11 and accompanying text (Part III.A).
95. See infra notes 112–14 and accompanying text (Part III.B).
96. See infra notes 115–39 and accompanying text (Part III.C).
A. Hopkins v. Clemson Agricultural College

*Hopkins* was discussed above in Part II to show that States have sovereign immunity in federal courts even from claims based on the Just Compensation Clause (as incorporated into the Fourteenth Amendment). Now this Article discusses *Hopkins’* suggestion that this immunity from just-compensation claims extends even to a State’s own courts.

Dr. Hopkins filed his lawsuit against the Clemson Agricultural College in a South Carolina state court, alleging that a dike built by the College had taken his property (by flooding it) without just compensation. The case came to the U.S. Supreme Court after the South Carolina Supreme Court relied on sovereign immunity to affirm the dismissal of Hopkins’ lawsuit. The U.S. Supreme Court held that Hopkins was entitled to recover damages payable out of the College’s own property. The Court appeared to agree with the College, however, that an award of damages could not be executed against state funds or state-owned property. In addition, the Court held that Hopkins was not entitled to an injunction requiring the College to tear down the dike that had caused Hopkins’ land to flood, because that dike sat on state-owned land. The decision in *Hopkins* thus strongly suggests that, when a state agent takes private property for public use without just compensation, the State has sovereign immunity in its own courts from relief that would be paid out of the state treasury or that would affect state-owned property.

Indeed, the only reason that Dr. Hopkins was able to get damages from the College was because the Court concluded that the College did not
share the State’s immunity.105 The Court treated the College the same as an individual officer or a unit of local government, neither of which, the Court had previously held, possessed sovereign immunity from personal damages for tortious conduct.106 Dr. Hopkins might not have been so lucky today. In recent years, the lower federal courts have consistently held that state universities are “arms of the State” that share the State’s sovereign immunity.107

I believe that *Hopkins* can be read narrowly, for reasons that I will now explain, and that it *should* be read narrowly, for reasons discussed in Part IV.108 *Hopkins* can be read as a case in which the State had immunity in its own courts only because that immunity did not prevent the plaintiff from getting adequate relief in those courts for the taking of his property. Under the Court’s decision, Dr. Hopkins could recover damages from the College, in periodic state-court suits, as long as the dike continued to take his land, because the College did not share the State’s sovereign immunity.109 Furthermore, the Court suggested that Dr. Hopkins would not be forced to bring periodic suits against the College indefinitely. Rather, the forces of nature would take care of the problem. The dike that had caused the flooding of Hopkins’ land had already washed away once, and the Court observed that Hopkins might be able to have the College enjoined from “further acts looking to the maintenance or reconstruction of the d[j]ike.”110 Such injunctive relief against a future taking, coupled with compensation from the College for the prior taking, arguably provided adequate relief in state court.111 For these reasons, *Hopkins* can be read as a case in which sovereign immunity did not prevent an adequate state remedy for a governmental taking of property.

105. *See Hopkins*, 221 U.S. at 642–46 (framing the question as “whether a public corporation can avail itself of the State’s immunity from suit” and holding that, although college was an agent of the State, it did not share State’s immunity in action alleging unconstitutional and tortious conduct).

106. *See id.* at 643–44 (relaying on suits against public officials to reject College’s claim that it had sovereign immunity); *id.* at 645 (citing Lincoln County v. Luning, 133 U.S. 529, 530–31 (1890) (holding that County did not share State’s sovereign immunity)).


110. *Id.* at 649.

111. *Cf.* Nat’l Private Truck Council, Inc. v. Okla. Tax Comm’n, 515 U.S. 582, 591 n.6 (1995) (stating in dicta that, despite general rule that 42 U.S.C. § 1983 does not authorize declaratory or injunctive relief against illegal state taxes if state law provides adequate remedy, declaratory relief might be available if necessary to spare taxpayer from having to bring multiple suits).
So read, *Hopkins* would not support a State’s use of sovereign immunity to bar an adequate remedy for a taking.

**B. Symmetry Between the States’ Immunity in Federal Court and Their Immunity in Their Own Courts**

As discussed in Part II, Supreme Court case law virtually compels the conclusion that States cannot be sued for just compensation in federal court. (They cannot be sued directly, nor can their officers be sued for an injunction requiring the payment of just compensation from the state treasury. 112) Other case law indicates that the State’s immunity in federal court is symmetrical to their immunity in their own courts. 113 Together, these two lines of cases forcefully suggest that unconsenting States cannot be sued for just compensation in their own courts.

The case law indicating that the State’s federal-court immunity and their state-court immunity are symmetrical dates back to the early 19th century. The Court has consistently relied on the States’ immunity in federal court to determine the scope of their immunity in their own courts, and vice-versa. As discussed below in Section D, the Court did so again in its recent decision in *Alden v. Maine* 114

**C. Symmetry Between the Sovereign Immunity of States and the Sovereign Immunity of the United States**

The Court has not only treated the States’ immunity in federal court as being symmetrical to the States’ immunity in their own courts. In addition, the Court has often treated the States’ sovereign immunity as symmetrical to the United States’ sovereign immunity. 115 Because of this

112. *See supra* notes 53–91 and accompanying text (Part II).

113. I have discussed this case law in an earlier Article and will not repeat the discussion here. For a discussion of earlier case law, see Seamon, *Sovereign Immunity, supra* note 104, at 375–88.


115. *See, e.g.*, Fla. Dep’t of State v. Treasure Salvors, Inc., 458 U.S. 670, 686 n.21 (1982) (Stevens, J., joined by Burger, C.J., and Marshall and Blackmun, J.J.) (citing with apparent approval precedent treating state sovereign immunity similarly to federal sovereign immunity); *Great N. Life Ins. Co. v. Read*, 322 U.S. 47, 53 (1944) (“The principle of immunity from litigation assures the states and the nation from unanticipated intervention in the processes of government.”); *Philadelphia Co. v. Stimson*, 223 U.S. 605, 620 (1912) (stating that principle permitting state officers to be sued for injunctive relief against violation of federal law is “equally applicable” to federal officers); *Tindal v. Wesley*, 167 U.S. 204, 213 (1897) (stating that, in analyzing whether suit against state officer was barred by sovereign immunity, “the question whether a particular suit is one against the State, within the meaning of the Constitution, must depend upon the same principles that determine whether a particular suit is one against the United States”); *Cunningham v. Macon & Brunswick...*
apparent symmetry between state and federal sovereign immunity, an analysis of whether the States are immune from just-compensation claims must consider the United States’ immunity from such claims.

The Court’s early case law suggests that the United States is immune from just-compensation claims except to the extent that it has waived that immunity in legislation such as the Tucker Act.116 As Professor Brauneis has noted, the “classic cases” supporting that suggestion are Schillinger v. United States117 and Lynch v. United States.118 Professor Brauneis adds that those cases are “somewhat dated.”119 Indeed, in light of later decisions of the Court, Schillinger and Lynch do not foreclose this Article’s thesis that the Due Process Clause sometimes overrides sovereign immunity.120

In Schillinger, the Court held that the plaintiff could not sue the United States under the Tucker Act for infringing on the plaintiff’s patented method of laying concrete pavement.121 The plaintiff alleged that, by infringing on his patent, the government had taken his property for public use without just compensation.122 The Court determined that this infringement claim did not fall into any of the categories of claims as to which the Tucker Act waived the federal government’s sovereign immunity; it was therefore barred by sovereign immunity.123 That

R.R., 109 U.S. 446, 451 (1883) (“It may be accepted as a point of departure unquestioned, that neither a State nor the United States can be sued as defendant in any court in this country without their consent, except in the limited class of cases in which a State may be made a party in the Supreme Court of the United States by virtue of the original jurisdiction conferred on this court by the Constitution.”); see also Seamon, Separation of Powers, supra note 59, at 173–74 (citing and discussing cases in which “the Court did not distinguish the sovereign immunity of the states from that of the United States”). But cf., e.g., Chisolm v. Georgia, 2 U.S. (2 Dall.) 419, 478 (1793) (opinion of Jay, C.J.) (stating that federal courts’ reliance on executive branch for enforcement of its judgments “place[s] the case of a state, and the case of the United States, in very different points of view” when it comes to sovereign immunity).

117. 155 U.S. 163 (1894).
119. Brauneis, supra note 20, at 138 n.343.
120. See infra notes 126–127, 134, 138 and accompanying text.
121. Schillinger, 155 U.S. at 166–72.
122. See id. at 168.
123. Id. at 167–72.
determination rested primarily on the Court’s view that the alleged violation of the Just Compensation Clause did not, in and of itself, state a “claim[] founded upon the constitution” within the meaning of the Tucker Act. At this time, the Court did not construe the Just Compensation Clause as creating a cause of action except when the circumstances of the taking implied a promise by the government to pay just compensation. The Court in later cases concluded that the Clause itself creates a cause of action, without regard to whether the government impliedly promised to pay for the property that it had taken. Thus, the predicate for the Court’s sovereign-immunity ruling in Schillinger is no longer valid. Even so, the portion of Schillinger that has been superceded by later case law concerned the existence of a cause of action, not the existence of sovereign immunity. Because these are two

124. Id. at 167 (observing that, under then-recent amendment, Court of Claims could hear “[a]ll claims founded upon the constitution”); id. at 168 (rejecting plaintiff’s argument that official’s tortious appropriation of private property for public use “creates a claim founded upon the constitution”).

125. See ROBERT MELTZ ET AL., THE TAKINGS ISSUE 40 n.20 (1999) (noting that “[e]arly decisions hung the [Court of Claims'] jurisdiction on the ‘implied contract with the United States’ phrase in the Tucker Act,” but later cases “shifted to the statute’s ‘claim . . . founded . . . under the Constitution’ phrase’”); Brauneis, supra note 20, at 137 n.342 (discussing Court’s transition from the view that Court of Claims’ jurisdiction over just-compensation claim rested on “implied contract” to modern view that such claims arise directly under the Constitution); Developments in the Law: Remedies Against the United States and Its Officials, 70 HARV. L. REV. 827, 876 (1957) (“Before 1927 the Supreme Court . . . took the view that a ‘taking’ action could be maintained [under the Tucker Act] only if a promise to make compensation could be attributed to the Government.”).

126. See, e.g., Preseault v. Interstate Commerce Comm’n, 494 U.S. 1, 12 (1990) (“If there is a taking, the claim is ‘founded upon the Constitution’ and within the jurisdiction of the [Claims Court] to hear and determine.”) (adding bracketed text and quoting United States v. Causby, 328 U.S. 256, 267 (1946)). In United States v. Dickinson, 331 U.S. 745, 748 (1947), the Court said:

But whether the theory of these suits be that there was a taking under the Fifth Amendment, and that therefore the Tucker Act may be invoked because it is a claim founded upon the Constitution, or that there was an implied promise by the Government to pay for it, is immaterial. In either event, the claim traces back to the prohibition of the Fifth Amendment . . . . See also Jacobs v. United States, 290 U.S. 13, 16 (1933) (holding that lower court erred in considering just-compensation claim to be based on implied contract, and stating: “A promise to pay was not necessary . . . . The suits were thus founded upon the Constitution.”); Seaboard Air Line Ry. v. United States, 261 U.S. 299, 304 (1923) (stating that property owner’s right to recover just compensation from United States “does not depend on contract” and that “[a] promise to pay is not necessary”); see also First English Evangelical Lutheran Church v. County of Los Angeles, 482 U.S. 304, 315 (1987) (observing that “it has been established at least since Jacobs . . . . that claims for just compensation are grounded in the Constitution itself”).

127. See Developments in the Law, supra note 125, at 876 (commenting that Court decisions rejecting takings claims that could not be supported on implied-contract rationale seemed to rest on Court’s assumption that Just Compensation Clause itself did not create cause of action).
distinct issues (as discussed above). In Schillinger, still constitutes precedent suggesting that the federal government would be immune from just-compensation suits in the absence of its consent.

In Lynch, the Court held that a federal statute violated substantive due process by repudiating the government’s obligation to honor contracts of war insurance. The statute was a wartime, cost-saving measure; it purported to repeal earlier statutes that gave low-cost death and disability insurance to people who fought in World War I. The Court determined that the repeal of those statutes would breach the government’s contracts with the insured veterans and thereby unconstitutionally deprive them of a property right. In addition to holding the repealer statute unconstitutional, the Court held that the statute did not withdraw the federal government’s consent to suits to recover money under the insurance contracts that the statute sought to repeal. In other words, the Court construed the statute to prevent the United States from using sovereign immunity to defeat a constitutional challenge to its contract repudiation. This result, if anything, implies that the federal government cannot use its sovereign immunity to prevent an adequate remedy for its unlawful deprivation of property, including its

128. See supra notes 32–37 and accompanying text.

129. Other cases, like Schillinger, suggest that the United States is immune from just-compensation claims based on a rationale the continued validity of which is dubious. See, e.g., United States v. N. Am. Transp. & Trading Co., 253 U.S. 330, 335–36 (1920); Tempel v. United States, 248 U.S. 121, 130–31 (1918); Russell v. United States, 182 U.S. 516, 535 (1901); see also United States v. Sioux Nation, 448 U.S. 371, 414 (1980) (arguably suggesting that Congress must “waive[] the Government’s sovereign immunity” for “courts to resolve a taking claim on the merits”).

130. See Lynch v. United States, 292 U.S. 571, 580 (1934) (holding that “Congress was without power to reduce expenditures by abrogating contractual obligations of the United States” that arose under wartime insurance policies issued by the government).

131. See id. at 574–79 (describing statutes that created wartime insurance program).

132. Id. at 576 (holding that insurance policies “are contracts of the United States”); id. at 579 (observing that repeal of statutes creating those policies would, “if valid, abrogat[e]” those contracts and “relieve[] the United States from all liability on the contracts without making compensation to the beneficiaries”); id. (holding that insurance contracts were protected by Just Compensation Clause; that, in absence of power to annul them, “the due process clause prohibits the United States from annulling them”; and that no such power had been identified).

133. See id. at 575 (government argued that lower courts lacked jurisdiction over the suit “because the consent of the United States to be sued had been withdrawn” by same statute that purported to repeal prior statutes creating wartime insurance); id. at 586 (rejecting government’s argument, stating: “[I]t does not appear that Congress wished to deny the [judicial] remedy if the repeal of the contractual right was held void under the Fifth Amendment.”).
takings of private property for public use without just compensation. 134

The Lynch Court did say in dicta that Congress could have withdrawn consent to suit on the insurance contracts as long as it did not repudiate its obligation to honor the contracts. 135 In light of the holding in Lynch, this dicta should be read to mean only that the government could fulfill its contractual obligations “without the interposition of” the courts. 136 It should not be read to mean that the United States need not provide a way of honoring those obligations or its obligation to pay just compensation when it takes private property for public use. 137 As the Court recently said, the government can violate the Constitution “either by denying just compensation in fact or by refusing to provide procedures through which compensation may be sought.” 138 In light of that principle, the Lynch dicta suggests only that the Constitution does not require that compensation procedures take the form of judicial procedures. 139

134. See Reg’l Rail Reorg. Act Cases, 419 U.S. 102, 134–36 (1974) (citing Lynch to support interpreting a federal statute not to preclude suit for just compensation under Tucker Act); Battaglia v. Gen. Motors Corp., 169 F.2d 254, 257 (2d Cir. 1948) (citing Lynch, among other cases, for proposition that, “while Congress has the undoubted power to give, withhold, and restrict the jurisdiction of [federal] courts other than the Supreme Court, it must not so exercise that power as to deprive any person of life, liberty, or property without due process of law or to take private property without just compensation”) (footnote omitted), cert. denied, 335 U.S. 887 (1948).

135. See Lynch, 292 U.S. at 581 (“Although consent to sue was thus given when the [war insurance] policy issued, Congress retained power to withdraw the consent at any time.”).

136. Id. at 582 (“So long as the contractual obligation is recognized, Congress may direct its fulfillment without the interposition of either a court or an administrative tribunal.”); cf. Beers v. Arkansas, 61 U.S. (20 How.) 527, 528–30 (1857) (rejecting constitutional challenge to state statute creating additional requirements for suing State in state court on its bond obligations and suggesting that even repeal of state statute permitting such suits would not imply that State repudiated the contract obligations upon which such suits were based).

137. Such a reading of the Lynch dicta would be at odds with not only the holding in Lynch but also the Court’s later decision in De La Rama Steam Ship Co. v. United States, 344 U.S. 386 (1953). In De La Rama, the Court refused to construe a federal statute to take away a federal district court’s jurisdiction to hear a claim against the United States based on a government insurance policy. De La Rama 344 U.S. at 387–91. The Court cited Lynch and observed that the effect of such a construction would be to prevent the plaintiff from enforcing the government’s contractual obligations. Id. at 382, 388–90; see also Graham v. Goodell, 282 U.S. 409, 431 (1931) (stating in dicta that Congress could not withdraw consent to suits against itself, and preclude suits against its officers, to prevent paying tax refunds, if it lacked power to enact legislation eliminating the right to the refunds).


139. See infra notes 182–195 and accompanying text (arguing that States may be able to use nonjudicial procedures to meet their obligation under Due Process Clause to provide procedure for payment of just compensation).
D. Alden v. Maine

The two dimensions of symmetry that have been discussed—(1) between the States’ immunity in federal court and their immunity in their own courts; and (2) between the States’ immunity and that of the United States—underlay Alden v. Maine.\(^{140}\) Alden therefore reinforces an argument that (1) if States are immune from just-compensation claims in federal court and (2) if the United States would be immune from just-compensation claims in the absence of the Tucker Act, then States should be immune from just-compensation claims in their own courts.\(^{141}\) Like the other case law discussed in this Part of the Article, Alden does not compel the conclusion that States are invariably immune from just-compensation suits brought in their own courts.

In Alden, the Court addressed an issue that arose in the wake of the Court’s decision three years earlier in Seminole Tribe v. Florida.\(^{142}\) In Seminole Tribe, the Court held that Congress cannot use its Article I powers to authorize private actions for retroactive monetary relief to be brought against unconsenting States in federal court.\(^{143}\) This naturally led to the question whether Congress could use Article I to subject unconsenting States to such suits in their own courts.\(^{144}\) Addressing this question so soon after Seminole Tribe, the Court seems to have regarded

\(^{140}\) 527 U.S. 706 (1999). In the interest of full disclosure, I note that I wrote an amicus brief in Alden on behalf of a group of state and local organizations. See Brief of the National Conference of State Legislatures [et al.], Alden v. Maine (No. 98-436). The Court adopted one of the arguments in that amicus brief. Compare id. at 15 (arguing that Court’s anti-commandeering precedent prevents Congress from “using the Commerce Clause . . . to turn the State against itself” by “pitting the executive branch of Maine against its judicial branch”) with Alden, 527 U.S. at 749 (“A power to press a State’s own courts into federal service to coerce the other branches of the State . . . is the power first to turn the State against itself and ultimately to commandeering the entire political machinery of the State against its will . . . .”).

\(^{141}\) For reasons discussed above, decisions of the United States Supreme Court virtually compel the conclusion that unconsenting States are immune from just-compensation claims in federal court. See supra notes 53–91 and accompanying text (Part II). In contrast, it is unclear from the Court’s case law whether the United States would be immune from just-compensation claims in federal court in the absence of the waiver of immunity in the Tucker Act. See supra notes 116–39 and accompanying text (arguing that older Court decisions suggesting that federal government is immune from just-compensation claims do not resolve the issue); see also infra notes 329–43 and accompanying text (arguing that, in the absence of Tucker Act, Due Process Clause of Fifth Amendment would require United States to provide adequate procedures for paying just compensation).

\(^{142}\) 517 U.S. 44 (1996).

\(^{143}\) Id. at 72–73.

\(^{144}\) Alden, 527 U.S. at 741 (“Whether Congress has authority under Article I to abrogate a State’s immunity from suit in its own courts is . . . a question of first impression.”).
it as a question of symmetry: for private causes of action authorized by federal statutes enacted under Article I, is a State’s immunity in its own courts symmetrical to its immunity in federal court? That the issue was one of symmetry was reinforced by the facts of *Alden*. The plaintiffs in that case initially sued the State of Maine in federal court to recover overtime wages under the federal Fair Labor Standards Act. After that federal-court suit was dismissed based on *Seminole Tribe*, the plaintiffs brought an identical suit against Maine in Maine’s own courts. The Court held in *Alden* that this state-court action was likewise barred by sovereign immunity.

The majority opinion in *Alden* consisted of four parts. Part I demonstrated that the States’ sovereign immunity derives, not from the Eleventh Amendment, but from the original Constitution. Part II demonstrated that Congress could not override this immunity using its Article I powers. Part III discussed limits on the States’ immunity, including the States’ ability to waive their immunity. Part IV rejected the plaintiffs’ argument that Maine had waived its immunity from the claims in that case. In discussing the nature of the States’ immunity in their own courts (in Part I) and Congress’s power to abrogate that immunity (in Part II), the Court repeatedly linked the States’ immunity in federal court to their immunity in their own courts and also linked the States’ immunity to that of the federal government.

Specifically, the Court determined in Part I that the original

---

145. *See* Jackson, *Seductions, supra* note 19, at 698 (asserting that, “[h]aving decided . . . in *Seminole Tribe* that the Eleventh Amendment barred the states from being sued in federal courts without their consent on Article I causes of action, coherence arguments for the same rule in state courts apparently were very attractive to the Court [in *Alden*]”).


147. *Alden*, 527 U.S. at 712.

148. *Id.* at 712, 754.

149. *Id.* at 713 (concluding that state sovereign immunity “neither derives from, nor is limited by, the terms of the Eleventh Amendment” but is instead “a fundamental aspect of the sovereignty which the States enjoyed before the ratification of the Constitution, and which they retain today . . . except as altered by the plan of the Convention or certain constitutional Amendments”).

150. *Id.* at 730 (stating that state sovereign immunity “is not directly related to the scope of the judicial power established by Article III, but inheres in the system of federalism established by the Constitution” and is therefore applicable to Congress’s exercise of legislative powers).

151. *Id.* at 755 (stating that “certain limits are implicit in the constitutional principle of sovereign immunity,” one of which is that “sovereign immunity bars suits only in the absence of consent”).

152. *Id.* at 757–58 (concluding that Maine “has not consented to suit”).
State Sovereign Immunity

Constitution “preserve[d] the States’ traditional immunity from private suits.”153 By “traditional” immunity, the Court meant the States’ immunity in their own courts, which, of course, predated the Constitution.154 The Court reasoned that the immunity that States had enjoyed in their own courts was preserved in the federal courts for which the Constitution provided.155 This “preservation” rationale implies that, under the original Constitution, States would enjoy the same immunity in the newly authorized federal courts as they had had in their own courts—i.e., that the States’ immunities in the two fora would be symmetrical.156

The Court began Part II of its opinion by asserting that state sovereign immunity restricts not only the powers of the federal judiciary but also those of Congress.157 In support of this assertion, the Court cited decisions, including Seminole Tribe, holding that Congress cannot use Article I to abrogate state sovereign immunity in federal court.158 In the Court’s view, “[t]he logic of the decisions . . . does not turn on the forum in which the suits were prosecuted but extends to state-court suits as well.”159 The Court also cited decisions in which it “described the States’ immunity . . . without reference to whether the suit was prosecuted in state or federal court.”160 These decisions “suggest[ed]” that States enjoy in their own courts an immunity “analogous” to that enjoyed in federal court.161 The Court reasoned that, because Congress cannot use Article I
to abrogate the States’ federal-court immunity, it cannot use Article I to abrogate their state-court immunity, either.162

After discussing its case law on the States’ federal-court immunity and finding that case law supported “analogous” state-court immunity, the *Alden* Court in Part II of its opinion considered “whether a congressional power to subject nonconsenting States to private suits in their own courts is consistent with the structure of the Constitution.”163 The Court found that the constitutional structure accords States a dignity that is offended by private suits “regardless of the forum.”164 Ultimately, the Court could not believe that Congress could “require state courts to entertain federal suits which . . . could not be heard in federal courts.”165

In addition to linking the States’ state-court immunity to their federal-court immunity, the Court in *Alden* linked the States’ sovereign immunity to that of the United States. Part I of the opinion discussed the founding generation’s “universal” acceptance of the English doctrine “that a sovereign could not be sued without its consent” without suggesting that this doctrine differed depending on whether the “sovereign” was a State or the federal government.166 Part II even more explicitly linked the States’ immunity to that of the United States:

It is unquestioned that the Federal Government retains its own immunity from suit not only in state tribunals but also in its own courts. In light of our constitutional system recognizing the essential sovereignty of the States, we are reluctant to conclude that the States are not entitled to a reciprocal privilege.167 Elsewhere in its opinion, moreover, the *Alden* Court quoted precedent in which it had equated the sovereign immunity of States to that of the

---

162. *See, e.g.*, *id.* at 735 (finding in history of Eleventh Amendment the suggestion that "the States' sovereign immunity was understood to extend beyond state-law causes of action"); *see also* Dan Braveman, *Enforcement of Federal Rights Against States: Alden and Federalism Non-sense*, 49 AM. U. L. REV. 611, 646 (2000) (remarking that *Alden* "appears to adopt a rather curious 'symmetrical' view of federalism" under which "state courts cannot be required to hear FLSA claims against the states because federal courts cannot be required to do so").

163. *Alden*, 527 U.S. at 748.

164. *Id.* at 749.

165. *Id.* at 754; *see also* *id.* at 752 (observing that it would be "anomalous" for "the National Government [to] wield greater power in the state courts than in its own judicial instrumentalities").

166. *Id.* at 715; *see also* *id.* at 716–17 (quoting Alexander Hamilton’s statement in Federalist No. 81 that "[i]t is inherent in the nature of sovereignty not to be amenable to the suit of an individual without its consent").

167. *Id.* at 749–50; *see also* *id.* at 735 ("[S]urely the dissent does not believe that sovereign immunity poses no bar to a state-law suit against the United States in federal court.").
Thus, if the Court were to conclude that States are immune from just-compensation suits brought in federal court (as argued in Part II of this Article) and that the United States would likewise be immune from such suits (as some early case law suggests), Alden would in two ways support the further conclusion that States are likewise immune from such suits in their own courts. First, Alden indicates that the States’ immunity in their own courts is coextensive with their immunity in federal court. Second, Alden indicates that the States’ immunity corresponds to that of the United States.

Nonetheless, Alden does not foreclose the thesis of this Article, which is that the Due Process Clause of the Fourteenth Amendment can subject unconsenting States to suits for just compensation in their own courts. As mentioned, Part I of the Alden opinion traces the States’ sovereign immunity to the original Constitution; it thus does not prevent the conclusion that the Fourteenth Amendment altered that immunity. Similarly, Part II of the Alden opinion discussed only Congress’s power under Article I to override the States’ sovereign immunity; it does not address whether Congress would have such power under Section 5 of the

168. Id. at 745–46:

It may be accepted as a point of departure unquestioned, that neither a State nor the United States can be sued as defendant in any court in this country without their consent, except in the limited class of cases in which a State may be made a party in the Supreme Court of the United States by virtue of the original jurisdiction conferred on this court by the Constitution. (quoting Cunningham v. Macon & Brunswick R.R., 109 U.S. 446, 451 (1883)); id. at 750 (“The principle of immunity from litigation assures the states and the nation from unanticipated intervention in the processes of government.”) (quoting Great N. Life Ins. Co. v. Read, 322 U.S. 47, 53 (1944)).

169. See supra notes 53–91 and accompanying text (discussing case law strongly suggesting that States are immune from just-compensation suits brought in federal court).

170. See supra notes 116–39 and accompanying text (discussing case law suggesting that United States is immune from just-compensation suits).

171. The Court in Alden determined that in some ways suits against the States in their own courts are more threatening to sovereignty interests than are suits against the States in federal court. Alden, 527 U.S. at 749. This determination would support an argument that the States’ immunity from just-compensation claims in their own courts follows a fortiori from their federal-court immunity from such claims. That argument, however, would be hard to reconcile with case law indicating that States bear the initial responsibility for providing just compensation when they take private property for public use. See, e.g., Williamson County Reg’l Planning Comm’n v. Hamilton Bank, 473 U.S. 172, 194 (1985).

172. See supra notes 149 and 153–156 and accompanying text; see also Alden, 527 U.S. at 726–27 (referring to the “original understanding of the States’ constitutional immunity from suit” by the “founding generation”); id. at 730 (saying that state sovereign immunity “inheres in the system of federalism established by the Constitution”).
Fourth Amendment. Most importantly, the issue before the Court in *Alden* concerned a State’s immunity from a private action based on an Article I statute, and the Court’s holding was expressly limited to that situation.174

The Court in *Alden* did mention the Fourteenth Amendment in Part III of its opinion, where it discussed limits on state sovereign immunity. The Court explained that “[t]he States have consented . . . to some suits pursuant to the plan of the Convention or to subsequent constitutional Amendments.” 175 The only amendment that the Court cited was the Fourteenth Amendment, of which the Court said:

We have held also that in adopting the Fourteenth Amendment, the people required the States to surrender a portion of the sovereignty that had been preserved to them by the original Constitution, so that Congress may authorize private suits against nonconsenting States pursuant to its § 5 enforcement power.176

It may be significant that the Court referred in this passage only to Congress’s power to abrogate state sovereign immunity under Section 5 of the Fourteenth Amendment. Perhaps this implies that the Fourteenth Amendment does not abrogate that immunity of its own force. 177 If the

---

173. See U.S. CONST. amend. XIV, § 5 (“The Congress shall have power to enforce, by appropriate legislation, the provisions of this article [i.e., this Amendment].”); see also *Alden*, 527 U.S. at 730 (stating at outset of Part II of the opinion: “In this case we must determine whether Congress has the power, under Article I, to subject nonconsenting States to private suits in their own courts.”); *id.* at 741 (stating at outset of Part II.B of its opinion: “Whether Congress has authority under Article I to abrogate a State’s immunity from suit in its own courts is, then, a question of first impression.”).

174. See *Alden*, 527 U.S. at 712 (“We hold that the powers delegated to Congress under Article I of the United States Constitution do not include the power to subject nonconsenting States to private suits for damages in state courts.”); *id.* at 754 (“In light of history, practice, precedent, and the structure of the Constitution, we hold that the States retain immunity from private suit in their own courts, an immunity beyond the congressional power to abrogate by Article I legislation.”).

175. *Id.* at 755.

176. *Id.* at 756.

177. See Boise Cascade Corp. v. Or. State Bd. of Forestry, 991 P.2d 563, 567 (Or. Ct. App. 1999) (stating that *Alden’s* “emphasis on positive acts of Congress under section five” might suggest that Fourteenth Amendment does not, of its own force, subject unconscioning States to just-compensation claims in their own courts). Compare the description of the Fourteenth Amendment in the passage of *Alden* quoted in the text with the description of the Fourteenth Amendment in *Fitzpatrick v. Bitzer*, 427 U.S. 445, 453–54 (1976). In *Bitzer*, the Court described the substantive provisions, as well as the enforcement provision, of the Fourteenth Amendment as “contemplat[ing] limitations on [the States’] authority.” *Id.* at 453. The *Bitzer* Court said that the Fourteenth Amendment’s substantive provisions, in particular, “[l]imited . . . duties” on the States “with respect to their treatment of private individuals.” *Id.*; see also *id.* at 456 (observing that, quite apart from Congress’s power to
Court were to so hold, it would reject the thesis of this Article, which is that the Fourteenth Amendment does indeed, of its own force, sometimes abrogate state sovereign immunity.

In any event, like the other precedent discussed in this Part of this Article, Alden supports but does not compel the conclusion that States are immune from just-compensation suits in their own courts. It is therefore hoped that this Part has shown that the Court’s precedent leaves room for the argument advanced in the next Part of this Article.

IV. THE DUE PROCESS CLAUSE CREATES REMEDIAL OBLIGATIONS THAT SOMETIMES REQUIRE STATES TO SUFFER JUST-COMPENSATION SUITS IN THEIR OWN COURTS

The Due Process Clause of the Fourteenth Amendment puts remedial obligations on the States.\(^{178}\) Those obligations arise whenever a State deprives a person of life, liberty, or property.\(^ {179}\) One such deprivation occurs when a State takes private property for public use.\(^ {180}\) The Court has said that the State’s remedial obligation for this sort of deprivation is to have made, at the time of its taking of the property, “reasonable, certain and adequate provision for obtaining compensation.”\(^ {181}\) This Part

---

178. This Part of the Article analyzes the nature of those obligations for claims seeking just compensation. The implications of that analysis for other due process claims are discussed infra notes 260–328 and accompanying text (Part V.B).

179. See U.S. CONST. amend. XIV, § 1 (“Nor shall any State deprive any person of life, liberty, or property, without due process of law.”).

180. Hudson v. Palmer, 468 U.S. 517, 539 (1984) (O’Connor, J., concurring) (“The Due Process and Takings Clauses of the Fifth and Fourteenth Amendments stand directly in opposition to state action intended to deprive people of their legally protected property interests. . . . When adequate remedies are provided and followed, no uncompensated taking or deprivation of property without due process can result.”).

argues that a State may well be able to meet this obligation in many cases without allowing itself to be sued in its own courts; in such cases, the State can retain its immunity from suits for just compensation in those courts. If a State fails to create an adequate, alternative way to pay just compensation, however, the Due Process Clause requires the State’s courts of general jurisdiction to be open to just-compensation suits against it. Under these circumstances, the Due Process Clause overrides state sovereign immunity.

A. A State May Be Able To Provide “Reasonable, Certain and Adequate Provision” for Just Compensation Without Involving Its Courts

We should take the Court at its word when it says that a State can meet its due process obligation by creating a just-compensation scheme that is “reasonable, certain, and adequate.” As the Court has often said, due process is “flexible.” The States, in particular, have wide discretion in devising procedures for providing due process. That discretion is part of the broader discretion that the people of every State have to structure their state government however they like, within the limits of the U.S. Constitution. This latitude includes the State procedures by which


See supra note 181 (citing cases that state “reasonable, certain, and adequate” standard).

E.g., Zinermon v. Burch, 494 U.S. 113, 127 (1990) (“Due process, as this Court often has said, is a flexible concept that varies with the particular situation.”); Gilbert v. Homar, 520 U.S. 924, 930 (1997) (describing due process as “flexible”).

See, e.g., Gilbert, 520 U.S. at 930 (observing that State has broad discretion in providing due process in connection with disciplining its employees).

property owners can get just compensation when their property is taken for public use.186 All that the Fourteenth Amendment demands is that the State’s compensation procedure be: (1) reasonable—i.e., that it not entail undue delay, expense, or effort;187 (2) certain—i.e., that there be no doubt about its existence or its potential to lead to an award of just compensation;188 and (3) adequate—i.e., that it actually and reliably result in an award of just compensation when the State has taken private property for public use.189

If a State has a nonjudicial compensation scheme that meets these criteria, the Due Process Clause should not be construed categorically to compel the State to let itself be sued for compensation in its own courts.190 For example, a State could create a claims commission in its

---

186. See e.g., Madisonville Traction, 196 U.S. at 252 (“Speaking generally, it is for the State, primarily and exclusively, to declare for what local public purposes private property, within its limits, may be taken upon compensation to the owner, as well as to prescribe a mode in which it may be condemned and taken.”); Chi., Burlington & Quincy R.R. v. City of Chicago, 166 U.S. 226, 236 (1897) (“The legislature may prescribe a form of procedure to be observed in the taking of private property for public use, but it is not due process of law if provision be not made for compensation.”); see also San Diego Gas & Elec. Co. v. City of San Diego, 450 U.S. 621, 660 (1981) (Brennan, J., dissenting) (stating that “the Constitution does not embody any specific procedure or form of remedy that the States must adopt” in meeting their obligations under Just Compensation Clause). Justice Brennan’s dissent in San Diego Gas was later adopted by a majority of the Court to the extent that it argued in favor of government liability for temporary regulatory takings. See First English Evangelical Lutheran Church v. County of Los Angeles, 482 U.S. 304, 315, 316 n.9, 318 (1987) (discussing and citing with approval Justice Brennan’s San Diego Gas dissent); see also Nollan v. Cal. Coastal Comm’n, 483 U.S. 825, 866–67 (1987) (Stevens, J., dissenting) (observing that Court in First English endorsed Justice Brennan’s proposal in San Diego Gas dissent that government owes just compensation for temporary regulatory takings); Lucas v. S.C. Coastal Council, 505 U.S. 1003, 1018 (1992) (describing with approval Justice Brennan’s observation in San Diego Gas dissent that, from property owner’s perspective, government’s prohibition of all beneficial use of land was tantamount to physical appropriation).


189. Cf. Crane v. Hahlo, 258 U.S. 142, 147–48 (1922) (holding that due process was not violated by statute that restricted judicial review of administrative determination of damages caused to property owner by city’s improvement of adjoining street).

190. See Daniel J. Meltzer, Legislative Courts, Legislative Power, and the Constitution, 65 Ind. L.J. 291, 298 (1990) (“[I]t seems hard to contend that a scheme providing a single, but entirely fair, administrative determination necessarily denies due process.”). The Due Process Clause may well require that judicial review be available for claims that a State’s non-judicial compensation scheme fails to meet the criteria of (1) reasonableness, (2) certainty, and (3) adequacy. See Richard H. Fallon, Jr., Some Confusions About Due Process, Judicial Review, and Constitutional Remedies, 93
A State could even authorize its legislature to rule on claims for just compensation and

COLUM. L. REV. 309, 334 (1993) [hereinafter Fallon, Confusions] (observing that judicial review is usually available for claim that an administrative scheme violates Due Process Clause). In addition, it is arguable that the Supremacy Clause or Article III, or the two in combination, require that some federal court be able to review a claim that a state has failed to provide just compensation for a taking of private property for public use. See Seamon, Sovereign Immunity, supra note 104, at 393 n.363 (suggesting that Supremacy Clause may require that United States Supreme Court be able to review state adjudications of federal claims); see also Meltzer, supra, at 298–99 (arguing that Article III is more appropriate basis for constitutional right to judicial review); cf., e.g., Crowell v. Benson, 285 U.S. 22, 46, 60 (1932) (suggesting that judicial review of administrative determinations affecting life, liberty, or property can sometimes be required by Due Process Clause or Article III); Ohio Valley Water Co. v. Ben Avon Borough, 253 U.S. 287, 289 (1920) (stating that, when property owner claims that rates that state entity allows it to charge are confiscatory, “the State must provide a fair opportunity for submitting that issue to a judicial tribunal for determination upon its own independent judgment as to both law and facts; otherwise the [rate] order is void because in conflict with the due process clause”); Erwin Chemerinsky, The Hypocrisy of Alden v. Maine: Judicial Review, Sovereign Immunity and the Rehnquist Court, 33 LOY. L. REV. 1283, 1305 (2000) (appearing to argue that due process generally requires judicial forum); Martin Redish, Congressional Power to Regulate Supreme Court Appellate Jurisdiction Under the Exceptions Clause: An Internal and External Examination, 27 VILL. L. REV. 900, 915 & n.61 (1982) (arguing that due process requires an “independent” court to review constitutional claims); see generally Woolhandler and Collins, supra note 58, at 628–40 (discussing early eminent domain and rate cases in which Court found constitutional right, including due-process right, to judicial review).

191 See, e.g., Austin v. Ark. State Highway Comm’n, 895 S.W.2d 941, 942–44 (Ark. 1995) (holding that sovereign immunity barred just-compensation claim, but property owner could file claim with state Claims Commission); see also ARK. CODE. ANN. §§ 19-10-201 to -210 (Repl. 1994) (authorizing certain claims against State to be heard by executive-branch commission, subject to review by Arkansas legislature); cf. Crane, 258 U.S. at 145–47 (rejecting due process challenge to state procedure under which damages to plaintiff’s property were assessed by municipal officials, subject to state-court review only on issues of jurisdiction or fraud and willful misconduct by officials); Comm’n of Road Improvement v. St. Louis S.W. Ry., 257 U.S. 547, 555 (1922) (stating that “due process of law does not necessarily require judicial machinery to fix values in condemnation”); Bauman v. Ross, 167 U.S. 548, 593 (1897) (upholding land-acquisition scheme under which value of land was determined by commissioners, subject to judicial review); Long Island Water Supply Co. v. Brooklyn, 166 U.S. 685, 694–95 (1897) (holding that due process did not require jury to assess damages but was instead satisfied by procedure that permitted award to be determined by commissioners, subject to judicial review); Shoemaker v. United States, 147 U.S. 282, 285 (1893) (same); United States v. Jones, 109 U.S. 513, 519 (1883) (upholding federal statute authorizing state courts to hear condemnation actions brought by federal officials and remarking that determination of just compensation “may be prosecuted before commissioners or special boards or the courts . . . . All that is required is that it shall be conducted in some fair and just manner . . . .”); Kohl v. United States, 91 U.S. 367, 378–79 (1875) (“The proceeding by the States, in the exercise of their right of eminent domain, is often had before commissioners of assessment or special boards appointed for that purpose.”). But cf. Monongahela Navigation Co. v. United States, 148 U.S. 312, 327 (1893) (describing amount of just compensation as “judicial” question); but see also United States v. Cors, 337 U.S. 325, 330–36 (1949) (independently reviewing whether just compensation resulted from application of administrative regulations prescribing compensation for property confiscated for wartime use).
to award compensation by private bills. Today, it may seem doubtful that a non-judicial compensation scheme—particularly a private-bill scheme—could meet the due process criteria of reasonableness, certainty, and adequacy. Problems with non-judicial procedures for adjudicating claims against the government have led in modern times to the prevalence of judicial procedures. Historically, however, compensation claims against the government were handled by the legislative branch, and to a lesser extent by the executive branch, long before they were assigned to the courts. Given this history and the flexibility of due process, it would be untenable to conclude that the Due Process Clause categorically compels judicial resolution of just-compensation claims.

B. If a State Does Not Provide an Adequate, Alternative Way To Pay Just Compensation, the State Can Be Sued for Just Compensation in Its Own Courts of General Jurisdiction

Section A argued that States can meet their due process obligation by creating non-judicial schemes of providing just compensation for private property that the State has taken for public use. The question remains: What does the Due Process Clause require if a State does not create an adequate non-judicial compensation scheme and retains its immunity from just-compensation suits in its own courts?

---


195. See Walters v. Nat’l Ass’n of Radiation Survivors, 473 U.S. 305, 326 (1985) (“The flexibility in approach in our due process cases is intended in part to allow room for other forms of dispute resolution [i.e., besides courtroom litigation].”); see also, e.g., Burnham v. Superior Court of California, 495 U.S. 604, 619 (1990) (noting that history is relevant to due-process analysis); Ingraham v. Wright, 430 U.S. 651, 672–76 (1977) (same); Crane, 258 U.S. at 147 (same, with specific reference to just-compensation proceedings); Hurtado v. California, 110 U.S. 516, 528–29 (1884) (indicating that historical practice is relevant to due-process analysis).
In considering that question, it is worth keeping in mind that a person whose property has been taken for public use can often get some judicial relief without suing the state directly. Specifically, the property owner can sue the responsible state officials for money damages out of their own pockets and an injunction to prevent them from continuing to take the property without just compensation. These remedies probably are available in both federal court and the State’s own courts. If these remedies are available in state court and they combine to produce just compensation in a particular case, the property owner in that case cannot complain that the State has failed to provide an adequate remedy.

Often, however, an officer suit will not produce just compensation. An award of damages for any temporary taking will frequently be barred by

196. See supra notes 59–72 and accompanying text (describing remedies historically available for uncompensated takings in officer suits). Today, the property owner could seek money damages from the responsible state officials under 42 U.S.C. § 1983. Cf., e.g., City of Monterey v. Del Monte Dunes, 526 U.S. 687, 694 (1999) (stating that federal-court suit seeking monetary award against municipality for alleged regulatory taking was brought under Section § 1983). Injunctive relief against the official would also be available under Section 1983, as well as under the doctrine of Ex parte Young, 209 U.S. 123, 149–68 (1908). See Mitchum v. Foster, 407 U.S. 225, 242 (1972) (“Congress plainly authorized the federal courts to issue injunctions in § 1983 actions . . . .”). See also Ex parte Young, 209 U.S. at 151–52 (citing United States v. Lee, 106 U.S. 196 (1882)). In Lee, the taking of property without just compensation was alleged to support the principle authorizing an officer suit for equitable relief despite state sovereign immunity. See supra notes 60–61 and accompanying text (discussing Lee). Although the Court applies “ripeness” rules that would generally prevent a federal-court suit against a state official alleging an uncompensated taking, those rules would not apply if, as posited in the text, the State failed to provide an adequate process for obtaining compensation in its own courts. See, e.g., Williamson County Reg’l Planning Comm’n v. Hamilton Bank, 473 U.S. 172, 194 (1985) (holding that, before suing in federal court, takings claimant must exhaust state remedies for obtaining just compensation as long as state provides “an adequate process for obtaining compensation”); see generally Max Kidalov & Richard H. Seamon, The Missing Pieces of the Debate Over Federal Property Rights Legislation, 27 HASTINGS CONST. L.Q. 1, 5–17 (1999) (describing Court’s “ripeness” rules for takings claims in federal court).


198. Cf. Parratt v. Taylor, 451 U.S. 527, 543–44 (1981) (rejecting plaintiff’s argument that tort remedy against state was inadequate because it did not allow suit against offending officer, because remedy “fully compensated” plaintiff for his loss); Bob Jones Univ. v. Simon, 416 U.S. 725, 746–48 (1974) (finding it unnecessary to decide whether Tax Injunction Act would violate due process as applied in some situations, because plaintiff in that case had adequate relief).
official immunity, and sometimes even injunctive relief could be barred. When that is true and the State has not created an adequate nonjudicial scheme for awarding just compensation, the State will have failed to make “reasonable, certain and adequate provision for obtaining compensation.”

Under these circumstances, the victim of a taking should be able to sue the State directly for just compensation in a state court of general jurisdiction. In such a suit, the plaintiff would have a cause of action under the Just Compensation Clause, as applicable to States under the Fourteenth Amendment. The plaintiff might have to prove that an officer suit in state court would not yield just compensation and that the

---

199. See 9 DAVID A. THOMAS, THOMPSON ON REAL PROPERTY, § 81.05(d), at 514–15 (2d ed. 1994) (discussing difficulties of overcoming official immunity in just-compensation suits based on Section 1983).

200. See THOMAS, supra note 199, §81.05(c)(2), at 510–11 (discussing difficulties of obtaining injunctive relief in takings suits); see also supra notes 73–88 and accompanying text (discussing Court’s refusal to uphold injunctive relief in Hopkins and Belknap because of its interference with government property); cf. Idaho v. Coeur d’Alene Tribe, 521 U.S. 261, 281–88 (1997) (holding that federal-court officer suit for injunctive and declaratory relief by Native American Tribe was barred by sovereign immunity because it was equivalent to a quiet title action to which State had not consented); Larson v. Domestic & Foreign Commerce Corp., 337 U.S. 682, 691 n.11 (1949) (noting that injunctive relief in officer suit would be barred “if the relief requested can not be granted by merely ordering the cessation of the conduct complained of but will require affirmative action by the sovereign or the disposition of unquestionably sovereign property”); Westside Mothers v. Haveman, 133 F. Supp. 2d 549, 560–75 (E.D. Mich. 2001) (holding that sovereign immunity barred Ex parte Young relief against state officials who administered Medicaid program).

201. See supra note 181 (citing cases articulating the “reasonable, certain, and adequate” standard).

202. Cf. United States v. Lee, 106 U.S. 196, 218 (1882) (describing claim based on Just Compensation Clause as being “of that character which it is intended the courts shall enforce”); Jackson, The Supreme Court, supra note 18, at 117 (arguing from Court’s precedent that “a remedy for unlawful government conduct,” including takings of private property for public use, should be provided “if there [is] a jurisdictional basis for doing so”). A state court is said to have “general jurisdiction” when it has power under state law “to hear a wide range of cases, civil or criminal, that arise within its geographic area.” BLACKS LAW DICTIONARY 856 (7th ed. 1999). All States have such courts. See P. BARNES, CONGRESSIONAL QUARTERLY’S DESK REFERENCE ON THE AMERICAN COURTS 173 (2000) (“All states have at least one court of general jurisdiction . . . .”); BUREAU OF JUSTICE STATISTICS, U.S. DEP’T OF JUSTICE, OFFICE OF JUSTICE PROGRAMS, STATE COURT ORGANIZATION: 1998 15–17 (2000) (Table 3). Although sovereign immunity ordinarily could be said to restrict a state court’s jurisdiction, that restriction would have to be ignored when the assertion of sovereign immunity would violate a State’s obligations under the Fourteenth Amendment. Cf. Testa v. Katt, 330 U.S. 386, 394 (1947) (holding that state court of otherwise competent jurisdiction was obligated to hear federal claim, despite state policy of not enforcing penal laws of foreign jurisdiction).

203. See First English Evangelical Lutheran Church v. County of Los Angeles, 482 U.S. 304, 315–16 (1987); see also supra notes 19–42 and 49 and accompanying text (discussing First English).
State did not provide any adequate, nonjudicial procedure for the plaintiff to get it.\textsuperscript{204} If those conditions are proven, the State should not be allowed to assert sovereign immunity as a defense. That is because, if a state court honored that defense, the state court would cause the State to deny “reasonable, certain, and adequate provision” for payment of just compensation.\textsuperscript{205} Because that result would violate the Fourteenth Amendment, the state court would have to reject the State’s sovereign immunity defense.\textsuperscript{206}

Some commentary has suggested that it is the Supremacy Clause,\textsuperscript{207} rather than the Due Process Clause of the Fourteenth Amendment, that requires state courts to entertain federal claims against their own States despite sovereign immunity.\textsuperscript{208} In light of \textit{Alden}, however, the Supremacy Clause cannot serve as a \textit{deus ex machina}\textsuperscript{209} that compels state courts to choose the just-compensation principle over the sovereign immunity principle on the theory that sovereign immunity is merely a creature of state law that is preempted by the federal just-compensation principle. \textit{Alden} establishes that a State’s immunity in its own courts

\begin{footnotesize}
\begin{enumerate}
\item[205.] See supra note 181 (citing cases articulating the “reasonable, certain, and adequate” standard).
\item[207.] \textit{U.S. CONST.} art. VI, cl. 2, states:
This Constitution, and the Laws of the United States which shall be made in Pursuance thereof; and all Treaties made, or which shall be made under the Authority of the United States, shall be the supreme Law of the Land; and the Judges in every State shall be bound thereby, any Thing in the Constitution or Laws of any State to the Contrary notwithstanding.
\item[208.] See, e.g., \textit{Grant}, supra note 19, at 199–200 (contending that, because “the sovereign Constitution stands supreme,” Just Compensation Clause “abrogates” federal sovereign immunity); \textit{Katz, supra} note 18, at 1492–93 (arguing that Supremacy Clause obligates state courts to hear federal claims against States because of “[t]he absence of a broad constitutionally based state immunity” in state court); \textit{Gordon & Gross, supra} note 18, at 1171–74 (similarly relying on Supremacy Clause to argue that state courts must hear federal claims against the State and state officers, notwithstanding sovereign immunity). \textit{But see} \textit{Alden v. Maine}, 527 U.S. 706, 730 (1999) (holding that States do have immunity in their own courts that is derived from U.S. Constitution).
\item[209.] A “\textit{deus ex machina}” is “a person or thing . . . that appears or is introduced suddenly and unexpectedly and provides a contrived solution to an apparently insoluble difficulty.” \textit{MERRIAM WEBSTER’S COLLEGIATE DICTIONARY} 316 (19th ed. 1993) (“\textit{deus ex machina},” definition 2).
\end{enumerate}
\end{footnotesize}
State Sovereign Immunity

derives from the U.S. Constitution, as does the just-compensation principle. Thus, the Supremacy Clause does not cause one principle to trump the other.

A state court must choose the just-compensation principle over the sovereign-immunity principle not because of the Supremacy Clause, but because sovereign immunity is “a part of the Constitution, of equal authority with every other, but no greater.” Sovereign immunity and the Fourteenth Amendment each get their due “authority” if the Fourteenth Amendment is construed as generally permitting a State to meet its remedial obligations under that Amendment in ways other than letting itself be sued directly for money. This is why it is appropriate, in particular, to interpret the Fourteenth Amendment to permit a State to provide just compensation through non-judicial procedures (as well as through officer suits). That interpretation preserves the essence of sovereign immunity, which, at its core, protects the sovereign from being called upon by a court to honor its monetary obligations. Sovereign immunity cannot eliminate those obligations, however, when those obligations, like sovereign immunity itself, arise from the Constitution. To avoid giving sovereign immunity higher “authority” than the Fourteenth Amendment obligation to pay just compensation, the immunity must yield when the State has not met that obligation in ways that do not implicate sovereign immunity. A contrary conclusion would pervert the federal system established by the Constitution. A central purpose of the federal system is to protect individual liberty through the diffusion of power to the States. Thus, state sovereignty is

210. See, e.g., Alden, 527 U.S. at 730 (saying that state sovereign immunity “inheres in the system of federalism established by the Constitution”); id. at 754 (“In light of history, practice, precedent, and the structure of the Constitution, we hold that the States retain immunity from private suit in their own courts, an immunity beyond the congressional power to abrogate by Article I legislation.”); see also supra notes 140–77 and accompanying text (discussing Alden).


212. See supra notes 179–95 (arguing that States can meet their obligation to provide procedures for paying just compensation without letting themselves be sued in their own courts).

213. See Fallon, Confusions, supra note 190, at 338 (“[T]he Constitution in general and the Due Process Clause in particular do sometimes require individually effective remediation for constitutional violations. In such cases, competing doctrines, including sovereign and official immunity, must give way.”) (footnotes omitted).

214. Cf. Jackson, The Supreme Court, supra note 18, at 114–16 (arguing that state sovereign immunity should be understood to create a preference against monetary remedies against the States, which would be overcome “where the supremacy of constitutional law” requires).

215. See Alden, 527 U.S. at 758 (attributing to Framers of Constitution “the unique insight that freedom is enhanced by the creation of two governments, not one”); Printz v. United States, 521 U.S. 898, 921 (1997) (“This separation of the two spheres is one of the Constitution’s structural
protected from federal incursion to preserve individual rights, not vice versa.216

One could argue that this conclusion gives more “authority” to the Fourteenth Amendment than to the States’ constitutionally based immunity. To the contrary, the conclusion reflects that sovereign immunity gives the States only the privilege of not being called upon by courts, at the instance of private plaintiffs, to honor their obligations. Sovereign immunity is a protective device that can be used against private lawsuits, not a destructive device that can be used to eliminate the constitutional rights of individuals.

C. Case Law Supports the Conclusion that an Unconsenting State Can Be Sued in Its Own Courts of General Jurisdiction If It Fails To Make an Adequate, Alternate Provision for Just Compensation

The most direct support for the analysis advanced in Section B comes from Supreme Court cases on state taxation. The Court has held that the Due Process Clause does not require a State to let a taxpayer challenge a state tax before it is collected. Instead, the State may compel the taxpayer to pay the tax first and challenge it later.217 When the State exerts that compulsion, however, the Court has said that the procedural component

---

216. If a state court, contrary to the analysis described in the text, erroneously accepted the State’s sovereign immunity defense, the property owner would be entitled, after pursuing appeals through the state-court system, to appellate review by the United States Supreme Court. See 28 U.S.C. § 1257(a) (authorizing United States Supreme Court to review final judgments rendered by “the highest court of a State in which a decision could be had” where “any title, right, privilege, or immunity is specially set up or claimed under the Constitution”); McKesson Corp. v. Div. of Alcoholic Beverages & Tobacco, 496 U.S. 18, 26–31 (1990) (holding that “[t]he Eleventh Amendment does not constrain the appellate jurisdiction of the Supreme Court over cases arising from state courts”); see also First English Evangelical Lutheran Church v. County of Los Angeles, 482 U.S. 304, 310–13 (1987) (holding that Court could review state-court decision denying compensation for temporary regulatory taking).

217. See, e.g., Nat’l Private Truck Council, Inc. v. Okla. Tax Comm’n, 515 U.S. 582, 587 (1995) (“As long as state law provides a clear and certain remedy, the States may determine whether to provide predeprivation process (e.g., an injunction) or instead to afford post-deprivation relief (e.g., a refund). . . .”) (internal quotation marks and citations omitted); Reich v. Collins, 513 U.S. 106, 110–11 (1994) (stating that Due Process Clause gives State “the flexibility to maintain an exclusively predeprivation remedial scheme” for wrongfully collected taxes, as well as “an exclusively postdeprivation regime”); McKesson, 496 U.S. at 37 (“[I]t is well established that a State need not provide predeprivation process for the exaction of taxes.”); id. at 51 (holding that, when State compels taxpayer to pay taxes before challenging them, State’s post-deprivation procedure must provide “a clear and certain remedy”) (internal quotation marks omitted).
of the Due Process Clause obligates the State to provide a “clear and certain remedy” for any erroneous tax collection. The Court has also said that this remedy must, in some cases, take the form of a refund of the erroneously collected taxes. Thus, the procedural due process requirements for erroneously collected taxes parallel those for governmental takings of private property for public use: The State does not have to provide a procedure for challenging the government’s deprivation of taxes or property (or its taking of private property for a public use) before the deprivation (or taking) occurs. The State does, however, have to provide a post-deprivation procedure that is “clear and certain” and that can result in an award of money from the state treasury. Ordinarily, States provide a post-deprivation remedy by letting themselves be sued for tax refunds in their own courts.

Admittedly, the Court’s tax decisions would have to be extended to support the conclusion that, in the absence of an adequate alternative, an unconsenting State can be sued for just compensation in its own courts. For one thing, the Court has not yet extended this case law outside the tax context. For another thing, even limited to the tax context, the

---


219. See Newsweek, 522 U.S. at 444–45 (holding that taxpayer was entitled to refund in light of state statutory provisions that appeared to authorize that remedy); Nat’l Private Truck Council, 515 U.S. at 587–92 (holding that, where state law provided for refund required by due process, state court could refuse to award declaratory and injunctive relief under 42 U.S.C. § 1983); Reich, 513 U.S. at 111–14 (taxpayer was entitled to refund in state court given statutory provisions authorizing that relief, despite later state court’s interpretation of those provisions as not authorizing refunds); McKesson, 496 U.S. at 39 (observing that, if state tax were beyond State’s power to impose, or imposed on an entity immune from taxation, “[t]he State would have had no choice but to ‘undo’ the unlawful deprivation by refunding the tax”).


221. See Reich, 513 U.S. at 112 (“States ordinarily prefer that taxpayers pursue only postdeprivation remedies—i.e., that taxpayers ‘pay first, litigate later.’”) (emphasis omitted).

222. See Fallon & Meltzer, supra note 18, at 1825–26 (observing that Court has not applied its decisions requiring refunds in tax cases to other contexts); see also The Supreme Court, 1989 Term:
Court’s tax cases need not be read to mean that the Fourteenth Amendment overrides state sovereign immunity. In all of the cases in which the Court has cited the “clear and certain remedy” principle to sustain tax-refund suits against States brought in state court, the State has waived its immunity by allowing itself to be sued for a refund in its own courts. Thus, the cases may mean only that the Due Process Clause “requires a State to provide the remedy it has promised.” The Court has never construed the Due Process Clause to require state courts to refund taxes paid to a State that has not waived its immunity from such suits. Likewise, the Court has never held that a state court can award just compensation from the treasury of an unconsenting State.

The extension of the tax cases urged here accords with case law holding or suggesting that sometimes a State can satisfy its due process obligations by letting its officers, rather than the State itself, be sued in

---

Leading Cases, 104 HARV. L. REV. 129, 197 (1990) (remarking that McKesson “may not be generalizable”).

223. See McKesson, 496 U.S. at 49 n.34 (noting that Florida “concede[d] that the State waived any sovereign immunity”); see also Seamon, Sovereign Immunity, supra note 104, at 399 n.394 (citing McKesson and later cases).

224. *Alden v. Maine*, 527 U.S. 706, 740 (1999) (emphasis added); accord *Hudson v. Palmer*, 468 U.S. 517, 539 (1984) (O’Connor, J., concurring) (“The due process requirement means that government must provide to the inmate the remedies it promised would be available.”) (emphasis added); see also *Richards v. Jefferson County*, 517 U.S. 793, 804 (1996) (“Whether acting through its judiciary or through its legislature, a State may not deprive a person of all existing remedies for the enforcement of a right, which the State has no power to destroy, unless there is, or was, afforded to him some real opportunity to protect it.”) (emphasis added) (quoting Brinkerhoff-Farris Trust & Sav. Co. v. Hill, 281 U.S. 673, 682 (1930)); cf. *FALLON*, supra note 19, 1999 SUPPLEMENT, at 135 (questioning whether *Alden* “suggests that the state could plead sovereign immunity in its own courts if no state remedy appeared to be available and the doors of the federal courts were similarly closed”).

225. Compare *Reich v. Collins*, in which the Court held that, once a taxpayer had paid taxes relying on existing state law that appeared clearly to authorize refund actions against the State, a State could not decide that no such remedy existed and that the taxpayer should have challenged the taxes before they were collected. *Reich*, 513 U.S. at 110–14. The Court considered this an unconstitutional “bait and switch” tactic. Id. at 111; see also *Brinkerhoff-Farris Trust*, 281 U.S. at 674–82 (holding that due process was violated by state-court decision that had the effect of preventing taxpayer from raising federal challenge to tax). As Professor Wells has observed, the State in *Reich* did not raise sovereign immunity as a defense to the refund action. See Michael Wells, *Suing States, supra note 18, at 779; see generally Ann Woolhandler, Old Property, New Property, and Sovereign Immunity, 75 NOTRE DAME L. REV. 919, 928–29 (2000) [hereinafter Woolhandler, *Old Property*] (demonstrating that, with possible exception of *Reich*, the Court has required states to provide remedies against themselves in their own courts only “where the state had substituted remedies against itself” for remedies against its officers).

226. See supra notes 53–91 and accompanying text (demonstrating, in Part II of this Article, that the Court has never decided whether sovereign immunity bars just-compensation claim against an unconsenting State).
As discussed above, officer suits could indeed yield just compensation in some cases. In those cases, the victim of a taking may not be constitutionally entitled to sue the State directly. In many cases, however, officer suits would not produce just compensation. Official immunity could prevent damage awards against officers for any taking that has already occurred, and, in any event, many officers would be judgment proof. Moreover, injunctive relief against continued takings would not always be available. Officer suits therefore would not always satisfy the Due Process Clause requirement that a State provide a “reasonable, certain, and adequate” procedure for paying just compensation every time it takes private property for public use. In cases in which the Court has held that post-deprivation officer suits satisfied due process, the plaintiff failed to show that this remedy was inadequate.

Moreover, the cases in which the Court has found that officer suits satisfy due process differ in two important ways from cases involving claims for just compensation. First, the nature of the government action that caused the deprivation in these cases differed from governmental takings of private property for public use. Specifically, they involved “random and unauthorized” deprivations by state officials. In contrast, a taking of private property triggers a right to just compensation only

227. See Zinermon v. Burch, 494 U.S. 113, 128 (1990) (“In some circumstances . . . the Court has held that . . . a common-law tort remedy for erroneous deprivation, satisfies due process.”).

228. See supra notes 196–98 and accompanying text.

229. See supra note 199 and accompanying text (discussing official immunity as limit on monetary relief in actions based on uncompensated taking).

230. See Owen v. City of Independence, 445 U.S. 622, 652 n.36 (1980); cf. Scott v. Donald, 165 U.S. 107, 109 (1897) (plaintiffs alleging that defendants “were wholly irresponsible financially, and unable to respond in damages” in successfully seeking federal court injunction against state officer’s enforcement of unconstitutional tax laws).

231. See supra note 200 and accompanying text (discussing limits on injunctive relief in officer suits based on uncompensated taking).

232. See Hudson v. Palmer, 468 U.S. 517, 535–36 (1984) (holding that due process was satisfied by existence of state tort remedy against officer who had destroyed plaintiff’s property); Ingraham v. Wright, 430 U.S. 651, 674–82 (1977) (holding that due process was satisfied by availability of state tort remedy for improper corporal punishment of public school student); see also N. Am. Cold Storage Co. v. City of Chicago, 211 U.S. 306, 314–16 (1908) (holding that plaintiff was not entitled to notice and hearing before health officials shut down facility suspected of harboring unwholesome food, while observing that plaintiff would be able to sue state officials afterwards).

when it is authorized by state law.\textsuperscript{234} It is arguably fairer to require the State itself to pony up for a deprivation that it authorizes than for a deprivation that it does not authorize.\textsuperscript{235} That is especially true when the state-authorized deprivation consists of a taking of private property for public use. That type of deprivation, by its nature, benefits the public.\textsuperscript{236} With that sort of deprivation, it is fair to make the public treasury pay the bill.\textsuperscript{237} More fundamentally, a violation of the Just Compensation Clause differs from all other constitutional violations in that the Just Compensation Clause “dictates” a monetary remedy for every governmental taking of private property for public use.\textsuperscript{238} Furthermore, the Clause seems to contemplate that this compensation will come from the public funds of government entities that authorize the taking, and not from its officers’ personal funds.\textsuperscript{239} Besides governmental takings of

---

\textsuperscript{234} See, e.g., United States v. N. Am. Transp. & Trading Co., 253 U.S. 330, 333 (1920) (“In order that the Government shall be liable [for just compensation] it must appear that the officer who has physically taken possession of the property was duly authorized so to do, either directly by Congress or by the official upon whom Congress conferred the power.”).

\textsuperscript{235} Cf., e.g., Bd. of County Comm’rs v. Brown, 520 U.S. 397, 403–04 (1997) (municipal liability under Section 1983 cannot be based on respondeat superior but instead generally requires proof of custom or policy). But cf. Home Tel. & Tel. Co. v. City of Los Angeles, 227 U.S. 278, 282–96 (1913) (holding that action under color of state law could violate Fourteenth Amendment even if it violated state law). Under Home Telephone, the fact that a state official’s conduct violates state law cannot matter in determining whether a violation of the Fourteenth Amendment has occurred, but I believe that this fact can play a role in determining the remedy required for that violation. This Article traces the States’ remedial obligation to the procedural component of the Due Process Clause of the Fourteenth Amendment and identifies the main goal of that Clause as ensuring accuracy. If one accepts that approach, then it seems sensible in crafting the remedy to consider whether the violation is the result of the conduct of a rogue official or more systemic.


\textsuperscript{237} See Armstrong v. United States, 364 U.S. 40, 49 (1960) (“The Fifth Amendment’s guarantee [of just compensation] . . . was designed to bar the government from forcing some people alone to bear public burdens which, in all fairness and justice, should be borne by the public as a whole.”); see also United States v. Dickinson, 331 U.S. 745, 748 (1947) (relying on principle that Just Compensation Clause “expresses a principle of fairness” to reject government’s argument that taking claim was barred by statute of limitations).

\textsuperscript{238} First English Evangelical Lutheran Church v. County of Los Angeles, 482 U.S. 304, 316 n.9 (1987).

\textsuperscript{239} See, e.g., United States v. Security Indus. Bank, 459 U.S. 70, 77 (1982) (stating that Just Compensation Clause requires government to use eminent domain power to take private property for public use “so that, through taxation, the burden of the relief afforded in the public interest may be borne by the public”) (quoting Louisville Joint Stock Land Bank v. Radford, 295 U.S. 555, 602 (1935)); Armstrong, 364 U.S. at 49 (stating that “public as a whole” is meant to bear burden—i.e., of paying compensation—when government takes private property for public use); United States v. Gettysburg Elec. Ry. 160 U.S. 668, 680 (1896) (reasoning that potential for government’s abuse of its taking power is diminished by fact that just compensation for taking “must be raised by taxation”); THE COMPLETE BILL OF RIGHTS § 11.3.1.1, at 377 (Neil H. Cogan ed. 1997) (quoting
State Sovereign Immunity

private property for public use, no other deprivation subject to the Due Process Clause is also subject to a constitutional provision that demands governmental compensation for every such deprivation.240

D. Due Process Can Create an Asymmetry Between a State’s Immunity from Suits Brought in Its Own Courts and Its Immunity from Suits Brought in Federal Court

This Part of the Article has argued that an unconsenting State can be sued for just compensation in its own courts of general jurisdiction if it fails to implement an adequate, alternative way of paying just compensation for property that it has taken for public use. That argument, if accepted, creates an asymmetry between the States’ immunity in their own courts and their immunity in federal court. It would subject States to suits in their own courts that, under the analysis advanced in Part II, could not be brought in federal court.241

Blackstone’s statement that, when government takes property involuntarily, the legislature must provide “full indemnification” to property owner); Joseph L. Sax, Takings and the Police Power, 74 YALE L.J. 36, 54 (1964) (quoting 17th century view of Grotius that “just satisfaction” for land taken by eminent domain should be paid “out of the common stock”); William B. Stoebuck, A General Theory of Eminent Domain, 47 WASH. L. REV. 553, 566–67, 572–88 (1972) (discussing historical basis for requirement of compensation from government for its taking of private property for public use).

240. This does not mean that the Just Compensation Clause creates the only substantive right for which the Due Process Clause requires a monetary remedy. As discussed supra notes 217–21 and accompanying text, and infra notes 300–314 and accompanying text, the Due Process Clause has been construed to mandate a monetary remedy in at least one other context—namely, as a remedy for state taxes collected in violation of federal law. The Clause may very well dictate monetary remedies in other contexts, too, but the identification of those contexts is complicated by the absence of explicit guidance in the text of the Constitution.

241. See supra notes 53–91 and accompanying text (arguing in Part II that sovereign immunity bars just-compensation suits brought against States in federal court); cf. Reich, 513 U.S. at 110 (noting that, while Constitution bars state courts from denying recovery of state taxes collected in violation of federal law, “the sovereign immunity States traditionally enjoy in their own courts notwithstanding,” the States’ Eleventh Amendment in federal court “does generally bar tax refund claims from being brought in that forum”) (dicta). Quite apart from the restriction on suits seeking monetary relief that is imposed by the doctrine of sovereign immunity, an additional restriction is imposed by the provisions in many states constitutions that, like the U.S. Constitution, prohibit payments from the treasury except by legislative appropriation. See James M. Hirschhorn, Where the Money Is: Remedies To Finance Compliance with Strict Structural Injunctions, 82 MICH. L. REV. 1815, 1837–38 n.120 (1984) (stating that most state constitutions forbid disbursements from state treasury except by appropriation statute); cf. U.S. CONST. art. I., § 9, cl. 7 (Appropriations Clause). Under the analysis proposed here, these state constitutional provisions could not prevent a state court of general jurisdiction from awarding just compensation against the State, in the absence of an adequate, alternative compensation procedure, for the same reason that the State’s sovereign immunity could not prevent such an award. A state court’s reliance on these state constitutional provisions to deny just compensation would violate the Fourteenth Amendment. See Fallon &
This asymmetry is justified by the differing roles of the state courts and federal courts when it comes to the States’ obligations under the Due Process Clause of the Fourteenth Amendment. The Due Process Clause obligates the sovereign that takes private property for public use to make “reasonable, certain and adequate provision” for awarding just compensation to the property owner. The “taking” sovereign can use its own courts to meet this obligation. The obligation cannot be met, however, by the courts of another sovereign. Thus, state courts have a primary role in ensuring their State’s compliance with the Just Compensation Clause, which is part of the broader role in safeguarding constitutional rights contemplated for those courts by the Framers of the Constitution.

Of course, a federal court can often provide effective remedies for a State’s violation of the Due Process Clause of the Fourteenth Amendment, despite the sovereign immunity that States enjoy from suits brought in federal court. Specifically, remedies against the state officials who are responsible for the violation may be available in federal court under 42 U.S.C. § 1983, which permits money judgments payable out of the officials’ own pockets, or under the doctrine of Ex parte Young, which permits injunctive relief against the officials. This Article should not be read to endorse any restriction on those federal-court remedies. Rather, its point is that the availability of federal-court remedies does not excuse a State’s failure to provide its own remedies—either in its courts or by some other means—to the extent required by the

Meltzer, supra note 18, at 1786 (“[T]he Supreme Court has sometimes compelled state courts to provide constitutional remedies despite a lack of state law authority for them to do so.”).

242. See supra note 181 (citing case law establishing requirement of “reasonable, certain, and adequate provision” for payment of just compensation); see also infra notes 260–325 (explaining that this requirement stems from doctrine of procedural due process).


244. Because Article III of the Constitution gave Congress discretion whether or not to create the lower federal courts, state courts were bound to have major responsibility for protecting individual constitutional rights. See Henry M. Hart, Jr., The Power of Congress to Limit the Jurisdiction of Federal Courts: An Exercise in Dialectic, 66 Harv. L. Rev. 1362, 1401 (1953) (stating that, under original Constitution, state courts were supposed to be “the primary guarantors of constitutional rights”).

245. See supra notes 53–91 and accompanying text (Part II).

246. Ex parte Young, 209 U.S. 123 (1908).

247. Nor should it be read to cast doubt on the U.S. Supreme Court’s exercise of appellate jurisdiction over decisions in state-court cases in which just compensation is sought from a State. See supra note 216.
V. IMPLICATIONS OF DUE PROCESS CONSTRAINTS ON STATE SOVEREIGN IMMUNITY

Part IV concluded that the Due Process Clause of the Fourteenth Amendment can override the States’ immunity from just-compensation suits brought in their own courts. This Part discusses the implications of that conclusion and of the analysis underlying it. As explained in Section A, the conclusion is directly relevant, because many States have not clearly waived their immunity from all just-compensation claims. In addition, as discussed in Section B, the analysis is relevant in analyzing the States’ ability to use sovereign immunity to avoid other obligations under the Due Process Clause of the Fourteenth Amendment. As Section C explains, because the proposed analysis rests on the Due Process Clause, it applies to the federal government as well as to the States. Finally, Section D discusses the implications of the analysis proposed.

248. This Part of the Article has argued that the States’ immunity in their own courts is overridden in some circumstances by their obligation under the procedural component of the Due Process Clause to establish procedures for meeting their substantive obligation to pay just compensation. It is the Article’s reliance on procedural due process, in combination with the substantive obligation imposed by the Just Compensation Clause, that results in the asymmetry between the State’s immunity in their own courts and their immunity in federal court. This asymmetry would not arise if—as many commentators believe the Court concluded in First English, see supra note 19 (citing commentary)—the Just Compensation Clause, standing alone, were interpreted to override state sovereign immunity. In that event, sovereign immunity presumably would not bar a suit against a State for just compensation regardless whether the suit were brought in federal court or in state court. (A just-compensation suit in federal court could, however, be barred by the Court’s “ripeness” rules. See note 196 supra (discussing Court’s ripeness rules).) This Article’s reliance on the Due Process Clause of the Fourteenth Amendment, rather than on the Just Compensation Clause standing alone, is consistent with the expression of doubt by a plurality of the Court in Del Monte Dunes whether sovereign immunity “retains its vitality in cases where the [Fourteenth] Amendment is applicable.” City of Monterey v. Del Monte Dunes, 526 U.S. 687, 714 (1999) (plurality opinion). This doubt suggests that sovereign immunity could be overridden to protect other substantive obligations applicable to the States under the Fourteenth Amendment. This Article’s analysis is also consistent with the case law of the Court that, this Article has argued in Part II, virtually compels the conclusion that the Just Compensation Clause does not, of its own force, override the States’ sovereign immunity from actions brought in federal court. See supra notes 53–91 and accompanying text (Part II). By proposing an analysis that is consistent with my understanding of precedent, I do not mean to imply that an alternative analysis would be illogical. As Professor Meltzer remarked in a personal communication with me, “the Court might take the view that (a) the express grant of a compensatory remedy for a takings should be viewed as incompatible with sovereign immunity, while (b) whatever remedies may be implied as appropriate under the Due Process Clause are subordinate to sovereign immunity.” E-Mail from Daniel J. Meltzer, Professor of Law, Harvard Law School, to Richard H. Seamon, Assistant Professor of Law, Univ. of S.C. (Mar. 5, 2001) (on file with author). I agree with this observation, while believing that such a ruling would be difficult to reconcile with the case law discussed in this Article.
here for Congress’s power to enforce the Fourteenth Amendment.

A. State Waivers of Immunity from Just-Compensation Claims

This Article addresses whether an unconsenting State can be sued for just compensation. That issue does not matter in States that have enacted legislation waiving the State’s immunity from all just-compensation suits in their courts. \(^{249}\) There are three other categories of States in which the issue does matter: (1) States in which it is unclear whether, as a matter of state law, the State has waived immunity from all just-compensation

\(^{249}\) At least two States, New York and Utah, expressly waive their immunity from just-compensation claims. See N.Y. CT. CL. ACT § 9(2) (McKinney 2001) (empowering court of claims to hear monetary claims “against the state for the appropriation of any real or personal property or interest therein”); UTAH CODE ANN. § 63-30-10.5(1) (1997) (providing that “immunity from suit of all governmental entities is waived for the recovery of compensation from the governmental entity when the governmental entity has taken or damaged private property for public uses without just compensation”). Other States have statutes that expressly authorize inverse-condemnation claims against the State under at least some circumstances. For example, MASS. GEN. LAWS ANN. ch. 79, § 10 (West 2001) authorizes an inverse-condemnation action when the real estate of any person has been taken for the public use or has been damaged by the construction, maintenance, operation, alteration, repair or discontinuance of a public improvement or has been entered for a public purpose, but such taking, entry or damage was not effected by or in accordance with a formal vote or order of the board of officers of a body politic or corporate duly authorized by law, or when the personal property of any person has been damaged, seized, destroyed or used for a public purpose. See also N.M. STAT. ANN. §§ 42A-1-2(H) & 42A-1-29 (Michie 1994) (authorizing inverse-condemnation actions against any “person,” including a governmental entity, that is authorized to exercise eminent domain); N.C. GEN. STAT. §§ 40A-3 & 40A-51 (1999) (authorizing inverse-condemnation actions against certain entities with eminent domain power); PA. CONS. STAT. ANN. tit. 26, §§ 1-502(e), 1-511 to 1-518 (West 1997) (authorizing condemnee whose land has been taken without institution of eminent domain proceeding to file petition triggering assessment of just compensation by “viewers,” subject to judicial review). Still other States have statutes that waive their sovereign immunity in more general terms. See, e.g., CAL. GOV’T CODE § 945 (West 2001) (“A public entity may sue and be sued.”); 705 ILL. COMP. STAT. 505/8 (1999) (authorizing claims against State to be asserted in court of claims); Mich. Comp. Laws 600.6419 (2000) (same); Ohio Rev. Code Ann. § 2743.02(A)(1) (Anderson 2000) (same); see also Nev. Rev. Stat. Ann. § 41.031 (Michie 1999) (broadly waiving State’s sovereign immunity from private actions in its own courts). Even in States with statutes that, specifically or in general terms, waive sovereign immunity from just-compensation claims, state courts may construe the statutes in a way that may preserve sovereign immunity from certain just-compensation claims. See, e.g., In re Condemnation of 2719, 21, 11 E. Berkshire St., 343 A.2d 67, 69 (Pa. Commw. Ct. 1975) (holding that Pennsylvania statute authorizing inverse-condemnation actions does not apply because plaintiffs failed to allege government defendant was lawfully exercising power of eminent domain); see also, e.g., Reynolds v. State, 471 N.E.2d 776, 778 (Ohio 1984) (stating that Ohio Court of Claims Act does not waive sovereign immunity from claims based on State executives’ exercise of discretionary authority).
suits brought in its courts;\textsuperscript{250} (2) States in which it is clear that the State has retained its immunity from at least some just-compensation claims brought against it in state court;\textsuperscript{251} and (3) States in which it has been the state courts, rather than the legislatures, that have decided to abrogate state-court immunity from just-compensation claims.\textsuperscript{252} It is obvious

\textsuperscript{250} See \textit{Mandelker}, \textit{supra} note 1, § 4A.02[5][d] ("[I]n many states the availability of a
compensation remedy in land use cases is not clear."); see also, e.g., \textit{New Port Largo, Inc. v. Monroe
County}, 985 F.2d 1488, 1493 n.12 (11th Cir. 1993) (noting uncertainty in Florida case law about
whether courts recognize cause of action in inverse condemnation for temporary takings), \textit{cert.
denied}, 510 U.S. 964 (1993); \textit{Austin v. City and County of Honolulu}, 840 F.2d 678, 681 (9th Cir.
1987) (observing that there was no Hawaii case law either establishing or rejecting existence of
("In deciding whether a State has waived its [sovereign immunity], we will find a waiver only where
stated by the most express language or by such overwhelming implications from the text as [will]
leave no room for any other reasonable construction.") (internal quotation marks omitted; brackets
supplied by the Court).

\textsuperscript{251} See \textit{Austin v. Ark. State Highway Comm’n}, 895 S.W.2d 941, 942–44 (Ark. 1995), \textit{cited
condemnation claim against State was barred by sovereign immunity), \textit{appeal denied} (Apr. 6, 1998).

\textsuperscript{252} See, e.g., \textit{Barber v. State}, 703 So. 2d 314, 317–22 (Ala. 1997) (reversing summary judgment
for State in case asserting inverse-condemnation claim; noting that such claims are not barred by
state constitutional provision establishing sovereign immunity); \textit{City of Kenai v. Burnett}, 860 P.2d
1233, 1238–39 (Alaska 1993) (apparently identifying state constitution’s just-compensation
provision as source for causes of action in inverse condemnation); \textit{Bd. of Comm’rs v. Adler}, 194 P.
621, 622–24 (Colo. 1920) (holding that state constitution’s just-compensation provision overrode
state’s sovereign immunity); \textit{Textron, Inc. v. Wood}, 355 A.2d 307, 311–13 (Conn. 1974) (holding
that sovereign immunity did not bar suits seeking declaration that State owed plaintiff just
compensation and stating in dicta that damages would also be available); \textit{State ex rel. Smith v.
0.24148, 0.23831 & 0.12277 Acres of Land}, 171 A.2d 228, 231 (Del. 1961) (holding that state
cstitution’s just-compensation provision was self-executing waiver of state’s sovereign immunity
from just-compensation claims); \textit{Columbia County v. Doolittle}, 512 S.E.2d 236, 237 (Ga. 1999)
(interpreting state constitution’s just-compensation provision as waiving sovereign immunity);
\textit{Deisher v. Kan. Dep’t of Transp.}, 958 P.2d 656, 662–63 (Kan. 1998) (discussing development of
inverse condemnation in Kansas); \textit{Dep’t of Highways v. Corey}, 247 S.W.2d 389, 389–91 (Ky. 1952)
(holding that sovereign immunity did not bar claim against state agency based on state constitution’s
just-compensation provisions); \textit{Foss v. Me. Turnpike Auth.}, 309 A.2d 339, 343–45 (Me. 1973)
(holding that state sovereign immunity did not bar claim for just compensation against state agency);
\textit{Dep’t of Natural Res. v. Welsh}, 521 A.2d 313, 315–19 (Md. 1986) (holding that sovereign immunity
did not bar quiet title action against state agency alleging unlawful taking of private property); \textit{State
ex rel. Peterson v. Bentley}, 12 N.W.2d 347, 357 (Minn. 1943) (holding that sovereign immunity did
not bar just-compensation claim for taking of land outside of State, and stating in dicta that
sovereign immunity would not bar just-compensation claim for taking of land inside State),
\textit{overruled in part on other grounds sub nom. State by Peterson v. Anderson}, 19 N.W.2d 70 (Minn.
1945); \textit{Williams v. Walley}, 295 So. 2d 286, 288 (Miss. 1974) (holding that sovereign immunity
did not protect governmental entity from just-compensation claim in light of self-executing nature of
state constitutional guarantee of just compensation); \textit{McGrew v. Granite Bituminous Paving Co.},
155 S.W. 411, 415 (Mo. 1912) (holding that state constitution’s just-compensation provision was
“self-enforcing” and therefore party whose property had been taken could resort to any common-law
remedy that would provide adequate relief for violation of provision; not addressing sovereign
immunity); \textit{Alexander v. State}, 381 P.2d 780, 781–82 (Mont. 1963) (rejecting sovereign-immunity
why the issue is relevant in the first two categories of States. This section explains why the issue is relevant to the third category.

State-court decisions holding that the State has waived its state-court immunity from just-compensation suits cannot necessarily be regarded as settling the issue, for two reasons. First, some of these decisions rest on the same mistake that many commentators have made: they fail to distinguish the issue of whether the Just Compensation Clause creates a monetary cause of action from the issue of whether that Clause overrides sovereign immunity.253 Even if that mistake were not obvious before (in defense to taking claim and holding that state constitutional guarantee of just compensation was self-executing); Nine Mile Irrigation Dist. v. State, 225 N.W. 679, 681–83 (Neb. 1929) (holding that sovereign immunity did not bar just compensation claim against State based on state constitution’s guarantee of just compensation); Sibson v. State, 282 A.2d 664, 665 (N.H. 1971) (holding that sovereign immunity did not bar claim for just compensation); DeBruhl v. State Highway & Pub. Works Comm’n, 102 S.E.2d 229, 238 (N.C. 1958) (quoting with approval case from another jurisdiction rejecting sovereign-immunity defense against payment of interest as component of just compensation); Minch v. City of Fargo, 297 N.W.2d 785, 789 (N.D. 1980) (construing state constitution’s just-compensation provision to create cause of action in inverse condemnation); Williams v. State, 998 P.2d 1245, 1248–52 (Okla. Civ. App. 2000) (describing inverse-condemnation action in Oklahoma as predominantly judicial creation); Boise Cascade Corp. v. Or. State Bd. of Forestry, 991 P.2d 563, 566–69 (Or. Ct. App. 1999) (holding that Due Process Clause of Fourteenth Amendment required state courts to entertain just-compensation claim against Oregon); Harris v. Town of Lincoln, 668 A.2d 321, 326–28 (R.I. 1995) (discussing case law holding that state and federal constitutions’ just-compensation provisions create cause of action in inverse condemnation); Horry County v. Ins. Reserve Fund, 544 S.E.2d 637, 640 (S.C. Ct. App. 2001) (stating that inverse-condemnation claim arises from constitution); Hurley v. State, 143 N.W.2d 722, 728–29 (S.D. 1966) (construing state constitution’s just-compensation provision to create inverse-condemnation remedy supplementing statutory remedies); Gen. Servs. Comm’n v. Little-Tex Insulation Co., 39 S.W.3d 591, 595–99 (Tex. 2001) (distinguishing breach-of-contract claim against State, which is governed by statute, from just-compensation claim against State, which is governed by case law construing state constitution’s just-compensation provision); Timms v. State, 428 A.2d 1125, 1126 (Vt. 1981) (citing case law establishing that state constitution’s just compensation clause overrides sovereign immunity); Wyo. State Highway Dep’t v. Napolitano, 578 P.2d 1342, 1347–51 (Wyo. 1978) (appearing to treat inverse-condemnation claim as arising from self-executing nature of state constitution’s just-compensation provision), appeal dismissed, 439 U.S. 948 (1978); cf. Waid v. Dep’t of Transp., 996 P.2d 18, 24–25 (Wyo. 2000) (suggesting that Napolitano may have been superseded by later statutes).

253. See, e.g., Dep’t of Agric. & Consumer Servs. v. Mid-Fla. Growers, Inc., 521 So. 2d 101, 103 n.2 (Fla. 1988) (holding that because of self-executing nature of state constitution’s just-compensation provision, no statutory authority was necessary for inverse-condemnation claim against State); Fielder v. Rice Constr. Co., 522 S.E.2d 13, 15 (Ga. Ct. App. 1999) (reasoning that sovereign immunity does not bar inverse-condemnation claims because inverse condemnation is form of eminent domain; citing state constitution’s just-compensation provision), cert. denied, No. S99C1722, 1999 Ga. LEXIS 1020 (Ga. Nov. 19, 1999); Renninger v. State, 213 P.2d 911, 916 (Idaho 1950) (holding that Idaho Constitution both created cause of action in inverse condemnation and waived sovereign immunity); Alexander, 381 P.2d at 781 (rejecting sovereign immunity defense in light of self-executing nature of Montana Constitution’s just-compensation provision); Hurley v. State, 143 N.W.2d 722, 729 (S.D. 1966) (reasoning that no consent to sue State was necessary given
light of decisions such as *Larson*, it is obvious now, in light of the recent plurality opinion in *Del Monte Dunes*, and it provides a strong justification for state courts to revisit the issue. Second, some state-court decisions holding that the State has waived its immunity appear to rest on the belief that a State’s immunity in its own courts exists solely as a matter of state law. Under this view, the “waiver” of state immunity is not the result of any voluntary choice by the State but is instead the result of preemption. This preemption analysis is incorrect after *Alden v. Maine*, which makes clear that a state’s immunity in its own courts is protected by the U.S. Constitution. With this error corrected, a state court might still conclude that the State had waived its immunity as a matter of state law. Even so, the incorrectness of the preemption analysis

---

254. See supra notes 32–35 and accompanying text (discussing *Larson v. Domestic & Foreign Commerce Corp.*, 337 U.S. 682 (1949)).
255. See supra notes 43–51 and accompanying text (discussing *City of Monterey v. Del Monte Dunes*, 526 U.S. 687, 714 (1999)).
256. In New Jersey, for example, the courts appear to treat sovereign immunity as a creation of state law. See *Willis v. Dep’t of Conservation & Econ. Dev.*, 264 A.3d 34, 35–38 (N.J. 1970). The New Jersey Supreme Court recently concluded that state law cannot constrain suits in inverse condemnation under *First English*. See *Greenway Dev. Co. v. Borough of Paramus*, 750 A.2d 764, 769–71 (N.J. 2000) (holding that suits in inverse condemnation were not subject to requirements of state tort claims statute). Other states have come to a similar conclusion. See e.g., *Jacoby v. Arkansas*, 962 S.W.2d 773, 775–78 (Ark. 1998) (holding that federal statute authorizing private actions in state court overrode State’s sovereign immunity under Arkansas Constitution), *vacated*, 527 U.S. 1031 (1999); *Foss*, 309 A.2d at 343–44 (holding that, while state legislature can authorize state agency to commit trespass or create nuisance, that power is subject to just compensation guarantee of federal constitution); *Alper v. Clark County*, 571 P.2d 810, 811–13 ( Nev. 1977) (holding that, because inverse-condemnation claim vindicated federal constitutional rights, claim was not subject to state statutory restrictions on suits against counties).
257. See, e.g., *Alden v. Maine*, 527 U.S. 706, 730 (1999) (saying that state sovereign immunity “inheres in the system of federalism established by the Constitution”), see also supra notes 207–210 (discussing *Alden’s* elimination of preemption rationale for concluding that just-compensation principle overrides sovereign-immunity principle).
warrants reconsideration of the issue. 258

In short, the issue addressed in this Article has direct relevance even though many States have mitigated the harshness of sovereign immunity by waiving their immunity to some extent. 259 In many States, the waiver is incomplete or unclear (or both). Moreover, although this Article proposes an analysis specifically for evaluating just-compensation claims against states, the analysis also has implications for other claims, as discussed in the sections that follow.

B. State Remedial Obligations Under the Due Process Clause for Deprivations Other than Takings of Private Property for Public Use

The analysis proposed in Part IV would sometimes subject States to just-compensation suits brought in their own courts, even though they would be immune if those suits were brought in federal court. The States’ exposure to just-compensation suits, however, is not necessarily the most significant feature of the proposed analysis. The proposed analysis extends beyond claims based on the Just Compensation Clause, because it does not rest solely on the Just Compensation Clause. 260 Rather, it rests on the interaction of that Clause and the procedural component of the Due Process Clause. This Article proposes that the remedial requirements of procedural due process can expose an unconsenting State to suits in its own courts that would be barred by sovereign immunity in federal court. That proposition includes, but is not necessarily limited to, suits for just compensation, as this section explains.

258. Cf. GA. CONST. art. 1, § 2, ¶ IX(f) (providing that waiver of sovereign immunity prescribed in that provision should not be “construed as a waiver of any immunity provided . . . by the United States Constitution”).


260. Cf. Beermann, supra note 18, at 315 (arguing that “[s]overeign and official immunities must be overruled by the takings clause”); supra note 19 (citing commentary asserting, in light of First English, that Just Compensation Clause, standing alone, overrides sovereign immunity).
1. The Error-Remediation Element of Procedural Due Process

At the heart of this Article is the Court’s repeated statement that both the States and the federal government must make a “reasonable, certain and adequate provision” for paying just compensation. This statement prescribes more than a substantive obligation to pay; it also demands the existence of an adequate procedure for honoring that substantive obligation. The Court has never identified the legal source of this procedural requirement. In my view, it stems from the procedural component of the Due Process Clause—i.e., from the doctrine of procedural due process.

Procedural due process requires States to use adequate procedures when they deprive people of life, liberty, or property. This is not just procedure for procedure’s sake. Rather, a key purpose of the procedures is to ensure that these deprivations are accurate—i.e., that

261. E.g., Williamson County Reg’l Planning Comm’n v. Hamilton Bank, 473 U.S. 172, 194 (1985); see also supra notes 181–205 and accompanying text (discussing requirement that States make “reasonable, certain, and adequate” provisions for paying just compensation).

262. See City of Monterey v. Del Monte Dunes at Monterey, Ltd., 526 U.S. 687, 717 (1999) (stating that government can violate Constitution “by denying just compensation in fact or by refusing to provide procedures through which compensation may be sought”).

263. The Court has construed the Due Process Clause of the Fourteenth Amendment to do three things.

First, it incorporates many of the specific protections defined in the Bill of Rights. Second, it contains a substantive component that bars certain arbitrary, wrongful government actions regardless of the fairness of the procedures used to implement them. Third, it provides a guarantee of fair procedure.


they comport with both federal and state law.\textsuperscript{267} This accuracy goal serves the systemic purpose of preserving the rule of law and the more particularized purpose of protecting the substantive rights of individuals in their life, liberty, and property.\textsuperscript{268}

To achieve accuracy, procedures must not only avoid errors but also provide a way to remedy the errors that inevitably occur.\textsuperscript{269} When an error is detected before it causes a deprivation of life, liberty, or property, the usual remedy is to prevent the deprivation from occurring, such as by an injunction.\textsuperscript{270} When the error is detected after it has caused a deprivation, some other remedy—including, in some situations, an award of compensatory damages—may be required to fulfill the accuracy goal of the Due Process Clause.\textsuperscript{271} In short, the doctrine of procedural due process contemplates that procedures will serve two functions: error-

\begin{itemize}
\item \textsuperscript{267} See, e.g., Cleveland Bd. of Educ. v. Loudermill, 470 U.S. 532, 543 (1985) (holding that procedural due process entitled municipal employee to pre-termination procedures to ensure that termination was based on accurate facts and was otherwise appropriate); Goss v. Lopez, 419 U.S. 565, 581 (1975) (holding that Due Process Clause entitled public school student to "rudimentary" procedures before being suspended from school to prevent suspensions based on "unfair or mistaken findings of misconduct"); Daniel S. Feder, Note, From Parratt to Zinermon: Authorization, Adequacy, and Immunity In a Systemic Analysis of State Procedure, 11 CARDOZO L. REV. 831, 844–45 (1990) (identifying accuracy as main goal of procedural due process, including accuracy in the "faithful implementation of state substantive laws").
\item \textsuperscript{268} Cf. Fallon, Confusions, supra note 190, at 311 ("The ultimate commitment of the law of due process remedies—alogous to that of procedural due process—is to create schemes and incentives adequate to keep government, overall and on average, tolerably within the bounds of law."); id. at 338 (stating that the Constitution in general and the Due Process Clause in particular do sometimes require individually effective remediation for constitutional violations").
\item \textsuperscript{269} See, e.g., Zinermon v. Burch, 494 U.S. 113, 126 (1990) (stating that inquiry under doctrine of procedural due process "would examine the procedural safeguards built into the statutory or administrative procedure of effecting the deprivation, and any remedies for erroneous deprivations provided by statute or tort law").
\item \textsuperscript{270} See Gen. Oil Co. v. Crain, 209 U.S. 211, 221–28 (1908) (holding that Court had jurisdiction to review state supreme court decision dismissing, on sovereign immunity grounds, suit against state officer challenging constitutionality of state tax law); Seamon, Sovereign Immunity, supra note 104, at 344–45, 381–83, 397–98 (discussing Crain and ultimately proposing that it be understood as a decision resting on Due Process Clause); see also Porter v. Investors Syndicate, 286 U.S. 461, 469–71 (1932) (holding that, to avoid due-process concerns, state statute should be construed to let state courts enjoin state administrative order revoking plaintiff’s permit to do business in State).
\item \textsuperscript{271} See, e.g., McKesson Corp. v. Div. of Alcoholic Beverages & Tobacco, 496 U.S. 18, 39 (1990) (holding that, when State chooses to remit taxpayers to post-collection relief from erroneous taxation, due process requires state to "provide taxpayers with, not only a fair opportunity to challenge the accuracy and legal validity of their tax obligation, but also a clear and certain remedy for any erroneous or unlawful tax collection") (internal quotation marks, citation, and footnote omitted); cf. Booth v. Churner, 121 S. Ct. 1819, 1823–24 (2001) ("[D]epending on where one looks, ‘remedy’ can mean either specific relief obtainable at the end of a process of seeking redress, or the process itself.").
\end{itemize}
avoidance and error-remediation.272

As between error-avoidance and error-remediation, the error-avoidance function is probably better known, perhaps because it was emphasized in the well-known decision Mathews v. Eldridge.273 In Mathews, the Court said that its precedent considered three factors in evaluating the adequacy of a procedure that deprived someone of life, liberty, or property:

[first, the private interest that will be affected by the official action; second, the risk of an erroneous deprivation of such interest through the procedures used, and the probable value, if any, of additional or substitute procedural safeguards; and finally, the Government’s interest, including the function involved and the fiscal and administrative burdens that the additional or substitute procedural requirement would entail.]274

In describing the second factor, the Court seemed primarily concerned with minimizing the risk of error—i.e., in error-avoidance. Given that concern, the three-factor inquiry as a whole seems to stress enhanced error-avoidance as the relevant goal of additional procedural safeguards (while also making clear that, even if additional procedural safeguards would minimize errors, they may not be constitutionally required in light of the relevant individual and governmental interests).

One of the best-known cases addressing the error-remediation (as distinguished from the error-avoidance) function of procedural due process is Parratt v. Taylor.275 In Parratt, the Court held that the requirements of procedural due process were satisfied by a state tort remedy for a prison official’s allegedly negligent loss of a prisoner’s property (a hobby kit).276 That holding emphasized not only a tort action’s potential for determining whether the deprivation was erroneous

---

272. Professor Fallon has argued that the post-deprivation remedial demands of the Due Process Clause should be considered distinct from the pre-deprivation procedures traditionally associated with “procedural due process.” See Fallon, Confusions, supra note 190, at 329–40. Without disagreeing with that argument, this Article discusses the remedial demands of the Due Process Clause as an aspect of procedural due process, in keeping with the Court’s current approach. See, e.g., Zinermon, 494 U.S. at 126 (stating that inquiry under doctrine of procedural due process “would examine . . . any remedies for erroneous deprivations’’); Parratt v. Taylor, 451 U.S. 527, 537 (1981) (“[W]e must decide whether the tort remedies which the State of Nebraska provides as a means of redress for property deprivations satisfy the requirements of procedural due process.”) (emphasis added).

273. 424 U.S. 319 (1976); see also supra note 266 (citing precedent that describes accuracy goal of procedural due process in terms of minimizing error).

274. Mathews, 424 U.S. at 335.


276. Id. at 536–44.
but also its potential for remedying an erroneous deprivation.\textsuperscript{277} In later cases, the Court confirmed that procedural due-process analysis examines not only the “procedural safeguards” against erroneous deprivations but also “any remedies for erroneous deprivations provided by statute or tort law.”\textsuperscript{278}

This Article relies principally on the error-remediation element of procedural due process. Of course, procedural due process requires a State to have an adequate process for accurately identifying instances in which it has taken private property for public use (this is an important requirement, considering how hard it can be accurately to identify regulatory takings). It is the error-remediation element of procedural due process, however, that sometimes overrides sovereign immunity by requiring state courts to award just compensation payable from the state treasury if there is no adequate, alternative compensation scheme. In short, a State meets its procedural due process obligations only when, having identified an instance in which private property has been taken for public use, it also provides the remedy—i.e., payment of just compensation—that is necessary to prevent that taking from being unconstitutional (and hence erroneous).

Because of this Article’s reliance on the error-remediation element of the doctrine of procedural due process, I must address commentary that denies the validity of that element. In particular, Professors Wells and Eaton have argued that the Court was wrong, and departed from precedent, when in \textit{Parratt} and later cases it construed the doctrine of procedural due process sometimes to require States to provide post-

---

\textsuperscript{277} See id. at 543–44 (observing that post-deprivation tort action provided a “means by which [the plaintiff] can receive redress for the deprivation” and be “fully compensated . . . for the property loss”).

\textsuperscript{278} Zinermon, 494 U.S. at 126; see id. at 125 (“the existence of state remedies is relevant” to procedural due-process analysis) (emphasis in original); id. at 129 (stating that, in \textit{Mathews}, “it was clear that the State, by making available a tort remedy that could adequately redress the loss, had given the prisoner the process he was due”); see also Fla. Prepaid Postsecondary Educ. Expense Bd. v. Coll. Sav. Bank, 527 U.S. 627, 643 (1999) (citing \textit{Parratt} for proposition that State can satisfy procedural due process requirement by providing post-deprivation remedy); Albright v. Oliver, 510 U.S. 266, 285 (1994) (Kennedy, J., concurring) (recognizing that “\textit{Parratt} rule” stems from procedural due process and would be satisfied in that case by availability of post-deprivation state tort suit for malicious prosecution); Williamson County Reg’l Planning Comm’n v. Hamilton Bank, 473 U.S. 172, 195 (1985) (relying on \textit{Parratt} to hold that violation of Just Compensation Clause was not complete until state denied just compensation); Hudson v. Palmer, 468 U.S. 517, 533 (1984) (relying on \textit{Parratt} to “hold that an unauthorized intentional deprivation of property by a state employee does not constitute a violation of the procedural requirements of the Due Process Clause of the Fourteenth Amendment if a meaningful postdeprivation remedy for the loss is available”); \textit{Logan}, 455 U.S. at 435–37 (holding that State violated procedural due process in that case; distinguishing \textit{Parratt}, while clearly treating it as valid procedural due process precedent).
deprivation remedies for erroneous deprivations; in their view, any requirement of a post-deprivation remedy must stem, if anywhere, from the doctrine of substantive due process. If they are right, the analysis proposed in this Article may be wrong, because the doctrine of substantive due process would be an odd basis for overriding a State’s sovereign immunity from just-compensation claims. When a court concludes that government action violates substantive due process, that conclusion ordinarily requires invalidation of the action. In contrast, when a court concludes that government action constitutes a taking of private property for public use, that conclusion does not invalidate the government action; it means only that the government must pay for the property that has been taken. Thus, the doctrine of substantive due process and the doctrine of just compensation work somewhat at cross-purposes.

In any event, I respectfully disagree with Professors Wells’ and Eaton’s view that Parratt and its progeny erroneously depart from procedural due process precedent. Precedent amply supports the Court’s conclusion in Parratt and later cases that procedural due process can require a state to provide post-deprivation remedies for erroneous deprivations. Support for the Court’s procedural due-process analysis

280. See, e.g., Foucha v. Louisiana, 504 U.S. 71, 80 (1992) (“[T]he due process clause contains a substantive component that bars certain arbitrary, wrongful government actions regardless of the fairness of the procedures used to implement them.”) (internal quotation marks and citation omitted) (emphasis added).
281. See, e.g., First English Evangelical Lutheran Church v. County of Los Angeles, 482 U.S. 304, 314–15 (1987) (stating that Just Compensation Clause “does not prohibit the taking of private property, but instead places a condition on the exercise of that power,” and that Clause “is designed not to limit the governmental interference with property rights per se, but rather to secure compensation”).
283. See also Feder, supra note 267, at 840 (observing that “courts and commentators consistently refuse to accept the Supreme Court’s own characterization” of its decision in Parratt as resting on procedural due process but defending Court’s characterization).
284. See Parratt v. Taylor, 451 U.S. 527, 543–44 (1981); see also supra note 278 (citing later cases in which Court described “Parratt rule” as making existence of state post-deprivation remedies relevant to procedural due-process analysis).
comes from: (1) *Ingraham v. Wright*, in which the Court held that procedural due process was satisfied by the availability of a state tort remedy against public school teachers who wrongfully imposed corporal punishment on their students; (2) a line of cases the most famous of which is *North American Cold Storage Co. v. City of Chicago*, in which the Court rejected procedural due process challenges to government deprivations that justifiably, in light of some exigency, occurred without meaningful pre-deprivation procedures but afforded post-deprivation remedies; and (3) the line of cases exemplified by *McKesson Corp. v. Division of Alcoholic Beverages & Tobacco*, in which the Court construed the doctrine of procedural due process to require States to provide either a pre-deprivation or post-deprivation remedy for erroneously collected taxes.

Professors Wells and Eaton find *Ingraham* and the *North American Cold Storage* line of cases inapposite because they involved government deprivations that under some circumstances would be justified. For example, *Ingraham* involved corporal punishment of public school students, which “may help maintain order in the classroom”; *North American Cold Storage* involved the government’s summary quarantining of food that officials suspected (perhaps erroneously) was unwholesome. Both cases thus presumed that the challenged government conduct could be justified under certain circumstances—i.e., if the beaten child was misbehaving and the confiscated food was unwholesome. *Parratt*, in contrast, involved the allegedly negligent loss of the plaintiff’s property by a government official, conduct that, the plaintiff claimed, could never be justified.

But just as the government can (according to the Court) properly beat a child who is believed to have misbehaved and confiscate food that is

286. Id. at 682.
287. 211 U.S. 306 (1908).
288. Id. at 314–19.
290. Id. at 31–51; see also supra notes 217–26 and accompanying text (discussing *McKesson* line of cases) and infra notes 295–97 (same).
293. Wells & Eaton, supra note 279, at 220 & n.85.
believed to be contaminated, it can properly destroy a prisoner’s property if, for example, the property is believed to be contraband. Those beliefs may turn out to be erroneous, and for that reason, the Court in cases involving these actions, including *Parratt*, construed procedural due process to require post-deprivation remedies. *Parratt* is not distinguishable from *Ingraham* and *North American Cold Storage*, however, merely because a prison official’s destruction of a prisoner’s property would violate substantive due process if motivated by sheer malice. So too, the malicious beating of a school student who was not believed to be misbehaving, and the malicious destruction of food that was not believed to be contaminated, would violate substantive due process. All this means is that, depending on the justification (or lack thereof) for governmental deprivation of liberty or property, the victim of the deprivation may be able to assert a substantive due process claim or only a procedural due process claim. But the issue of justification does not distinguish *Ingraham* and *North American Cold Storage* from *Parratt*.

Professor Wells finds the *McKesson* line of cases inapposite because, in light of the federal Tax Injunction Act, 295 those cases concern a state-court remedy that “may be the only recourse for unconstitutionally collected taxes.” 296 The problem with this basis for distinguishing the Court’s reliance on procedural due process in the *McKesson* line of cases from its reliance on procedural due process in *Parratt* and its progeny is that the Court itself in the *McKesson* line of cases did not rely on the unavailability of federal-court remedies in concluding that due process mandated a state-court remedy for erroneous taxes. More fundamentally, as argued above, the availability of a federal-court remedy does not excuse a State’s failure to provide its own remedies to the extent required by the Due Process Clause. 297

In sum, the Court has held that procedural due process requires

295. 28 U.S.C. § 1341 (1994) (“The [federal] district courts shall not enjoin, suspend or restrain the assessment, levy or collection of any tax under State law where a plain, speedy and efficient remedy may be had in the courts of such State.”).

296. Michael Wells, *Suing States*, supra note 18, at 778 (“The constitutionally-implied [*sic*] remedy issue arises in the tax context and not elsewhere because other remedies, including § 1983 suits, are typically available for other constitutional violations, while a state court [tax] refund action, implied directly from the Due Process Clause, may be the only recourse for unconstitutionally collected taxes.”); see also id. at 793 (“But an equally essential premise for the state court cause of action recognized in *Reich* [a decision following *McKesson*] is that Congress may validly bar taxpayer access to the federal courts, as it has by the tax injunction act.”); id. at 800 (describing principle established in *Reich* to be that “states must provide effective remedies for constitutional violations when federal remedies are not available”).

297. See supra notes 241–48 and accompanying text.
procedures that not only avoid but also remedy erroneous deprivations of life, liberty or property. This Article contends that the remedial element of procedural due process underlies a State’s obligation to create a procedural scheme for awarding just compensation when it takes private property for public use. If a State does not create an adequate alternative to just-compensation suits against itself in its own courts, it cannot rely on sovereign immunity to defeat those suits.

2. The Interaction of the Remedial Element of Procedural Due Process and the Just Compensation Clause

Although procedural due process requires error-remediation, it does not require a remedy for every error, nor does it require the same type of remedy for every type of error.\(^{298}\) This Article has focused on deprivations that consist of governmental takings of private property for public use. Such a deprivation is erroneous only when it occurs without the payment of just compensation. Once that error (the failure to pay just compensation) is identified, the nature and invariability of the remedy—the payment of just compensation for every such taking—is dictated by the text of the Just Compensation Clause.\(^{299}\) Thus, it is procedural due process that requires some remedy, but it is the Just Compensation Clause that specifies what the remedy must be in every instance.

The Court has held that the procedural component of the Due Process Clause also requires a compensatory remedy—but does not invariably do so—for the erroneous collection of taxes.\(^{300}\) The collection of taxes implicates procedural due process because it deprives taxpayers of their property (namely, their money).\(^{301}\) Procedural due process accordingly requires the State to have adequate procedures to ensure that the taxes are collected accurately—i.e., consistently with state and federal law. This

\(^{298}\) See, e.g., Martinez v. California, 444 U.S. 277, 284 n.9 (1980) (“It must be remembered that even if a state decision does deprive an individual of life or property, and even if that decision is erroneous, it does not necessarily follow that the decision violated that individual’s right to due process.”).

\(^{299}\) See First English Evangelical Lutheran Church v. County of Los Angeles, 482 U.S. 304, 316 (1987) (“in the event of a taking, the compensation remedy is required by the Constitution”); San Diego Gas & Elec. Co. v. City of San Diego, 450 U.S. 621, 658 n.24 (1981) (Brennan, J., dissenting) (“[T]his is not a case involving implication of a damages remedy—the words of the Just Compensation Clause are express.”).

\(^{300}\) See supra notes 217–32 and accompanying text (arguing that Court’s tax cases support conclusion that Due Process Clause can override States’ sovereign immunity from just-compensation claims).

State Sovereign Immunity

requirement includes the need for procedures that prospectively prevent or retroactively remedy erroneous collections. As to the timing of those procedures, “the States may determine whether to provide predeprivation process (e.g., an injunction) or instead to afford post deprivation relief (e.g., a refund).”302 When a State chooses to provide post-deprivation relief, instead of pre-deprivation relief, for erroneous taxes, a refund is not invariably required. For example, if a state tax violates the Commerce Clause by discriminating between out-of-state taxpayers and their in-state competitors, the State might be able to satisfy the Due Process Clause by retroactively increasing the taxes of the in-state competitors.303 In contrast, procedural due process would require a refund of state taxes that were erroneously collected from people who were exempt from the taxes under a federal statute.304 Thus, the remedial requirements imposed by the doctrine of procedural due process, like other requirements imposed by that doctrine, depend partly on the nature of the individual harm caused by an erroneous deprivation.305

For this reason, the Due Process Clause might sometimes require a State to refund taxes that had been collected in violation of state, rather than federal, law.306 The nature of a taxpayer’s harm does not vary depending on whether the State has erroneously collected taxes from him in violation of (say) a state-law exemption or a federal-law exemption. Of course, there may be considerations besides individual harm that would justify construing the Due Process Clause to require differing remedies for federal-law and state-law violations, including the role of those remedies in ensuring the supremacy of federal law.307 The point is

304. See Ward v. Bd. of County Comm’rs, 253 U.S. 17, 22–25 (1920); see also McKesson, 496 U.S. at 39 (citing Ward as example of situation in which State “would have had no choice” but to refund taxes).
307. See McKesson, 496 U.S. at 105 (stating that due process “requires a clear and certain remedy for taxes collected in violation of federal law”) (emphasis added); cf. Pennhurst State Sch. & Hosp. v. Halderman, 465 U.S. 89, 114 (1984) (stating that Court’s development of Ex parte Young doctrine had reflected “accommodation” between “the need to promote the supremacy of federal law” and “the constitutional immunity of the States,” and holding that doctrine does not authorize federal-court relief for violations of state law); Jackson, The Supreme Court, supra note 18, at 52–61 (arguing that federal courts should be more willing to grant relief for States’ violations of federal law than for their violations of state law); Woolhandler, Old Property, supra note 225, at 937
that, if a state court decided that a state tax had been collected from a taxpayer in violation of state law, but denied a refund as barred by sovereign immunity, the taxpayer could plausibly argue that the state court’s decision violated the doctrine of procedural due process, and the argument would pose a question about the remedial requirements of that doctrine. 308

The Court’s tax cases show that the remedial requirement of procedural due process can override state sovereign immunity, but the cases should not be read to mean that the remedial requirement always does. Although the doctrine of procedural due process requires States to create a system for avoiding and remedying erroneous deprivations of life, liberty, and property—the doctrine does not require the system to detect all errors, 309 to remedy all errors detected, 310 or to make the post-deprivation remedy required for an erroneous deprivation a judicial award of money from the sovereign’s treasury. 311 That is because the

308. See Cent. of Ga. Ry. v. Wright, 207 U.S. 127, 136–42 (1907) (holding that Due Process Clause was violated by state laws that authorized state taxes to be assessed against taxpayer without giving taxpayer a chance to challenge official’s valuation of taxable property).

309. See, e.g., Walters v. Nat’l Ass’n of Radiation Survivors, 473 U.S. 305, 320 (1985) (due process “does not require that the procedures used to guard against an erroneous deprivation . . . be so comprehensive as to preclude any possibility of error”) (internal quotation marks omitted).

310. See, e.g., Martinez v. California, 444 U.S. 277, 284 n.9 (1980) (“It must be remembered that even if a state decision does deprive an individual of life or property, and even if that decision is erroneous, it does not necessarily follow that the decision violated that individual’s right to due process.”); Mackey v. Montrym, 443 U.S. 1, 11 (1979) (holding that due process did not require hearing before suspension of drivers’ license, while recognizing that post-suspension detection of error would not yield a complete remedy); Bd. of Curators v. Horowitz, 435 U.S. 78, 92 & n.8 (1978) (indicating that state medical school’s decision did not violate due process even if it violated school’s own rules); Ingraham v. Wright, 430 U.S. 651, 694 (1977) (White, J., dissenting) (observing that, under majority’s holding, public school student who suffered corporal punishment on basis of mistaken facts would have no remedy as long as officials acted reasonably); Fallon, Confusions, supra note 190, at 311:

The dictum of Marbury v. Madison notwithstanding, there is no right to an individually effective remedy for every constitutional violation. The ultimate commitment of the law of due process remedies—analagous to that of procedural due process—is to create schemes and incentives adequate to keep government, overall and on average, tolerably within the bounds of law.

311. See Gilbert v. Homar, 520 U.S. 924, 929–36 (1997) (holding that State did not violate doctrine of procedural due process by suspending plaintiff without pay, even though this entailed loss of income during period of suspension); see also Hague v. Comm. for Indus. Org., 307 U.S. 496, 529 (1939) (“There are many rights and immunities secured by the Constitution . . . which are not capable of money valuation . . . .”); Walter E. Dellinger, Of Rights and Remedies: The
doctrine’s demand for error-avoidance and error-remediation vary depending on the individual and governmental interests at stake.\textsuperscript{312} For example, the Due Process Clause may demand greater accuracy for permanent deprivations of property than for temporary ones.\textsuperscript{313} Furthermore, the invariability with which a remedy is required for an erroneous deprivation may depend not only on the nature of the individual harm caused by an erroneous deprivation but also on other factors, such as whether or not the error is one of federal law; if an error is one of federal law, whether it is a statutory error or a constitutional one; and, if the error is constitutional, what type of constitutional right is affected.\textsuperscript{314}

An illustrative case is \textit{Ingraham v. Wright},\textsuperscript{315} in which the Court held that due process does not require procedural safeguards before a public school official beats a student.\textsuperscript{316} (Corporal punishment implicates procedural due process because it interferes with the student’s liberty interest in being free from physical restraint and pain.)\textsuperscript{317} The \textit{Ingraham} Court held that due process was satisfied by the availability of a post-punishment state tort suit against the responsible officials.\textsuperscript{318} Obviously, these tort suits would not detect every erroneous beating. Moreover, as the dissent observed, tort suits would not ensure a remedy for every

\textit{Constitution as a Sword}, 85 HARV. L. REV. 1532, 1544 (1972) (asserting that, contrary to implication in \textit{Bivens}, 403 U.S. at 395, the federal courts’ “exercise of remedial power to create a damage action directly from the Constitution [was] virtually unprecedented”); John C. Jeffries, Jr., \textit{Disaggregating Constitutional Torts}, 110 YALE L.J. 259, 279 (2000) (“The strictures in the Constitution were not conceived as predicates for money damages.”).

\textsuperscript{312} See, for example, \textit{Mathews v. Eldridge}, 424 U.S. 319, 335 (1976), where the Court described precedent on procedural due process to require consideration of:

\begin{itemize}
  \item First, the private interest that will be affected by the official action;
  \item second, the risk of an erroneous deprivation of such interest through the procedures used, and the probable value, if any, of additional or substitute procedural safeguards; and
  \item finally, the Government’s interest, including the function involved and the fiscal and administrative burdens that the additional or substitute procedural requirement would entail.
\end{itemize}

\textsuperscript{313} See, e.g., \textit{Gilbert}, 520 U.S. at 932 (stating that procedural due-process analysis considers finality of deprivation); \textit{Mathews}, 424 U.S. at 340 (determining that individual interest at stake in that case was in uninterrupted flow of benefits, rather than in absolute amount of benefits, because, if termination found erroneous, benefits would be awarded retroactively).

\textsuperscript{314} Cf. Jeffries, \textit{supra} note 311, at 280 (arguing that “the [official] liability rule for money damages should vary with the constitutional violation at issue”).

\textsuperscript{315} 430 U.S. 651 (1977).

\textsuperscript{316} \textit{Id.} at 683.

\textsuperscript{317} \textit{See id.} at 673–74.

\textsuperscript{318} \textit{Id.} at 683 (concluding that due process was satisfied by state’s “preservation of common-law constraints and remedies”).
beating that was found, in those suits, to have been erroneous.\textsuperscript{319}

As with erroneous tax collections, the need for a compensatory remedy for an erroneous liberty deprivation may depend on whether the error is one of state law or federal law.\textsuperscript{320} For example, compensation may not be constitutionally required for the beating of a student that was erroneous because it violated a state statute.\textsuperscript{321} In contrast, compensation may be required if the beating violated substantive due process because of its arbitrariness.\textsuperscript{322} The differing results would reflect that the remedial demands of the Due Process Clause can be influenced by the Supremacy Clause.\textsuperscript{323}

In short, the remedial element of procedural due process produces some hard cases, but cases involving governmental takings of private property for public use are not among them. It is difficult to explain, for example, the case law indicating that compensation from the State is often required for erroneous taxation but seldom for erroneous liberty deprivations.\textsuperscript{324} In contrast, there is an explanation for the nature and invariability of the remedy required for governmental takings of private property for public use: Although the Due Process Clause requires a remedy for that sort of deprivation, the nature and invariability of that remedy (just compensation for every such taking) is prescribed by the Just Compensation Clause. The Constitution does not prescribe the

\textsuperscript{319} Id. at 693–94 & n.11 (White, J., dissenting).

\textsuperscript{320} See supra notes 306–08 and accompanying text.

\textsuperscript{321} Cf. Paul v. Davis, 424 U.S. 693, 701 (1976) (stating that Fourteenth Amendment should not be interpreted as a “font of tort law”).

\textsuperscript{322} Cf. County of Sacramento v. Lewis, 523 U.S. 833, 848–51 (1998) (explaining that doctrine of substantive due process does not invariably create constitutional liability for official conduct that would constitute a tort).

\textsuperscript{323} The Supremacy Clause can likewise influence whether a State’s remedial procedures must include judicial review. See supra note 190 (observing that, even if Due Process Clause would not, standing alone, require state to involve its courts in adjudicating just-compensation claims, Supremacy Clause may mandate judicial review in some circumstances).

\textsuperscript{324} See Fallon & Meltzer, supra note 18, at 1825–28 (discussing possible reasons for differences between Court’s decisions in tax cases and its decisions in other due process cases and finding none of them satisfactory). It is likewise difficult for this author to understand the Court’s conclusion that a government official’s negligence can never constitute a “depriv[ation]” within the meaning of the Due Process Clause. See Davidson v. Cannon, 474 U.S. 344, 348 (1986) (“[T]he protections of the Due Process Clause, whether procedural or substantive, are just not triggered by lack of due care by prison officials.”); see also Daniels v. Williams, 474 U.S. 327, 329–36 (1986) (holding that Due Process Clause is not implicated by a state official’s negligent destruction of property because such conduct does not “deprive” a person of the property, within the meaning of the Due Process Clause). In my view, if the negligence is caused by inadequate procedures (for example, the procedures for training or supervising the negligent official), the Court should recognize a deprivation.
remedies required for other erroneous deprivations of life, liberty, or property. In the absence of textual guidance, difficult questions arise.325


This section has not essayed a grand theory of due process remedies or even suggested that such a theory is appropriate.326 Rather, it has discussed the implications of Part IV’s reliance on the doctrine of procedural due process, plus the Just Compensation Clause, as the legal bases for overriding state sovereign immunity from just-compensation claims. The discussion has emphasized that procedural due process obligates States to have procedures designed not only to avoid but to remedy erroneous deprivations of life, liberty, or property. There is no single answer to the question of what remedy the Due Process Clause

325. Although the Just Compensation Clause has been interpreted to require compensation for every taking of private property for public use, this interpretation does not necessarily give the right to just compensation privileged status among constitutional rights. This becomes clear when one considers that injunctive relief is not available to prevent the government from taking private property for public use. See Seamon, Separation of Powers, supra note 59, at 210–11. In contrast, injunctive relief is often available to protect other constitutional rights, such as those secured by the Fourth Amendment. See, e.g., City of Indianapolis v. Edmond, 531 U.S. 32, 36 (2000) (upholding preliminary injunction against drug checkpoint program found to violate Fourth Amendment). By the same token, the government could not force someone to submit to an unreasonable search and seizure by tendering just compensation. Cf. Brauneis, supra note 20, at 61 (observing that 19th century case law invalidated laws that authorized uncompensated takings, treating them similarly to warrants issued without probable cause, in violation of Fourth Amendment); Dellinger, supra note 311, at 1562–63 (arguing that protection of Fourth Amendment would be weakened if exclusionary rule were abolished in favor of requiring government compensation for Fourth Amendment violations). In this respect, as Professor Brauneis has observed, just-compensation rights are "uniquely weak"; they provide only liability-rule protection instead of the stronger property-rule protection provided by other constitutional rights. See E-Mail from Robert Brauneis, Associate Professor of Law, The George Washington University Law School, to Richard H. Seamon, Assistant Professor of Law, Univ. of S.C. (Jan. 24, 2001) (on file with author); see also Brauneis, supra note 20, at 113 (observing that, in contrast to governmental takings, which trigger Just Compensation Clause’s provision for monetary relief, violations of other constitutional provisions that flatly and unqualifiedly limit government power "should arguably be seen as more serious than the mere failure to pay compensation for what could be forcibly taken so long as it was paid for"); Robert Brauneis, "The Foundation of Our ‘Regulatory Takings’ Jurisprudence”: The Myth and Meaning of Justice Holmes’s Opinion in Pennsylvania Coal v. Mahon, 106 YALE L.J. 613, 672 n.268 (1996) (discussing distinction between a “liability rule” and “property rule” in context of takings jurisprudence); Jackson, The Supreme Court, supra note 18, at 93 (observing that, because monetary awards for government’s violation of certain rights could be perceived as “depreciat[ing] the value of those rights,” injunctive relief is preferable for such violations).

326. Cf. Jeffries, supra note 311, at 259 (arguing that it is not appropriate to determine the availability of money damages from state officials under 42 U.S.C. § 1983 without considering the nature of the constitutional violation for which the damages are sought).
requires for a State’s erroneous deprivation of life, liberty, or property. The Court’s decisions make clear, however, that for some deprivations—including, but not limited to, those involving takings of private property for public use—the Clause requires compensation. This Article argues that, when that is the remedy required by the Fourteenth Amendment, and when it is sought in a state court of otherwise competent jurisdiction, a state court cannot deny the remedy because of sovereign immunity if the State has created no adequate alternative for providing just compensation. This is true even though that judicial remedy would be barred by the State’s sovereign immunity if it were sought in a federal court. Thus, the Due Process Clause creates asymmetries between a State’s immunity in its own courts and its immunity in federal court.

C. The United States’ Obligations Under the Due Process Clause of the Fifth Amendment

Just as the analysis offered here is not limited to claims for just compensation, it is not limited to claims against the States. The United States has remedial obligations under the Due Process Clause of the Fifth Amendment. Under that Clause, when the federal government takes private property for public use, it, like the States, must have made “reasonable, certain, and adequate provision for obtaining compensation.” Currently, the federal government meets this obligation by authorizing just compensation suits to be brought against it

327. Cf. Hart, supra note 244, at 1366–67 (discussing breadth of congressional discretion to prescribe remedies consistent with Constitution).

328. I thus join Professor Jackson, and part company with Professor Vázquez, in believing that a State cannot always satisfy the Fourteenth Amendment by letting its officers be sued. See Vicki C. Jackson, Principle and Compromise in Constitutional Adjudication: The Eleventh Amendment and State Sovereign Immunity, 75 NOTRE DAME L. REV. 953, 979 n.91 (2000) (stating that she is “not yet persuaded by the argument that the Due Process Clause of the 14th Amendment could always be satisfied with relief against officers”); cf. Vázquez, What Is, supra note 18, at 1770 (proposing reinterpretation of McKesson “as establishing that individuals injured by a state’s violation of mandatory federal obligations have a right to damages not from the state itself, but from state officials”); Vázquez, Alden Trilogy, supra note 18, at 1947–48 (asserting that Court’s due process precedent “could be reconciled, and other doctrinal conundrums solved, if McKesson were interpreted as holding that the remedy required by the Due Process Clause is a remedy against state officials rather than the state itself”).

329. U.S. CONST. amend. V.

330. See supra note 181 (citing cases articulating the “reasonable, certain, and adequate” standard, many of which involved asserted takings by federal government).
State Sovereign Immunity

under the Tucker Act. As discussed above, some early decisions of the Court—including Schillinger and Lynch—have been read to suggest that, in the absence of the Tucker Act, the United States would have sovereign immunity from just-compensation suits. As also discussed above, however, the Schillinger line of cases has been undermined by later precedent, and Lynch is better read as prohibiting, rather than authorizing, the United States to use sovereign immunity to prevent payment of just compensation. Under the analysis proposed here, the Due Process Clause would bar Congress from repealing the Tucker Act unless there were an adequate, alternative procedure by which the federal government would pay just compensation for private property that it took for public use.

The Court hinted as much in Russian Volunteer Fleet v. United States. In that case, a federal agency appropriated two shipbuilding contracts, as well as the completed ships, from a Russian corporation. The federal statute authorizing the appropriation also authorized people whose property had been appropriated to sue the United States for just compensation. The Court rejected the government’s argument that the provision authorizing these suits did not permit suits to be brought by companies incorporated in a country the government of which was not recognized by the United States. The Court determined that the government’s interpretation, “to say the least, would raise a grave question as to the constitutional validity of the act.” That determination implies that the Constitution requires the United States to have an adequate scheme for paying just compensation when it takes private property for public use.

The same implication arises from the Court’s decisions on the

331. 28 U.S.C. §§ 1346(a), 1491 (1994); see also Reg’l Rail Reorg. Act Cases, 419 U.S. 102, 149 (1974) (holding that availability of Tucker Act remedy for taking alleged in that case ensured that “procedural due process is satisfied”). Claims for just compensation of $10,000 or less may be brought in either the Court of Claims or the federal district courts, under the “Little Tucker Act.” See 28 U.S.C. § 1346(a)(2).

332. See supra notes 115–139 and accompanying text (Part III.C).

333. 282 U.S. 481 (1931).

334. Id. at 486–87 (explaining that contracts were appropriated by United States Shipping Board Emergency Fleet Corp. under federal statute and executive order).


336. Id. at 492.

337. Id., quoted in Developments in the Law, supra note 125 at 878 n.338.

338. See also Reg’l Rail Reorg. Act Cases, 419 U.S. 102, 134 (1974) (stating that there would “clearly [be] grave doubts” about constitutionality of federal statute at issue in that case if statute were construed to withdraw Tucker Act remedy for governmental taking of plaintiffs’ property).
authority of federal district courts to address just-compensation claims against the federal government. The Court has held that the district courts cannot exercise federal-question jurisdiction over just-compensation claims that can be asserted in the United States Court of Federal Claims under the Tucker Act. The Court has also indicated, however, that a district court can exercise federal-question jurisdiction over a just-compensation claim if Congress has withdrawn the Tucker Act remedy for that claim. The Court has never suggested that the federal government’s sovereign immunity bars the district courts from hearing just-compensation claims that cannot be asserted under the Tucker Act. To the contrary, the Court’s decisions suggest that the federal district courts act as a backstop for the Court of Claims by resolving just-compensation claims that do not fall within the latter’s jurisdiction under the Tucker Act.


340. See E. Enters., Inc. v. Apfel, 524 U.S. 498, 520–22 (1998) (plurality opinion) (holding that takings claim was properly addressed by federal district court because Tucker Act remedy had presumptively been withdrawn for plaintiff’s claim); Reg’l Rail Reorg. Act Cases, 419 U.S. at 124–25, 138 (holding that takings claim was ripe given uncertainty about availability of Tucker Act remedy).


342. See Kidalov & Seamon, supra note 196, at 43–44 (discussing power of federal district courts to decide whether Tucker Act remedy is available for alleged takings by federal government). For the reasons discussed in analyzing just-compensation claims against the States, an adequate remedy for just-compensation claims against the United States cannot always be furnished through an officer
State Sovereign Immunity

In short, just as the state courts enable the States to provide due process, so too do the United States Court of Federal Claims and the United States District Courts enable the federal government to provide due process. And, just as the Due Process Clause would prevent state courts from honoring a defense of sovereign immunity when doing so would deny just compensation, so would the Clause prevent the federal courts from doing so. 343

D. Congress’s Power Under Section 5 of the Fourteenth Amendment

Section 5 of the Fourteenth Amendment empowers Congress to

343. As discussed in Part III.C above, supra notes 115–39 and accompanying text, there is early Court case law suggesting that the United States would be immune from just-compensation claims in the absence of a statutory waiver such as the Tucker Act. It is possible that, contrary to the conclusion for which this Section of the Article contends, the Court could rely on that early case law to hold that the United States is immune from just-compensation claims while also concluding that the Fourteenth Amendment overrides the States’ sovereign immunity from such claims. This differing treatment of the United States and the States could be defended on at least three grounds: (1) that the Fifth Amendment, unlike the Fourteenth Amendment, was not meant to alter the sovereign immunity that existed under the original Constitution, see Seminole Tribe v. Florida, 517 U.S. 44, 65–66 (1996) (stating that Congress’s power under Section 5 of Fourteenth Amendment to abrogate state sovereign immunity rested on the rationale that “the Fourteenth Amendment, adopted well after the adoption of the Eleventh Amendment and the ratification of the Constitution, operated to alter the pre-existing balance between state and federal power achieved by Article III and the Eleventh Amendment”); (2) that the separation-of-powers doctrine puts a limit on federal courts’ power to enter money judgments against members of the federal political branches, cf. Chisholm v. Georgia, 2 U.S. (2 Dall.) 419, 477 (1793) (opinion of Jay, C.J.) (stating that federal courts’ reliance on executive branch for enforcement of its judgments “place[s] the case of a State, and the case of the United States, in very different points of view” when it came to sovereign immunity); or (3) that the Framers intended in the original Constitution to distinguish the suitability of the States from the suitability of the United States, see James E. Pfander, History and State Suability: An “Explanatory” Account of the Eleventh Amendment, 83 CORNELL L. REV. 1269, 1370–73 (1998) (discussing historical basis for differential treatment of federal and state sovereign immunity). I thank Professors Joshua Schwartz and Robert Brauneis for the insight that the United States’ immunity from just-compensation claims could, consistently with case law, be found to differ from that of the States. See Posting of Environmental Policy Project, (Dec. 17, 1999) (forwarding comments of Joshua I. Schwartz, Professor of Law, The George Washington Univ. Law School, and Robert Brauneis, Associate Professor of Law, The George Washington Univ. Law School) (copy on file with author).
“enforce” the Due Process Clause by “appropriate legislation.”\textsuperscript{344} Congress’s Section 5 power has been addressed in several recent decisions of the Court.\textsuperscript{345} This Part does not attempt a comprehensive analysis of those decisions. Instead, this Part describes two ways in which the due-process analysis proposed in this Article sheds light on Congress’s Section 5 power. First, the proposed analysis clarifies that the States’ obligation to remedy erroneous deprivations of life, liberty, or property stems from the procedural component of the Due Process Clause.\textsuperscript{346} Congress can use Section 5 to enforce those remedial obligations, but it is not necessarily “appropriate” for Congress to do so by regulating the state conduct that merely gives rise to those obligations.\textsuperscript{347} Second, the proposed analysis clarifies that federal courts cannot help the States meet their remedial obligations under the Due Process Clause.\textsuperscript{348} Accordingly, Congress’s power to enforce the States’ remedial obligations under Due Process Clause of the Fourteenth Amendment is not diminished by the availability of federal-court remedies.\textsuperscript{349}

Because the procedural component of the Due Process Clause puts remedial obligations on the States, Congress can use Section 5 to “enforce” those remedial obligations by “appropriate” legislation.\textsuperscript{350} To be “appropriate,” however, the legislation must be tailored to remedying or preventing the States’ failure to meet their remedial obligations.\textsuperscript{351} A federal law may not satisfy this “tailoring” requirement if, instead of

\begin{itemize}
\item \textsuperscript{344} U.S. CONST. amend. XIV, § 5 (“The Congress shall have power to enforce, by appropriate legislation, the provisions of this article.”).
\item \textsuperscript{346} See supra notes 178–248, 260–325 and accompanying text (discussing States’ remedial obligations under the doctrine of procedural due process).
\item \textsuperscript{347} See supra Part IV.D.
\item \textsuperscript{348} See infra notes 392–94 and accompanying text.
\item \textsuperscript{349} See infra note 393 and accompanying text.
\item \textsuperscript{350} See U.S. CONST. amend. XIV, § 5; see also City of Boerne v. Flores, 521 U.S. 507, 519 (1997) (stating that “[t]he ‘provisions of this article,’ to which § 5 refers, include the Due Process Clause of the Fourteenth Amendment,” and holding that Congress had power under Section 5 to enforce Free Exercise Clause, which had been held to be incorporated into Due Process Clause).
\item \textsuperscript{351} See Fla. Prepaid Postsecondary Educ. Expense Bd. v. Coll. Sav. Bank, 527 U.S. 627, 639 (1999) (stating that, “for Congress to invoke § 5, it must identify conduct transgressing the Fourteenth Amendment’s substantive provisions, and must tailor its legislative scheme to remedying or preventing such conduct”).
\end{itemize}
regulating state remedial procedures, the law regulates state conduct that could give rise to remedial obligations. The “appropriate” way for Congress to remedy or prevent defective state remedial procedures would ordinarily be to regulate the remedial procedures themselves.\textsuperscript{352}

This appears to be why the Court in \textit{Florida Prepaid Postsecondary Education Expense Board v. College Savings Bank}\textsuperscript{353} concluded that the federal statute challenged there exceeded Congress’s power under Section 5. \textsuperscript{354} \textit{Florida Prepaid} concerned the federal Patent and Plant Variety Protection Remedy Clarification Act (“Patent Remedy Act,” or “Act”). The Act made States civilly liable for infringing patents.\textsuperscript{355} The Court in \textit{Florida Prepaid} observed that a State’s infringement of a patent could cause a deprivation of property, if the infringement was intentional.\textsuperscript{356} Thus, the Patent Remedy Act targeted State conduct that could give rise to state remedial obligations under the doctrine of procedural due process. The Court added, however, that, “[i]n procedural due process claims, the deprivation by state action of a constitutionally protected interest…is not in itself unconstitutional; what is unconstitutional is the deprivation of such an interest without due process of law.”\textsuperscript{357} Thus, a State’s intentional infringement of patents violates procedural due process “only where the State provides no remedy, or only inadequate remedies, to injured patent owners for its infringement of their patent.”\textsuperscript{358} Nonetheless, Congress “barely considered the availability of state remedies for patent infringement” when it was drafting the Patent Remedy Act.\textsuperscript{359} As a result, the statute that Congress enacted was not limited to “cases involving arguable constitutional

\textsuperscript{352} The qualification in the text (“ordinarily”) reflects that Congress can use Section 5 to enact “reasonably prophylactic legislation” in response to “[d]ifficult and intractable” violations of the Fourteenth Amendment. \textit{Kimel v. Fla. Bd. of Regents}, 528 U.S. 62, 88, 91 (2000).

\textsuperscript{353} \textit{Id.} at 646–47.

\textsuperscript{354} \textit{Id.} at 646–47.

\textsuperscript{355} \textit{Id.} at 646–47.

\textsuperscript{356} \textit{Id.} at 646–47.

\textsuperscript{357} \textit{Id.} at 646–47.

\textsuperscript{358} \textit{Id.} at 646–47.

\textsuperscript{359} \textit{Id.} at 646–47.
violations, and hence did not fall within Section 5. In short, the Patent Remedy Act was not “appropriate” legislation for enforcing the procedural Due Process Clause because it was aimed at state conduct that could give rise to remedial obligations under the Clause—i.e., at the States’ infringement of patents—rather than at state conduct that violated those remedial obligations—i.e., at the States’ failure adequately to avoid or remedy their intentional patent infringements.

I respectfully disagree with commentators who have criticized _Florida Prepaid_ for failing to analyze the Patent Remedy Act as a law that enforced substantive due process. It is true that the plaintiff in _Florida_...
State Sovereign Immunity

*Florida Prepaid* was not complaining about the procedure by which the State infringed its patent; the plaintiff wanted a remedy for that infringement. Yet, as previously discussed, a State’s obligation to remedy an erroneous deprivation of property, including a State’s intentional infringement of a patent, stems from the procedural component of the Due Process Clause. Thus, a plaintiff’s desire for a remedy for a State’s deprivation of his or her property interest does not automatically identify the plaintiff’s claim as one grounded in substantive due process.

Likewise, a plaintiff’s claim that a deprivation of property was erroneous does not automatically make that claim a substantive due process claim. To plead a substantive due process claim, the plaintiff must allege that the deprivation was not merely erroneous, but so flawed that it was “arbitrary in the constitutional sense.” The plaintiff in *Florida Prepaid* did not allege constitutional arbitrariness. The plaintiff did assert that the State’s infringement of its patent violated the federal law against patent infringement. That assertion was probably correct; even after *Florida Prepaid*, States probably still have to obey the federal patent laws.

---

363. See supra note 362, at 1675 (asserting that “the nature of the underlying claim in *Florida Prepaid* seems to be substantive rather than procedural”).

364. See supra notes 261–97 and accompanying text.

365. County of Sacramento v. Lewis, 523 U.S. 833, 846 (1998) (“Only the most egregious official conduct can be said to be arbitrary in the constitutional sense.”) (internal quotation marks omitted); Fallon, *Confusions*, supra note 190, at 310 (“In its commonest form, substantive due process doctrine reflects the simple but far-reaching principle . . . that government cannot be arbitrary.”) (footnotes omitted).

366. The plaintiff in *Florida Prepaid* contended that the State’s infringement was “willful.” *Fla. Prepaid*, 527 U.S. at 632–33. It is doubtful that even most “willful” infringements would violate substantive due process, given the Court’s repeated insistence that “only the most egregious official conduct can be said to be arbitrary in the constitutional sense.” *Lewis*, 523 U.S. at 846 (internal quotation marks omitted).

367. *Fla. Prepaid*, 527 U.S. at 631 (“College Savings claims that, in the course of administering its tuition prepayment program, Florida Prepaid directly and indirectly infringed College Savings’ patent. College Savings brought an infringement action under 35 U.S.C. § 271(a) . . . .”).

368. The Court in *Florida Prepaid* concluded “that the Patent Remedy Act cannot be sustained under § 5 of the Fourteenth Amendment.” Id. at 647. This conclusion does not resolve whether, to the extent that the Act imposes only a substantive obligation on States not to infringe patents, it can be justified under the Patent Clause, U.S. CONST. art. I, § 8, cl. 8. In *Alden*, the Court clearly
ordinarily would violate substantive due process (as distinguished from violating the federal patent laws) only if the infringement were “arbitrary in the constitutional sense”—i.e., “egregious.” And the fact that a State’s conduct violates the federal patent laws does not make the conduct “arbitrary” for purposes of substantive due process (or otherwise unconstitutional). Congress cannot use Section 5 to alter the Court’s definition of “arbitrary” for purposes of substantive due process any more than Congress can use Section 5 to alter the meaning of the Free Exercise Clause.

The analysis is the same if one treats the plaintiff’s claim in Florida Prepaid as alleging that, by infringing the plaintiff’s patent, the State was taking private property for a public use. A State’s taking of private property for public use does not violate the Just Compensation Clause, as applicable to States under the Fourteenth Amendment. Instead, a violation of the Just Compensation Clause occurs only if the State fails to pay just compensation for the taking. Thus, the assertion of a taking

believed that States continue to have an obligation to pay the wages prescribed in the FLSA, even though they cannot be sued under that statute for retroactive monetary relief. See Alden v. Maine, 527 U.S. 706, 754–60 (1999) (discussing alternative ways of enforcing federal laws against the States and observing that Maine “has not questioned Congress’s power to prescribe substantive rules of federal law to which it must comply”). By parity of reasoning, States would also still have to obey the federal patent laws.

369. Lewis, 523 U.S. at 846.

370. See Fla. Prepaid, 527 U.S. at 643 (“[A] State’s infringement of a patent . . . does not by itself violate the Constitution.”).

371. See City of Boerne v. Flores, 521 U.S. 507, 532 (1997) (stating that “Congress does not enforce a constitutional right by changing what the right is” and holding that, as applied to States, Religious Freedom Restoration Act of 1993, 42 U.S.C. §§ 2000bb et seq. (1996), exceeded Congress’s enforcement power under Section 5, because Act “appears to attempt a substantive change in constitutional protections.”).

372. In fact, the plaintiff in Florida Prepaid did argue that the Patent Remedy Act fell within Congress’s power under Section 5 because it was “appropriate” for enforcing the States’ obligations under the Just Compensation Clause. See Fla. Prepaid, 527 U.S. at 641–42. The Court declined to consider that argument, finding no evidence that, in enacting the Patent Remedy Act, Congress intended to rely on the Just Compensation Clause. See id. at 642 n.7; see also Kidlov & Seamon, supra note 196, at 73–75 (discussing significance of Court’s refusal to address the just-compensation defense of the statute).

373. See, e.g., First English Evangelical Lutheran Church v. County of Los Angeles, 482 U.S. 304, 315 (1987) (stating that Just Compensation Clause “does not prohibit the taking of private property”).

374. See, e.g., Suitum v. Tahoe Reg’l Planning Agency, 520 U.S. 725, 734 (1997) (stating that “only takings without ‘just compensation’ infringe” the Just Compensation Clause). A person whose patent had been taken for public use by the State could sue the State in state court alleging a claim in inverse condemnation under First English. See supra notes 19–52 and accompanying text (discussing First English). The claim would not be barred by the federal statute that gives federal
State Sovereign Immunity

does not establish a violation of a State’s substantive obligation under the Fourteenth Amendment to pay just compensation. Of course, a State violates the procedural component of the Due Process Clause if it fails to provide adequate procedures for paying just compensation.375 But violations of neither the substantive obligation to pay just compensation nor the procedural obligation to have a scheme for doing so are targeted by a federal statute that creates civil liability for all state infringements of patents, even those for which the State provides just compensation. Such a statute would not target state violations of the substantive obligation to pay just compensation, because state takings of private property for public use do not violate that component unless they go uncompensated.376 And the statute would not target state violations of

375. See, e.g., City of Monterey v. Del Monte Dunes at Monterey, Ltd., 526 U.S. 687, 717 (1999) (stating that government violates the Constitution “either by denying just compensation in fact or by refusing to provide procedures through which compensation may be sought”). In Part IV, I argue that a State can establish adequate procedures for payment of just compensation without waiving immunity in its own courts, but, if (and only if) it fails to do so, its use of sovereign immunity to prevent the recovery of just compensation would violate the procedural component of the Due Process Clause of the Fourteenth Amendment. In my view, this argument accords with the Court’s precedent, both old and new. I therefore respectfully disagree with Professor Vázquez’s view that the Court’s decision in Coll. Sav. Bank v. Fla. Prepaid Postsecondary Educ. Expense Bd., 527 U.S. 666 (1999), can be read to mean “that a state’s insistence on sovereign immunity is never a violation of the Fourteenth Amendment.” Professor Vázquez stated,

[The Court in College Savings Bank, 527 U.S. at 675] indicated that the statute involved in that case was a prophylactic measure because it was a prohibition of States’ sovereign-immunity claims, which are not in themselves a violation of the Fourteenth Amendment. This suggests that a state’s insistence on sovereign immunity is never a violation of the Fourteenth Amendment . . . .

Carlos Manuel Vázquez, Eleventh Amendment Schizophrenia, 75 NOTRE DAME L. REV. 859, 898 (2000) [hereinafter Vázquez, Schizophrenia] (internal quotation marks and footnote omitted; emphasis added). College Savings indicates only that a State’s reliance on sovereign immunity does not necessarily violate the Fourteenth Amendment, a principle that is fully consistent with the analysis in this Article. See Coll. Sav. Bank, 527 U.S. at 675 (remarking that State’s assertion of sovereign-immunity defenses to patent-infringement claims “are not in themselves violations of the Fourteenth Amendment”).

376. Professors Heald and Wells assert that most patent infringements by the States are caused by officials who lack the power to exercise eminent domain. See Heald & Wells, supra note 19, at 870–71, 888. They appear to believe that the infringement of a patent by a state official who lacks eminent-domain power would violate the Fourteenth Amendment if the infringement constituted a taking of property for which just compensation was due. Id. at 888 (contrasting such an infringement
the procedural component of the Due Process Clause, because it would operate without regard to whether the state had an adequate procedure for paying just compensation for its takings.\footnote{Because Florida Prepaid addressed Congress’s power to enforce the States’ procedural due process obligations, I respectfully disagree with commentary suggesting that Florida Prepaid conflicts with Home Telephone, which was a substantive due process case. See David L. Shapiro, The 1999 Trilogy: What is Good Federalism?, 31 RUTGERS L.J. 753, 757 (2000). In Home Telephone, a phone company claimed that telephone rates set by a city violated the Fourteenth Amendment because they were so low as to be confiscatory. Home Tel. & Tel. Co. v. City of Los Angeles, 227 U.S. 278, 281 (1913). This claim asserted a violation of substantive due process, not procedural due process. See, e.g., Stephen A. Siegel, Understanding the Lochner Era: Lessons from the Controversy Over Railroad and Utility Rate Regulation, 70 VA. L. REV. 187, 255 (1984) (identifying substantive due process as doctrine underlying Court’s decisions reviewing reasonableness of utility rates). The Court concluded in Home Telephone that the company had adequately alleged a substantive due process violation, even if the challenged rates violated state law and could have been set aside on that basis. Home Tel., 227 U.S. at 282–95. This conclusion reflects that a violation of substantive due process can occur before state remedies for the violation have been exhausted. That principle does not conflict with Florida Prepaid. Florida Prepaid reflects that, unlike a State’s violation of substantive due process, a State’s violation of procedural due process sometimes does not occur until after adequate state remedies have been exhausted. See Fla. Prepaid, 527 U.S. at 642–43 (relying on cases involving “procedural due process claims”); see also Zinermon v. Burch, 494 U.S. 113, 125–26 (1990) (explaining that violation of substantive due process “is complete when the wrongful action is taken,” whereas violation of procedural due process “is not complete when the deprivation occurs; it is not complete unless and until the State fails to provide due process.”). If a plaintiff can plead a violation of substantive due process, the plaintiff can seek a}

\textit{See id.} If I understand their argument correctly, I respectfully disagree. It is true that a state official’s infringement of a patent could immediately give rise to a cause of action in inverse condemnation, if the infringement constituted a taking of private property for public use. See id. at 888. A cause of action in inverse condemnation is not the same, however, as a cause of action for a violation of the Just Compensation Clause. See Kidalov & Seamon, supra note 196, at 51–52. A cause of action in inverse condemnation arises as soon as a taking occurs, whereas a cause of action for a violation of the Just Compensation Clause does not arise until the State fails to pay just compensation for the taking. See, e.g., Williamson County Reg’l Planning Comm’n v. Hamilton Bank, 473 U.S. 172, 195 (1985) (recognizing that “a property owner has not suffered a violation of the Just Compensation Clause until the owner has unsuccessfully attempted to obtain just compensation through the procedures provided by the State for obtaining such compensation”). Thus, a state official’s infringement of a patent does not violate the Constitution just because it causes a taking of private property for public use. This conclusion is not altered by the official’s lack of eminent-domain power. As long as the official’s act of infringement fell within the general scope of his or her authority—i.e., as long as it was not ultra vires—the official’s conduct could cause a taking for which just compensation was due. See Del-Rio Drilling Programs, Inc. v. United States, 146 F.3d 1358, 1362–63 (Fed. Cir. 1998). If the official’s infringement was ultra vires, it could not constitute a taking for which just compensation is due. See id. If the infringement was “arbitrary in the constitutional sense,” Lewis, 523 U.S. at 846 (internal quotes omitted), it could violate substantive due process, and be actionable on that basis. As discussed in the text, however, the official’s conduct would not be constitutionally arbitrary merely because it violated federal patent laws.
This explanation of *Florida Prepaid* as a decision about Congress’s power to enforce the procedural component of the Due Process Clause should show that the decision does not necessarily “undo” the Court’s decision in *Alden v. Maine*.378 In *Alden*, state employees sued the State of Maine in a state court in Maine for backpay under the federal Fair Labor Standards Act (FLSA), a statute enacted under Congress’s Article I, Commerce Clause power.379 The Court in *Alden* held that the suit was barred by the State’s immunity in its own courts and that Congress could not use its Article I powers to abrogate that immunity.380 Professor Vázquez has argued that these holdings may be insignificant if a state employee’s right to the wages prescribed by the FLSA constitutes a “property” interest protected by the doctrine of procedural due process.381 The premise of this argument is that the States’ use of sovereign immunity to defeat a state-court action for backpay under the FLSA would violate the Due Process Clause.382

It is not necessarily true, however, that a State would violate the Due Process Clause by using sovereign immunity to bar a state-court action against it for backpay under the FLSA. To begin with, a State’s violation of the FLSA is no more tantamount to a violation of substantive due process than is a State’s violation of the federal patent laws. Neither sort of violation is necessarily “arbitrary” as the Court has defined that term

---

378. *Florida Prepaid*, 527 U.S. 706 (1999), discussed *supra* notes 140–77 and accompanying text; see Vázquez, *Alden Trilogy*, *supra* note 18, at 1928 (asserting that *Florida Prepaid* “may contain the seed of *Alden’s substantial undoing*”).


380. *Alden*, 527 U.S. at 754.

381. Vázquez, *Alden Trilogy*, *supra* note 18, at 1974–80 (arguing that comparison of *Alden* and *Florida Prepaid* leads to conclusion “either that, under the holding of *Florida Prepaid*, the *Alden* plaintiffs would have been entitled under the Due Process Clause to the relief that the Court in *Alden* denied them, or that the Court now rejects its ‘new property’ jurisprudence”).

382. Vázquez, *Alden Trilogy*, *supra* note 18, at 1974 (asserting that, if the right to wages under FLSA constituted property, plaintiffs in *Alden* “would have been entitled under the Due Process Clause to the relief the Court in *Alden* denied them”). The plaintiffs in *Alden* sought back pay for approximately a four-year period. See *Joint Appendix* at 50–51, *Alden* (No. 98-436) (unpublished decision of state trial court in *Alden* case, stating that plaintiffs “are seeking damages for the period of December 21, 1989 to February 6, 1994”).
for purposes of substantive due process.\footnote{383} It is true that, if the right of a state employee to the wages prescribed by the FLSA is, like a patent, a property interest, a State’s intentional violation of the FLSA would constitute a deprivation of property.\footnote{384} The State would therefore have a procedural due process obligation to establish a system for ensuring that its payment of wages complied with the FLSA. That system would have to ensure accurate payment of wages and a way to remedy erroneous payments.

The question remains what remedy the State would have to provide when its system detected a post-payment error. That question is one of procedural due process, for reasons already discussed. Procedural due process might generally require retroactive compensation from the State for FLSA violations; the Court’s tax cases arguably support that requirement.\footnote{385} If the Due Process Clause did require retroactive compensation from the State for FLSA violations, then \textit{Alden} would indeed lose much significance.

It is far from clear, however, that the Due Process Clause would require retroactive compensation from the State for violations of the FLSA.\footnote{386} One could plausibly argue that prospective correction of the

\footnote{383. See \textit{supra} text accompanying notes 365–71, \textit{supra} (explaining why State’s violation of patent laws does not violate substantive due process).}

\footnote{384. An employee’s right to the wages prescribed by the FLSA appears to be “property” as the Court has usually defined the term for purposes of the Due Process Clause. \textit{See, e.g., Bd. of Regents v. Roth}, 408 U.S. 564, 577 (1972) (stating that property interest exists for due process purposes if “existing rules or understandings that stem from” sources of law independent of the Constitution support “legitimate claims of entitlement” to a monetary benefit). In a companion case to \textit{Florida Prepaid}, however, the Court appeared to adopt a narrower definition of “property” in holding that the Lanham Act, 15 U.S.C. § 1125(a), did not give the plaintiff a “property” interest in being free from a state entity’s making a false statement about a product that the State marketed in competition with the plaintiff’s product. \textit{See Coll. Sav. Bank v. Fla. Prepaid Postsecondary Educ. Expense Bd.}, 527 U.S. 666, 673–75 (1999). To reconcile \textit{College Savings} with precedent, the result in that case is best understood to reflect that, when a State makes false claims about its own product, any resulting competitive injury is too indirect to cause the owner of a competing product to suffer a “deprivation” within the meaning of the Due Process Clause. \textit{Id.} at 675. \textit{Cf. Martinez v. California}, 444 U.S. 277, 285 (1980) (State’s decision to parole person who murdered plaintiffs’ decedent did not “deprive” decedent of life because murder was “too remote a consequence”). Admittedly, this was not the reason the Court gave for its decision in \textit{College Savings}.}

\footnote{385. See \textit{supra} notes 217–26 and 300–14 (discussing Court’s tax cases); \textit{see also} Brief for Petitioners at 29–31, \textit{Alden} (No. 98-436) (arguing that Court’s tax cases supported conclusion that plaintiffs were entitled to retroactive relief for State’s FLSA violations as a matter of due process).}

\footnote{386. Even if the Due Process Clause does require retroactive compensation from the State for violations of the FLSA, the Clause does not necessarily require the State to make that remedy available in its courts. Instead, the State may be able to satisfy the Due Process Clause by creating an adequate nonjudicial scheme for awarding that compensation.}
error that led to the violations, such as through injunctive relief, would satisfy the Due Process Clause. Alternatively, one could argue that due process requires retroactive compensation only for egregious violations and, even then, only compensation payable by the officials responsible for the violations. These arguments would emphasize the difference between a State’s erroneous collection of a tax—which deprives people of money (or other property) that they already have received—and a State’s failure to pay the prescribed wage—which deprives people of money that they have not yet received. In addition to this difference between the individual interest at stake in the two settings, there may be other differences that justify interpreting the Due Process Clause to mandate different remedies for erroneous taxation and erroneous payment of wages. For example, perhaps States have greater incentives to avoid erroneous wage payments than they do to avoid erroneous tax collections. The point is that, although the procedural Due Process Clause may require States to remedy its violations of the FLSA, the Clause does not necessarily require that remedy to take the form of a state-court award of retroactive monetary relief payable out of the state treasury, which is the only type of remedy that implicates the sovereign immunity recognized in Alden. The broader point is that,

---

387. But cf. Jackson, The Supreme Court, supra note 18, at 113 (arguing that Court’s precedent barring federal courts from awarding retroactive monetary relief against States for federal statutory violations might be justified if state courts were available to grant that relief).


389. This is, of course, speculative. People, and especially businesses, can vote with their feet by leaving or avoiding States with defective tax systems, an ability that may create a disincentive to erroneous taxation. Similarly, States might be deterred from violating the FLSA because they compete for employees with both private employers and other public sector employers. Cf. Fallon & Meltzer, supra note 18, at 1828 (arguing that threat of refunds “are surely useful, and indeed may be necessary, to keep state officials within the bounds of law”).

390. Professor Vázquez believes that the Court may have decided Alden without considering the implications of its decision in that case for its decision the same day in Florida Prepaid, and that the “most important[]” cause of this oversight was that “the litigants in Alden failed to invoke the due process principle.” Vázquez, Alden Trilogy, supra note 18, at 1929–30. In fairness to the plaintiffs in
although due process can override the States’ immunity in their own courts, it will not always do so; the occurrence of an override depends on, among other things, whether such an override is necessary to achieve the accuracy demanded by the doctrine of procedural due process.391

Finally, the due-process analysis proposed in this Article clarifies that Congress’s use of Section 5 to remedy or prevent violations of the Fourteenth Amendment does not depend on the availability of federal-court remedies for those violations.392 As discussed above, the federal courts cannot help a State meet its remedial obligations under the procedural component of the Due Process Clause.393 The presence of federal-court remedies is therefore irrelevant to Congress’s power to ensure that the States themselves meet their remedial obligations, just as the absence of federal-court remedies is irrelevant to whether the Due Process Clause requires the States themselves to provide those remedies. Furthermore, when a State violates, not its remedial obligations, but its substantive obligations under the Due Process Clause—for example, by violating substantive due process restrictions on arbitrary state conduct

391. See supra note 268 and accompanying text (stating that procedural due process serves systemic goal of enforcing rule of law and goal of protecting individual’s interest in life, liberty, and property).

392. But cf. Vázquez, Schizophrenia, supra note 375, at 893. Professor Vázquez believes that the Court’s recent decisions on Section 5 “appear to hold that an abrogation is valid under Section 5 . . . only if a withdrawal of Eleventh Amendment immunity is ‘genuinely necessary’ to assure compliance by the states with their federal obligations.” Id.; see also id. at 895 (stating that Court’s recent Section 5 decisions “suggest that, at a minimum, Congress must consider the adequacy of . . . alternative means before it resorts to means barred by the Eleventh Amendment”). Professor Vázquez bases this reading primarily on the Court’s statement in College Savings Bank that Congress may use Section 5 to abrogate state sovereign immunity where “genuinely necessary” to prevent States from violating the Fourteenth Amendment. See id. at 861 & n.11 (quoting Coll. Sav. Bank v. Fla. Prepaid Postsecondary Educ. Expense Bd., 527 U.S. 666, 675 (1999)). Professor Vázquez then juxtaposes this statement with Justice Scalia’s statement in another case that federal-court remedies already supply “the necessary judicial means to assure [state] compliance” with the Constitution. See id. at 859–60 & n.2 (quoting Pennsylvania v. Union Gas Co., 491 U.S. 1, 33 (1989) (Scalia, J., concurring in part and dissenting in part)). Based on these statements, he suggests that the Court could conclude that Congress never has the power under Section 5 to abrogate state sovereign immunity, because that step is not genuinely necessary in light of existing federal-court remedies. See id. at 893 (“If the Court adhered to the supremacy strain [evinced in recent Section 5 decisions], abrogation would never be appropriate . . . .”). Contrary to that conclusion, the existence of federal-court remedies does not diminish Congress’s Section 5 power, for reasons discussed in the text.

393. See supra notes 241–44 and accompanying text.
State Sovereign Immunity

and on state conduct that unduly burdens fundamental liberties—that substantive violation occurs whether or not there is a remedy in federal court (or, for that matter, whether or not there is a remedy in state court) for the violation. The existence of judicial remedies for a State’s violations of its substantive obligations under the Fourteenth Amendment does not excuse those violations, and it therefore cannot dilute Congress’s power to remedy or prevent them under Section 5.

VI. CONCLUSION

The state courts and the federal courts are not fungible. Each can enable its respective sovereign to meet the remedial obligations of the Due Process Clause. By the same token, neither the state courts nor the federal courts can discharge the remedial obligations of the other sovereign. Thus, when a State fails to meet its remedial obligations, it can subject itself, in its own courts and without its consent, to suits that would be barred from federal court by sovereign immunity. One such situation arises when the State violates its obligation under the Due Process Clause to create a “reasonable, certain, and adequate” procedure to pay just compensation for private property that the State has taken for public use.

394. See Home Tel. & Tel. Co. v. City of Los Angeles, 227 U.S. 278, 282–95 (1913) (holding that existence of state-court remedies for alleged violation of Fourteenth Amendment did not defeat claim); see also Pac. Tel. & Tel. Co. v. Kuykendall, 265 U.S. 196, 204–05 (1924) (holding that plaintiff asserted present injury by claiming that state agency’s rate order was confiscatory and therefore violated Due Process Clause of Fourteenth Amendment).
Abstract: In Festo Corp. v. Shoketsu Kinzoku Kogyo Kabushiki Co., the Federal Circuit adopted a strict approach to prosecution history estoppel and thereby limited the availability of the doctrine of equivalents to patentees suing for infringement by after-arising technology. The court held that when a narrowing claim amendment creates prosecution history estoppel, the amended claim element maintains no range of equivalents and therefore the patentee is completely barred from applying the doctrine of equivalents. The court rejected the flexible approach, which allows a scope of equivalents even after a narrowing claim amendment. This Note argues that the Supreme Court should overrule the Festo decision and adopt the flexible approach to prosecution history estoppel in cases of infringement by after-arising technology. The strict approach adopted by Festo hinders the doctrine of equivalents in protecting patentees from infringement by after-arising technology. The Festo majority misinterpreted the Supreme Court’s opinion in Warner-Jenkinson Co. v. Hilton Davis Chemical Co., which implies that a scope of equivalents is available after a narrowing claim amendment has been made for patentability reasons. In addition, the Festo decision leads to incongruous results with respect to after-arising technology, as it protects unamended claims but not amended claims, and predictable arts but not unpredictable arts. Furthermore, Festo conflicts with patent policy, which rewards pioneer inventions with a broad range of equivalents. Finally, this flexible approach would give patentees fair protection from infringement by after-arising technology under the doctrine of equivalents.

The United States patent system is designed to promote the public disclosure of innovation. In exchange for publicly disclosing a new invention, a patentee gains the right to exclude others from making, using, or selling the invention.1 The U.S. Supreme Court and the Federal Circuit have recognized that copyists should not be able to avoid infringement by using after-arising technology to make insubstantial substitutions in a patented invention.2 The courts have defined after-arising technology as technological developments known after issuance of a patent.3 In order to protect patentees from such insubstantial substitutions, the courts developed the doctrine of equivalents. Under the doctrine of equivalents, a patentee can claim infringement when an accused device is not an exact copy of the patented invention, but an

---

3. Warner-Jenkinson, 520 U.S. at 37; Al-Site Corp. v. VSI Int’l, Inc., 174 F.3d 1308, 1320 (Fed. Cir. 1999); Litton, 140 F.3d at 1457; see also Martin J. Adelman, Patent Law Perspectives § 3.4[1], at 3-40.17 to 3-40.18 (2d ed. 2001).
element of the accused device performs substantially the same function in substantially the same way to obtain the same result.  

In *Festo Corp. v. Shoketsu Kinzoku Kogyo Kabushiki Co.*, the Federal Circuit, sitting en banc, resolved several issues involving the doctrine of equivalents. In the most divisive issue, the court held that a narrowing claim amendment creates a “complete bar” to the application of the doctrine of equivalents to the amended claim element. The majority reasoned that this would promote the public notice function of patents and create more certainty in the patent system. Several dissenting judges sharply disagreed, arguing that the decision contradicted Supreme Court and Federal Circuit precedent, would promote copying of patented inventions, and failed to address its effect on the role played by the doctrine of equivalents in preventing infringement by after-arising technology.

This Note argues that the Supreme Court should overrule *Festo’s* “complete bar” rule and adopt a flexible approach to prosecution history estoppel in cases of infringement by after-arising technology. Parts I and II give an overview of patent law and the doctrine of equivalents. Part III examines the relationship between the doctrine of equivalents and after-arising technology. Part IV summarizes the facts and opinions of *Festo*. Part V argues that *Festo’s* strict approach prevents the doctrine of equivalents from accommodating after-arising technology, that the *Festo* majority opinion leads to incompatible results when applied in cases involving after-arising technology, and that the decision conflicts with the patent policy of rewarding pioneer inventions with broad claim scope. This Note concludes that the Supreme Court should permit the flexible approach in all cases of infringement by after-arising technology.

5. 234 F.3d 558 (Fed. Cir. 2000) (en banc), *cert. granted*, 69 U.S.L.W. 3779 (June 18, 2001) (No. 00-1543).
6. *Id.* at 563–64.
7. *Id.* at 569.
8. *Id.* at 575.
9. *Id.* at 598, 612 (Michel, J., dissenting), 620 (Linn, J., dissenting), 630 (Newman, J., dissenting).
10. *Id.* at 616 (Michel, J., dissenting), 627 (Linn, J., dissenting), 635–36 (Newman, J., dissenting).
11. *Id.* at 619 (Rader, J., dissenting).
I. PATENT LAW AND THE DOCTRINE OF EQUIVALENTS

An inventor begins the process of obtaining a patent by sending the Patent and Trademark Office (PTO) an application explaining an invention in detail. A patent issues to the inventor once the PTO determines that the application meets certain statutory requirements. After issuance of a patent, the inventor can sue for patent infringement those who make, use, or sell the invention without authority. During litigation, the court will compare the patent to the accused device. The court can find infringement if the accused device is an exact copy of the patented invention (literal infringement) or if the accused device has elements similar enough to the claimed invention to be equivalent (doctrine of equivalents). In order to prevent a patentee from abusing the doctrine, the doctrine of equivalents is subject to various limitations.

A. The Patent Application Process

The Federal government has exclusive jurisdiction over patents as enumerated by the U.S. Constitution and codified by the 1952 Patent Act. The process of issuing patents is governed by the Patent and Trademark Office (PTO). A patent gives its holder, the patentee, the right to exclude others from making, using, or selling the patented invention for twenty years after the patent application is filed. The PTO grants patents to persons who invent something that is new, useful, and nonobvious.

Under the Patent Act, a patent application must meet three Section 112 requirements of patentability: written description, enablement, and best mode. The specification must contain a written description of the invention in sufficient detail so that one skilled in the art could reasonably conclude that the inventor had possession of the

13. U.S. CONST. art. I, § 8 ("The Congress shall have Power . . . [t]o promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries . . . ").
15. Id. § 153.
16. Id. § 154.
17. Id. §§ 101–03.
18. Id. § 112.
invention as of the filing date. 19 Under the enablement requirement, the specification must describe the invention in enough detail to enable a person skilled in the art to make and use the invention without “undue experimentation.” 20 Finally, the specification must also include the “best mode” that the inventor knows to carry out the invention. 21

In addition, the applicant’s invention must meet the statutory requirements of novelty and nonobviousness. 22 Novelty and nonobviousness are measured against “prior art,” which includes issued patents, patent applications pending at the PTO, and publications disclosing technological discoveries. 23 Novelty requires the invention to be something new. 24 Nonobviousness requires the invention be more than an obvious variation of the prior art as measured by a person with ordinary skill in the art. 25

The process of applying for a patent involves a series of negotiations between an applicant and a PTO Examiner. Applications that do not meet statutory requirements are rejected by the Examiner. 26 If the Examiner rejects an application, he or she prepares a document informing the applicant that the application has been rejected and stating the reasons for rejection. 27 The applicant may then respond to the Examiner’s rejection, either by arguing that the application complies with statutory requirements, or by amending the rejected claims, or both. 28 Nearly all applications are amended during the application process. 29 The Examiner will consider the applicant’s arguments and amendments and either issue a patent or reject the claims again. 30 This process continues until the

---

22. Id. § 102–03.
23. Id. § 102; ALAN L. DURHAM, PATENT LAW ESSENTIALS: A CONCISE GUIDE 80 (1999).
24. Id. § 102.
26. Id. § 103.
28. Id. § 1.104(a)(2).
29. Id. § 1.111.
Examiner grants an allowance or the Examiner issues a Final Rejection.\(^{31}\) A Final Rejection may first be appealed to the PTO Board of Patent Appeals, and then to the Court of Appeals for the Federal Circuit.\(^{32}\) Once a patent issues, the record of the proceedings before the PTO, also called the “prosecution history,” is made available to the public.\(^{33}\) The prosecution history is reviewed during patent litigation in determining whether a patent has been infringed.\(^{34}\)

**B. Patent Infringement**

A patent gives its holder the right to sue those who infringe the patented invention. In order to determine whether a patent has been infringed, courts must employ a two-step analysis.\(^{36}\) First, a court must interpret the meaning of the claims by examining the claim language, the patent specification, and the prosecution history.\(^{37}\) Once a court has determined the meaning of the claims, it must determine whether the patent has been infringed by comparing each element of the patent’s claims to the accused device.\(^{38}\) At this second step, called infringement analysis, a court may find infringement under one of two theories: literal infringement or the doctrine of equivalents.\(^{39}\) The accused device must contain each and every element of the claimed invention, either literally or equivalently, to infringe the patent.\(^{40}\)

To prove literal infringement, a patentee must show that an accused device contains every element of a claimed invention.\(^{41}\) Literal infringement has the benefit of providing the public clear notice as to

---

\(^{31}\) *Id.* §§ 1.113 (final rejection), 1.311 (notice of allowance).

\(^{32}\) *Id.* §§ 1.191 (appeal to Board of Patent Appeals), 1.301 (appeal to U.S. Court of Appeals for the Federal Circuit).

\(^{33}\) 5A DONALD S. CHISUM, CHISUM ON PATENTS: A TREATISE ON THE LAW OF PATENTABILITY, VALIDITY, AND INFRINGEMENT § 18.05, at 18-413 (2000).

\(^{34}\) 37 C.F.R. § 1.11.

\(^{35}\) Markman v. Westview Instruments, Inc., 52 F.3d 967, 980 (Fed. Cir. 1995) (claim interpretation); CHISUM, supra note 33, § 18.05, at 18-413 to 18-415.

\(^{36}\) Markman, 52 F.3d at 976.

\(^{37}\) *Id.* (referring to this process as claim interpretation).


\(^{39}\) *Id.*

\(^{40}\) *Id.*

\(^{41}\) Litton Sys., Inc. v. Honeywell, Inc., 140 F.3d 1449, 1454 (Fed. Cir. 1998) (“[A]ny deviation from the claim precludes a finding of literal infringement.”).
what subject matter a patent covers. However, literal infringement fails to
prevent copyists from pirating claimed inventions by making insubstantial substitutions to the invention’s claim elements. Therefore, the courts developed the doctrine of equivalents to prevent such copying.

The doctrine of equivalents provides patentees fair protection against infringement, freeing claim language from its literal meaning. Under the doctrine of equivalents, a patentee can claim infringement by showing that elements of the accused device are equivalent to a claim limitation. In *Graver Tank & Manufacturing Co. v. Linde Air Products Co.*, the Supreme Court recognized the “triple identity” test for equivalency. Under the test, elements of an accused device are equivalent to a claim limitation if they perform “substantially the same function in substantially the same way to obtain the same result.”

The doctrine of equivalents strikes a balance between two competing policies of the patent system: providing public notice of patented inventions and providing patentees protection for their inventions. On the one hand, the doctrine provides patentees greater protection than that afforded under literal infringement by preventing “unscrupulous copyists” from avoiding infringement by making insubstantial changes to a patented invention. This greater protection is necessary to encourage public disclosure of inventions. If inventors are not confident that patent law will protect their inventions, they will be more likely to hide.

43. *Id.*
44. *Id.*
45. *Id.*
46. *Id.* at 610.
their discoveries, thereby inhibiting other researchers from making further innovations.\footnote{See id.}

On the other hand, the doctrine of equivalents can blur the boundaries defining what a patent protects. When applied broadly, to cover a large scope of equivalents, the doctrine “conflicts with the definitional and public-notice functions of the statutory claiming requirement.”\footnote{Warner-Jenkinson, 520 U.S. at 29.} Adequate notice is necessary to permit sound predictions about the outcome of infringement litigation,\footnote{Martin J. Adelman & Gary L. Francione, The Doctrine of Equivalents in Patent Law: Questions That Pennwalt Did Not Answer, 137 U. Pa. L. Rev. 673, 682 (1989).} thereby allowing other researchers to make developments in the same area without fear of litigation.\footnote{Id. at 682–83 (this situation is referred to as “design around”).} Therefore, while the doctrine of equivalents provides patentees valuable infringement protection, it should not provide such protection by completely sacrificing the notice function of claims.\footnote{Chisum, supra note 50, at 7.}

### II. LIMITATIONS ON THE DOCTRINE OF EQUIVALENTS

The courts have developed certain limitations on the doctrine of equivalents to prevent the doctrine from eliminating the notice function of claims. Among these limitations is prosecution history estoppel, which limits the availability of the doctrine of equivalents to a patentee in infringement litigation. In addition, courts have crafted other limitations, which include the “all elements” rule and the “prior art” limitation.

#### A. Prosecution History Estoppel and the Warner-Jenkinson Presumption

Prosecution history estoppel prevents a patent holder from using the doctrine of equivalents in litigation to reach subject matter surrendered during the patent application process. The Supreme Court created the \textit{Warner-Jenkinson} presumption to handle situations in which the reason for a claim amendment is unclear from the prosecution history. In such a case, the Court allows no scope of equivalents for the amended claim element.
Prosecution history estoppel prevents a patentee from claiming subject matter that the patentee surrendered during prosecution of the patent. During prosecution, a patentee may have to make narrowing claim amendments to overcome patentability rejections by the PTO, and thereby surrender subject matter that the claims initially covered. In such a case, the patentee is estopped from reaching this subject matter during litigation if the record shows that it was relinquished during prosecution.

Prior to Festo, Federal Circuit precedent supported both a “flexible” and a “strict” approach to the scope of equivalents available after prosecution history estoppel. The vast majority of cases took the flexible approach. Under the flexible approach, a narrowing claim amendment does not automatically bar application of the doctrine of equivalents to a claim element. Instead, a court examines the prosecution history to determine what scope of equivalents remains. After examining the record, a court may still find that the doctrine of equivalents is precluded if the potential infringer clearly failed to reach the amended element’s remaining subject matter. Two Federal Circuit cases employed a strict approach. Under the strict approach, a court refuses to examine the reasons for the claim amendment in the

58. Cybor Corp. v. FAS Techs., Inc., 138 F.3d 1448, 1460 (Fed. Cir. 1998).
59. Id.
62. CHISUM, supra note 33, § 18.05[3][b][i], at 18-497, 18-503.
63. Hughes, 140 F.3d at 1476–77.
64. Id. at 1476.
65. Id. at 1477.
66. Kinzenbaw, 741 F.2d 383, 389; Prodyne, 743 F.2d 1581, 1583; see also Festo Corp. v. Shoketsu Kinzoku Kogyo Kabushiki Co., 234 F.3d 558, 610 (Fed. Cir. 2000) (en banc) (Michel, J., dissenting) (stating that the two cases following the strict approach can be reconciled with the flexible approach cases), cert. granted, 69 U.S.L.W. 3779 (June 18, 2001) (No. 00-1543); CHISUM, supra note 33, § 18.05[3][b][i], at 18-496.
prosecution history and simply precludes applying the doctrine of equivalents to the amended claim element. 67

2. The Warner-Jenkinson Presumption

The Supreme Court crafted an important corollary to prosecution history estoppel in Warner-Jenkinson Co. v. Hilton Davis Chemical Co. 68 Warner-Jenkinson involved a patent owned by Hilton Davis that covered a process for purifying food and drug dyes. 69 The patent claimed that the process operated “at a pH from approximately 6.0 to 9.0.” 70 Subsequent to the issuance of the Hilton Davis patent, Warner-Jenkinson developed a similar filtration process that operated at a pH of 5.0. 71 As a result, Hilton Davis sued Warner-Jenkinson for patent infringement under the doctrine of equivalents. 72

At issue in the case was the reason Hilton Davis had limited its patent to a pH level of 6.0. When Hilton Davis initially filed its patent application with the PTO, it did not contain a specific pH level limitation. 73 The pH limit of 6.0 was later added to the claims. 74 Under prosecution history estoppel, amendments made for patentability reasons bar the patentee from claiming subject matter given up by the amendment. Yet, because the record did not explain why Hilton Davis had added the lower pH limit of 6.0, it was unclear whether Hilton Davis should be allowed to reach a pH level of 5.0 through the doctrine of equivalents. 75

In ruling on the case, the Supreme Court crafted what has become known as the “Warner-Jenkinson presumption.” 76 When no explanation for a claim amendment is established, a court should presume that the applicant amended the claim due to a “substantial reason related to

67. CHISUM, supra note 33, § 18.05[3][b], at 18-492, 18-496.
68. 520 U.S. 17 (1997).
69. Id. at 21.
70. Id. at 22.
71. Id. at 23.
72. Id.
73. Id. at 22.
74. Id.
75. Id. at 41.
patentability.\textsuperscript{77} In such cases, a court should completely bar the patentee from claiming any equivalents for that particular claim element unless the patentee can show an “appropriate reason for a required amendment.”\textsuperscript{78} Prosecution history estoppel precludes the application of the doctrine of equivalents to that amended claim element.\textsuperscript{79} \textit{Warner-Jenkinson} did not specifically rule on whether a claim element maintains a scope of equivalents after prosecution history estoppel.\textsuperscript{80} However, Federal Circuit cases,\textsuperscript{81} as well as commentators,\textsuperscript{82} have interpreted \textit{Warner-Jenkinson} as promoting a scope of equivalents after a narrowing claim amendment made for reasons related to patentability. In \textit{Sextant Avionique, S.A. v. Analog Devices, Inc.},\textsuperscript{83} the Federal Circuit closely examined \textit{Warner-Jenkinson}’s language and rejected the notion that the Supreme Court meant that all claim amendments for reasons related to patentability created a “complete bar” to the doctrine of equivalents.\textsuperscript{84} Instead, the “complete bar” rule of the presumption applies only when the patentee fails to establish a reason for the claim amendment.\textsuperscript{85}

\textbf{B. Other Limitations on the Doctrine of Equivalents}

The courts have developed other limitations that prevent a patentee from abusing the doctrine of equivalents by stretching claim language beyond what the inventor has created. Under the “all elements” rule, each and every element of a claimed invention must be found in an

\begin{itemize}
\item \textsuperscript{77} \textit{Warner-Jenkinson}, 520 U.S. at 33.
\item Id.
\item Id.
\item CHISUM, supra note 33, § 18.05[3][c], at 18-508 to 18-509.
\item Sextant Avionique, 172 F.3d at 831; Litton Sys., Inc. v. Honeywell, Inc., 140 F.3d 1449, 1455 (Fed. Cir. 1998) (“[T]he entire context of the Warner-Jenkinson opinion shows that the Supreme Court approved the PTO’s practice of requesting amendments with the understanding that the doctrine of equivalents would still apply to the amended language.”); \textit{id.} at 1457 (“Nowhere did the Supreme Court suggest that such after-arising equivalents would be barred as a matter of law whenever a claim is amended to overcome prior art before the examiner.”); see also Hughes Aircraft Co. v. United States, 140 F.3d 1470, 1476 (Fed. Cir. 1998) (“In evaluating the reason behind an amendment [related to patentability], a court must determine what subject matter the patentee actually surrendered.”) (citing \textit{Warner-Jenkinson}, 520 U.S. at 33 n.7)).
\item Adelman, supra note 3, § 3.5, at 3-42.49 to 3-42.50; Chisum, supra note 50, at 57–58.
\item 172 F.3d 817 (Fed. Cir. 1999).
\item Id. at 831.
\item Id.
\end{itemize}
accused device. The doctrine of equivalents must be applied to individual elements of the claim and not to the invention as a whole. A “one-to-one correspondence” between elements of the claimed invention and the accused device is not necessary. An element of the patent claim can be infringed by a single component of the accused device or a series of components can be combined to make up the claim element. The prior art limitation prevents a patentee from trying to claim something through the doctrine of equivalents that the patent holder could not claim during the patent application process. The scope of equivalents of a claim element may not reach inventions already disclosed by the prior art.

III. INFRINGEMENT PROTECTION AGAINST AFTER- ARISING TECHNOLOGY

Supreme Court and Federal Circuit precedent underscore the need for the doctrine of equivalents to protect patentees from after-arising technology. The courts define after-arising technology as equivalents that were not known at the time of patent issuance. Related to after-arising technology and the doctrine of equivalents is the pioneer invention doctrine. A pioneer invention is an endeavor in a new field. Pioneer inventions are usually infringed by after-arising technology and, consequently, often rely on the doctrine of equivalents for infringement protection. Lastly, inventions in the unpredictable arts depend on the doctrine of equivalents to cover infringement by after-arising technology.

A. A Primary Purpose of the Doctrine of Equivalents Is To Provide Infringement Protection Against After-Arising Technology

The Supreme Court and the Federal Circuit have both emphasized that a primary purpose of the doctrine of equivalents is to protect patentees

86. Pennwalt Corp. v. Durand-Wayland, Inc., 833 F.2d 931, 935 (Fed. Cir. 1987) (en banc); see also Hughes, 140 F.3d at 1474.
from after-arising technology. In *Warner-Jenkinson*, the Supreme Court expressly rejected the argument that equivalents should not extend to after-arising technology, by fixing the time when equivalency should be measured. The Court ruled that equivalency should be measured at the time of infringement, not as of the patent issue date. Because knowledge of interchangeability is objective, equivalency is not limited to what was known in the art at the time the patent was issued. Therefore, the doctrine of equivalents can apply to after-arising technology.

The Federal Circuit has applied the doctrine of equivalents to after-arising technology. In doing so, it has stressed that “[p]atent policy supports application of the doctrine of equivalents to a claim element . . . in the case of ‘after-arising’ technology because a patent draftsman has no way to anticipate and account for later developed substitutes for a claim element.” According to the Federal Circuit, this is necessary to prevent infringers from pirating an invention by using new technology to make an insubstantial change in the claimed invention.

For example, in *Hughes Aircraft Co. v. United States*, the Federal Circuit found the after-arising technology presented by the accused device equivalent to the claimed invention under the doctrine of equivalents. *Hughes* involved a patent disclosing a device for controlling the orientation of a spacecraft from a ground control station. The accused device employed microprocessors, not available at the time the patent application was filed, to compute some of the

---


94. Id.

95. Id.

96. Id.

97. See, e.g., *Pall Corp.*, 66 F.3d at 1220; *Hughes*, 140 F.3d at 1477.

98. *Al-Site Corp. v. VSI Int’l*, Inc., 174 F.3d 1308, 1321 n.2 (referring to means-plus-function claims).


100. 140 F.3d 1470 (Fed. Cir. 1998).

101. *Hughes*, 140 F.3d at 1477.

102. Id. at 1475.
positioning requirements on board. The court determined that the change was the result of technology not available until after the patent issued. The only difference was that the function was being performed on board the satellite instead of on the ground. Therefore, the accused device infringed the patent under the doctrine of equivalents.

**B. “Pioneer” Inventions**

A pioneer invention is one that creates an entirely new function or makes a distinctive leap in science. Broad literal claims can be written for pioneer inventions because, by definition, pioneer inventions are not restricted by prior art. Accordingly, pioneer inventions will have no prosecution history relating to prior art rejections. However, pioneer inventions are susceptible to rejections under Section 112 because the patent drafter and the Examiner are trying to interpret a new field of technology. As a result, applications for pioneer patents are frequently amended because language in the patent application typically needs refinement as the inventor works toward patent issuance.

Infringement protection against after-arising technology is a particular concern for pioneer inventions. Pioneer inventions are often infringed by after-arising technology because they are advances in a new area of science. Infringement will occur from developments made after the invention becomes public. Therefore, pioneer inventions need the

---

103. *Id.* at 1472–73; *see also* CHISUM, supra note 33, § 18.04(3)(c), at 18-386.
104. *Id.* at 1475.
105. *Id.*
106. *Id.* at 1477.
110. Festo Corp. v. Shoketsu Kinzoku Kogyo Kabushiki Co., 234 F.3d 558, 622 (Fed. Cir. 2000) (en banc) (Linn, J., dissenting) (“For inventions in rapidly evolving fields, application filings are often made while the inventions are still in nascent stages, i.e., early in the evolutionary process, necessitating inevitable and sometimes considerable fine tuning of claim language after the initial application filings have been made.”); *id.* at 624 (Linn, J., dissenting) (“[E]liminating all flexibility in the scope of claim limitations amended for a statutory purpose reflects unjustified faith in the draftsperson to select language to perfectly describe a new and unobvious invention at an early stage of the development process.”).
doctrine of equivalents for fair protection against infringers who utilize after-arising technology to avoid literal infringement.

Pioneer inventions gain a broad range of equivalents during an infringement suit because application of the doctrine of equivalents is not limited by prior art or prosecution history estoppel.112 Pioneer inventions are not analyzed under a different standard than other inventions, but the nature of pioneer inventions usually results in broad equivalents.113 A broader range of equivalents for pioneering inventions serves as an incentive for greater leaps in innovation. This incentive is necessary because research into a new area is expensive and time-consuming, and potentially unsuccessful.114 Hence, the inventor who succeeds in developing a pioneering invention deserves to be rewarded with a broad range of equivalents.115

C. Unpredictable Arts

Inventions in the unpredictable arts cannot capture after-arising technology through broad claims because of a failure to meet the enablement requirement of patentability. Due to the difficulty in meeting the enablement requirement, patents of unpredictable arts typically must have narrower claims than those covering the predictable arts.116 To meet the statutory enablement requirement, the inventor must describe the invention in enough detail to enable a person skilled in the art to make and use the invention without “undue experimentation.”117 One of the factors to consider when determining whether a disclosure requires

112. Augustine Med., 181 F.3d at 1301.

113. Texas Instruments, Inc. v. U.S. Int’l Trade Comm’n, 846 F.2d 1369, 1370 (Fed. Cir. 1988); see also Sun Studs, Inc. v. ATA Equip. Leasing, Inc., 872 F.2d 978, 987 (Fed. Cir. 1989) (“The ‘pioneer’ is not a separate class of invention, carrying a unique body of law. The wide range of technological advance between pioneering breakthrough and modest improvement accommodates gradations in scope of equivalency.”).


116. See In re Fisher, 427 F.2d 833, 839 (C.C.P.A. 1970); see also Enzo Biochem, Inc. v. Calgene, Inc., 188 F.3d 1362, 1376 (Fed. Cir. 1999) (finding that broad claims to genetic antisense technology, which had been rejected 10 times by PTO for lack of enablement before being allowed, were invalid for lack of enablement).

“undue experimentation” is the predictability or unpredictability of the
art of the invention.118

The Federal Circuit has routinely found broad claims in the
unpredictable arts to be invalid because they fail to meet the enablement
requirement.119 To meet the enablement requirement, the unpredictable
arts require a higher degree of disclosure.120 For example, biotechnology
and chemistry are considered unpredictable arts because scientists cannot
predict how minor chemical changes will affect chemical reactions or
physiological activities.121 Because of the unpredictability of the art, the
claims must more closely resemble the specific embodiments disclosed
in the specification.122 The inventor must present several working
examples of an unpredictable art to show that no undue experimentation
is required for a person skilled in the art to make and use the claimed
invention. A patent on an unpredictable art cannot support a broad claim
because application of the disclosed embodiments involves processes
that science does not yet fully understand.123 Thus, a patent application in
the unpredictable arts with broad claims will most likely fail the
enablement requirement.

In contrast, patents covering predictable arts require less specificity to
meet the enablement requirement. Inventions in the predictable arts, such
as mechanical or electrical inventions, can enable broad claims.124 One
embodiment of a predictable art supports a broad claim because
disclosure of one variation of the invention will enable a person skilled in
the art to substitute other variations without undue experimentation.125 In

118. In re Wands, 858 F.2d 731, 737 (Fed. Cir. 1988). The factors include:“(1) the quantity of
experimentation necessary, (2) the amount of direction or guidance presented, (3) the presence or
absence of working examples, (4) the nature of the invention, (5) the state of the prior art, (6) the
relative skill of those in the art, (7) the predictability or unpredictability of the art, and (8) the
breadth of the claims.”Id.

119. See Margaret Sampson, Comment: The Evolution of the Enablement and Written Description
Requirements Under 35 U.S.C. § 112 in the Area of Biotechnology, 15 BERKELEY TECH. L.J. 1233,

120. In re Vaeck, 947 F.2d 488, 496 (Fed. Cir. 1991) (stating that higher level of disclosure is
required where “diverse and poorly understood” subject matter is involved).

121. Sampson, supra note 119, at 1240 (citing In re Fisher, 427 F.2d 833, 839 (C.C.P.A. 1970)).

122. DURHAM, supra note 23, at 72.

123. Sampson, supra note 119, at 1248.

124. In re Fisher, 427 F.2d at 839; see also In re Spectra-Physics, Inc. v. Coherent, Inc., 827 F.2d
1524, 1533 (Fed. Cir. 1987).

125. In re Fisher, 427 F.2d at 839; Spectra-Physics, 827 F.2d at 1533.
the predictable arts, the claims do not have to closely mimic the disclosed embodiments to satisfy the enablement requirement.

IV. AFTER FESTO, PROSECUTION HISTORY ESTOPPEL PRECLUDES THE APPLICATION OF THE DOCTRINE OF EQUIVALENTS

In Festo Corp. v. Shoketsu Kinzoku Kogyo Kabushiki Co., the Federal Circuit held that a narrowing claim amendment that creates prosecution history estoppel completely bars application of the doctrine of equivalents to the amended claim element. The court reasoned that the complete bar would ensure the notice function of claims by eliminating the uncertainty in determining whether an element maintains a scope of equivalents after a narrowing claim amendment. The Federal Circuit’s strict approach gave rise to numerous dissents. The dissenting judges argued that the majority decision contradicted precedent, promoted copying, and discouraged technological innovation. The dissenters also stressed that the majority decision would hinder a primary justification for the doctrine of equivalents: providing patentees infringement protection against after-arising technology.

A. Procedural History of Festo

Festo involved an infringement lawsuit concerning two patents owned by the Festo Corporation (Festo). Festo sued Shoketsu Kinzoku Kogyo Kabushiki Co., Ltd. (also known as the SMC Corporation) and SMC Pneumatics, Inc. (collectively referred to as SMC), in the United States District Court for the District of Massachusetts for patent infringement. Festo’s patents disclosed an invention consisting of rodless cylinders that are magnetically coupled to a yoke or other structure. The district court

126. 234 F.3d 558 (Fed. Cir. 2000) (en banc), cert. granted, 69 U.S.L.W. 3779 (June 18, 2001) (No. 00-1543).
127. Id. at 564.
128. Id. at 576.
129. See infra Parts IV.C.1 and C.3.
130. See infra Part IV.C.2.
132. Id.
found infringement of Festo’s patents under the doctrine of equivalents. SMC appealed the findings of infringement to the Court of Appeals for the Federal Circuit, which in turn affirmed the district court’s judgments.

The United States Supreme Court granted a petition for a writ of certiorari by SMC. The Supreme Court vacated the Federal Circuit decision, and remanded the case to the Federal Circuit for further consideration in light of its holding in *Warner-Jenkinson*. On remand, the Federal Circuit affirmed the summary judgment with respect to one of Festo’s patents, but vacated the judgment in regard to the other, and remanded the case to the district court for re-determination.

The Federal Circuit granted a petition to rehear the appeal en banc. The court’s previous judgment was vacated and the accompanying judgment withdrawn. Although the court was presented with five questions, this Note addresses the third question: If a claim amendment creates prosecution history estoppel, under *Warner-Jenkinson* what range of equivalents, if any, is available under the doctrine of equivalents for the claim element so amended?

---

133. *Id.* at 1365.
134. *Id.*
135. *Id.*
136. *Id* at 1374.
137. *Id.* at 1380–81.
139. *Id.*
140. *Festo* Corp. v. Shoketsu Kinzoku Kogyo Kabushiki Co., 234 F.3d 558 (Fed. Cir. 2000) (en banc), *cert. granted*, 69 U.S.L.W. 3779 (June 18, 2001) (No. 00-1543). The court also addressed four additional questions:

First, the court held that “a narrowing amendment made for any reason related to the statutory requirements for a patent will give rise to prosecution history estoppel with respect to the amended claim element.” *Id.* at 566.

Second, the court stated that “a voluntary amendment that narrows the scope of a claim for a reason related to the statutory requirements for a patent will give rise to prosecution history estoppel as to the amended claim element.” *Id.* at 568.

Next, “when no explanation for a claim amendment is established, no range of equivalents is available for the claim element so amended.” *Id.* at 578.

Finally, the Federal Circuit declined to answer the fifth question because the court found no infringement under the doctrine of equivalents. *Id.* at 578.
B. The Majority Opinion

The *Festo* majority reasoned that “when a claim amendment creates prosecution history estoppel with regard to a claim element, there is no range of equivalents available for the amended claim element. Application of the doctrine of equivalents to the claim element is completely barred.” Since narrowing claim amendments were made during prosecution of the patents and *Festo* could not establish reasons unrelated to patentability for the amendments, no range of equivalents was available for the amended claim elements. Thus, SMC did not infringe either of *Festo*'s patents.

The majority reviewed Supreme Court precedent, but found no cases that had decided whether a claim element that was narrowed by an amendment that gave rise to prosecution history estoppel was entitled to a range of equivalents. According to the majority, *Warner-Jenkinson* did not speak directly to the issue. Rather, the majority found that *Warner-Jenkinson*’s holding applied only to a range of equivalents that is available when there is no explanation for the amendment in the prosecution history. Thus, the majority concluded that the Supreme Court had never ruled on the precise issue before the *Festo* court.

Consequently, the majority turned to Federal Circuit precedent for guidance, and identified two lines of authority—the strict approach and the flexible approach. The majority rejected the flexible approach for several reasons. First, it stated that after nearly twenty years of experience handling patent appeals, the flexible approach was unworkable because it failed to produce consistent results that could be relied upon by the marketplace. Second, the flexible approach opposed the notice function of claims and created uncertainty in patent law. In contrast, the strict approach would create certainty—once a narrowing

---

141. *Id.* at 563–64.
142. *Id.* at 564.
143. *Id.*
144. *Id.* at 571.
145. *Id.* at 569.
146. *Id.*
147. *Id.* at 571.
148. *Id.* at 573–74.
149. *Id.* at 574–75.
150. *Id.*
amendment for patentability is made, the patentee and the public have notice that the element is limited to literal infringement. 151 Hence, neither the public nor the patentee would need to pay for expensive litigation to determine the remaining scope of equivalents after a narrowing amendment. 152 Third, the strict approach would encourage technological innovation. 153 The certainty and predictability created by the strict approach would minimize the fear of litigation surrounding an amended claim’s remaining scope of equivalents. 154 Consequently, this would stimulate investment in research and lead to greater innovation. 155

C. The Dissents

The dissenting judges disagreed with the majority’s decision regarding the scope of equivalents available after prosecution history estoppel. The dissenters argued that Supreme Court and Federal Circuit precedent advocated a flexible approach. They also predicted that Festo would lead to an increase in blatant copying of patents, a decrease in the value of issued patents, a decline in technological innovation, and a limitation on the ability of the doctrine of equivalents to protect patentees from infringement by after-arising technology.

1. The Majority Opinion Contradicts Supreme Court and Federal Circuit Precedent

The dissenting judges rejected the majority’s claim that the Supreme Court had never spoken on the issue of the scope of equivalents after prosecution history estoppel. 156 Judge Michel reviewed eight Supreme Court cases dating back to the nineteenth century and concluded that they established that a claim element may retain a range of equivalents even after a patent applicant makes a narrowing claim amendment. 157

151. Id. at 577.
152. Id.
153. Id.
154. Id.
155. Id. at 577–78.
157. Id. at 602–09 (Michel, J., dissenting). Judge Michel noted that the Supreme Court cited these cases in the Warner-Jenkinson decision, thereby endorsing their analysis and holdings. Id. at 609 (Michel, J., dissenting).
The dissenting judges also argued that the majority decision conflicted with *Warner-Jenkinson*. According to the dissenters, the Supreme Court rejected a bright-line rule in *Warner-Jenkinson*, which would have precluded the application of the doctrine of equivalents to a claim element that had been narrowed during prosecution.

Several of the dissenting judges argued that Federal Circuit precedent overwhelmingly supported the flexible approach to prosecution history estoppel. Judge Michel characterized the majority position as a “sudden shift” in the court’s precedent because nearly all cases since 1983 took a flexible approach. According to Judge Michel, only two cases took the strict approach, and even those cases did not actually hold that narrowing amendments are always a complete bar to infringement by equivalents. Rather, the courts in those two cases took a flexible approach but found that the patentees had surrendered all of the pertinent subject matter.

2. *The Strict Approach Frustrates the Infringement Protection Against After-Arising Technology Provided by the Doctrine of Equivalents*

In his dissent, Judge Rader contended that the majority decision ignored the purpose of the doctrine of equivalents—to protect patentees from infringement by after-arising technology. Judge Rader stressed that a primary justification for the doctrine of equivalents is to protect patentees from infringement by after-arising technology. Without the doctrine of equivalents, copyists could easily circumvent infringement by substituting new technology for a claim element. Furthermore, he described *Warner-Jenkinson* as acknowledging that the doctrine of equivalents...
equivalents accommodates after-arising technology.\textsuperscript{166} Finally, Judge Rader argued that the majority opinion “defies logic” because “[a]ll patent protection for amended claims is lost when it comes to after-arising technology, while the doctrine of equivalents will continue to accommodate after-arising technology in unamended claims.”\textsuperscript{167} Such an outcome was illogical according to Judge Rader because after-arising technology, by definition, does not exist at the time of patent issuance; therefore, an applicant could not have given up such subject matter when a claim was amended.\textsuperscript{168}

3. \textit{The Strict Approach Will Promote Copying of Patented Inventions and Stifle Innovation}

The dissenters argued that the strict approach provides individuals with a way to copy patented inventions without fear of infringement liability.\textsuperscript{169} In Judge Michel’s view, a savvy copyist could search the prosecution history of a competitor and exploit an amended claim element by substituting an equivalent in his own product.\textsuperscript{170} In effect, copyists get a free ride, using the prosecution history as a roadmap to pirate the patentee’s invention.\textsuperscript{171}

The dissenters also contended that potential copying will decrease the value of issued patents and increase the cost of prosecuting a new patent.\textsuperscript{172} Unexpired patents, written under the assumption that courts would take a flexible approach to prosecution history estoppel, would no longer prevent copying.\textsuperscript{173} According to the dissenters, the cost of obtaining a patent will increase because it will become exceedingly difficult to obtain a patent protected by the doctrine of equivalents.\textsuperscript{174} Presently, patent applications frequently receive patentability rejections and the applicant makes narrowing amendments in response.\textsuperscript{175} However,
according to the majority’s opinion, an applicant who makes narrowing claim amendments also relinquishes the doctrine of equivalents in future litigation.\textsuperscript{176} Therefore, an applicant who wishes to overcome any patentability rejections and maintain the infringement protection provided by the doctrine of equivalents is faced with two expensive alternatives: appeal rejections to the PTO and the court\textsuperscript{177} or engage in extensive pre-filing prior art searches.\textsuperscript{178}

The dissenters warned that the ease of copying patented inventions could decrease technological innovation.\textsuperscript{179} According to the dissenters, inadequate protection of patents would lead to a loss of potential profits, which in turn would lead to a decrease in investment for new technological research.\textsuperscript{180} In addition, incentives to license inventions would be eliminated because the potential licensee could simply copy the invention without fear of infringement.\textsuperscript{181} Furthermore, the costs involved in an application appeals process would force some inventors to abandon their inventions or maintain them as trade secrets.\textsuperscript{182} In particular, these extra costs of prosecution would hurt the individual inventors and startup companies that the patent system was meant to protect.\textsuperscript{183} In the end, according to the dissenters, the extra cost would also harm the public, as it is the public who suffers when an inventor fails to disclose an invention.\textsuperscript{184}

V. \textit{Festo} UNDERMINES THE PURPOSE OF THE DOCTRINE OF EQUIVALENTS AND LEAVES PATENTEES WITHOUT RE COURSE AGAINST INFRINGEMENT BY AFTER-ARISING TECHNOLOGY

The \textit{Festo} court erred in implementing the strict approach to prosecution history estoppel as it virtually eliminates infringement protection against after-arising technology provided by the doctrine of

\begin{itemize}
  \item \textsuperscript{176} Id. at 563–64.
  \item \textsuperscript{177} Id. at 618 (Michel, J., dissenting).
  \item \textsuperscript{178} Id. at 624 (Linn, J., dissenting).
  \item \textsuperscript{179} See id. at 624 (Linn, J., dissenting), 641 (Newman, J., dissenting).
  \item \textsuperscript{180} Id. at 627 (Linn, J., dissenting), 641 (Newman, J., dissenting).
  \item \textsuperscript{181} Id. at 619 (Michel, J., dissenting).
  \item \textsuperscript{182} Id. at 618 (Michel, J., dissenting).
  \item \textsuperscript{183} Id. at 624 (Linn, J., dissenting).
  \item \textsuperscript{184} Id. at 618 (Michel, J., dissenting); see also id. at 621, 624 (Linn, J., dissenting), 641 (Newman, J., dissenting).
\end{itemize}
equivalents. The *Festo* majority misread *Warner-Jenkinson* regarding the scope of equivalents after prosecution history estoppel because *Warner-Jenkinson* implies that a scope of equivalents is available after a narrowing claim amendment has been made for a patentability reason. In addition, *Festo* leads to absurd results in the context of after-arising technology. Moreover, the *Festo* decision conflicts with patent law’s policy of rewarding pioneer inventions. The Supreme Court should overrule *Festo* and apply the flexible approach to prosecution history estoppel in cases of infringement by after-arising technology.

A. *Festo* Leaves Patentees Vulnerable to Infringement by After-Arising Technology

*Festo*’s strict approach to prosecution history estoppel frustrates the ability of the doctrine of equivalents to protect patentees against after-arising technology. A primary purpose of the doctrine of equivalents is to provide protection against after-arising technology. This protection is necessary because, by definition, after-arising technology is unknown at the time the claims are drafted.

Under *Festo*, a patentee loses the protection provided by the doctrine of equivalents if the claims are amended during prosecution of the patent. Yet, patent applications are amended routinely in present patent practice. Therefore, most patents will not enjoy the protection against after-arising technology provided by the doctrine of equivalents.

It is unclear whether the *Festo* majority considered the effect of its decision on infringement protection against after-arising technology. Nowhere in its opinion does the majority address the clear need for the doctrine of equivalents to accommodate after-arising technology. The court failed to correctly apply Supreme Court and Federal Circuit precedent, which states that the doctrine of equivalents protects patentees from infringement by after-arising technology. In addition, the majority failed to respond to Judge Rader’s dissent, which warned of the

---


186. *Festo*, 234 F.3d at 563–64.

187. *See supra* note 92 and accompanying text.
absurd results that *Festo* would have when applied to cases of after-arising technology.  

The *Festo* majority’s assertion that the increase in notice by the strict approach would promote innovation is incorrect. The *Festo* majority concluded that the “complete bar” rule would promote innovation by giving researchers certainty as to the scope of a patent’s claims.  

A claim that was narrowed during prosecution would be limited to its literal terms, thereby permitting inventors to design around the patent without fear of infringement litigation over the remaining scope of equivalents.

However, the “complete bar” rule will actually encourage copying of patented inventions and thus hinder future innovation. The purpose of the patent system is to encourage technological innovation by protecting a patentee’s inventions. One dissenting judge cited empirical studies showing that a loss of potential profits due to weak patent protection reduces the amount of investment in new technology research. By restricting the availability of the doctrine of equivalents, *Festo* decreases the protection provided by a patent and, thus, diminishes the financial value of a patent. Consequently, the incentive to obtain a patent is reduced, discouraging public disclosure of innovation.

Moreover, the *Festo* majority’s “complete bar” rule fails to increase the notice function of claims in the context of after-arising technology. The notice function of claims is inherently murky, especially when involving after-arising technology. The *Festo* court concluded that when claims are restricted to their literal terms, uncertainty regarding infringement is eliminated. By definition, after-arising technology was not known at the time the claims were drafted. Therefore, even when only literal infringement applies, it is unclear whether the claims cover

---

188. *See id.* at 619–20 (Rader, J., dissenting).
189. *Id.* at 577.
190. *Id.*
195. *Festo*, 234 F.3d at 575–76.
after-arising technology because the claims were drafted without after-arising technology in mind.

B. Festo Did Not Appropriately Apply Warner-Jenkinson

The Festo majority failed to consider the language of the Warner-Jenkinson decision regarding the scope of equivalents available after prosecution history estoppel. The majority read Warner-Jenkinson as only addressing cases of unexplained amendments, and concluded that the opinion did not speak to the issue of a range of equivalents after prosecution history estoppel. In such a situation, the doctrine of equivalents would be completely barred. The patentee could only avoid the complete bar rule by establishing that the amendment was made for reasons unrelated to patentability. In effect, the Festo majority drew Warner-Jenkinson’s “complete bar” line between reasons related to patentability and reasons unrelated to patentability. According to the Festo majority, a patentee must prove that a claim element was amended for reasons unrelated to patentability; if the patentee fails this burden, then the prosecution history bars the doctrine of equivalents for that element.

Warner-Jenkinson holds that the complete bar to the doctrine of equivalents applies only when the Warner-Jenkinson presumption arises and cannot be rebutted. In other words, when a patentee cannot prove why a claim element was narrowed, a court should presume that the claim was amended for a reason related to patentability and therefore the claim should have no scope of equivalents. If neither the prosecution history nor the patentee can show the reason for an amendment, a court has no information to determine what subject matter was relinquished by

196. Id.
197. Id.
198. Id. at 586.
199. Id. Judge Gajarsa, a member of the Festo majority, argued in another case that Warner-Jenkinson equates amendments made for unknown reasons with amendments made for patentability reasons, both of which should create a complete bar to the doctrine of equivalents. Litton Sys., Inc. v. Honeywell, Inc., 145 F.3d 1472, 1475 (Fed. Cir. 1998) (Gajarsa, J., dissenting).
200. Festo, 234 F.3d at 586.
202. Id. at 33.
the patent applicant. In such a situation, in deference to the policy of notice, the court must assume the worst case scenario—that the applicant gave up all scope of equivalents for that claim element.

In addition, the *Festo* majority ignored what *Warner-Jenkinson* suggested regarding the scope of equivalents available for a claim element following a patentability amendment. As the majority correctly stated, *Warner-Jenkinson*’s holding did not directly address this issue. However, as Federal Circuit cases and commentators have recognized, *Warner-Jenkinson* implies that a scope of equivalents is available after a narrowing claim amendment has been made for reasons related to patentability. The Supreme Court rejected a bright-line rule invoking a complete bar regardless of the reasons for a claim amendment. Such a rule would upset the expectation of patent applicants and the PTO that a flexible approach to estoppel would be available. The PTO rejects patent applications based on the expectation that narrowing amendments made by the patent applicant will not automatically preclude all range of equivalents. The Court avers that the reasons for a claim amendment must be explored.

C. The *Festo* Decision Leads to Odd Results in Protecting Patentees from Infringement by After-Arising Technology

*Festo*’s complete bar rule leads to absurd results when applied to after-arising technology. While two claims may have identical scope, an amended claim will be precluded from the doctrine of equivalents while an unamended claim will not. Also, *Festo* creates inequality between the predictable arts and the unpredictable arts. Because patent applicants in the unpredictable arts cannot write broad claims that meet the enablement requirement, these inventions will receive less protection than predictable arts.


204. *Sextant Avionique*, 172 F.3d at 831–32.

205. *See supra* notes 145–147 and accompanying text.

206. *See supra* note 81.

207. *See supra* note 82.

208. *See* *Warner-Jenkinson*, 520 U.S. at 32.

209. *See id.* at 32 n.6.

210. *See id.* at 32.

211. *See id.* at 33 n.7.
Festo and After-Arising Technology

1. **Festo Protects Unamended Claims but Not Amended Claims**

The primary purpose of the doctrine of equivalents is to protect patentees from infringement by after-arising technology.\(^{212}\) Festo creates incongruous results because it protects patents with unamended claims against after-arising technology, but not patents with amended claims.\(^{213}\) Patentees cannot rely on literal infringement to protect their rights against technology that was not known at the time the patent was drafted.

Patentees should be able to bring an infringement suit under the doctrine of equivalents against future technologies regardless of whether a claim was amended. A patent applicant could not have given up subject matter that was not known at the time of the claim amendment. A patent system must include methods to handle after-arising equivalents that were not known, and thus not claimable, at the time the meaning of the claims was fixed.\(^{214}\) Because the meaning of claims is fixed at the issue date, the doctrine of equivalents must be available to all patentees to protect themselves against after-arising technology.

2. **After Festo, Predictable Arts Can Still Use Literal Infringement for Protection from After-Arising Technology, but Unpredictable Arts Cannot**

A patentee barred by Festo from applying the doctrine of equivalents may still find infringement under literal infringement if the invention is a predictable art, but would have no such opportunity if the art is unpredictable. Festo creates unequal protection for different types of technologies. The complete bar rule of Festo leaves patentees of unpredictable arts with no means of enforcing their rights against infringers using after-arising technology.

Patents in the unpredictable arts depend on the doctrine of equivalents for protection against infringement by after-arising technology. The patentee cannot disclose the after-arising technology because it was not known at the time of patent issuance. Because the claims of unpredictable arts are limited to the specific embodiments disclosed in the specification, a patent in the unpredictable arts cannot reach after-

\(^{212}\) See Adelman, supra note 185, at 1023.

\(^{213}\) 234 F.3d 558, 619 (Fed. Cir. 2000) (en banc) (Rader, J., dissenting), cert. granted, 69 U.S.L.W. 3779 (June 18, 2001) (No. 00-1543).

\(^{214}\) See Adelman, supra note 185, at 1023.
arising technology through literal infringement.\textsuperscript{215} The patent holder of an unpredictable art will have to rely on the doctrine of equivalents for protection from an equivalent based on after-arising technology. If a claim of an unpredictable art has been amended during prosecution for reasons related to patentability, \textit{Festo} would completely bar the application of the doctrine of equivalents. If the invention was in a predictable art, inventors may still be able to protect their rights through literal infringement of a broad claim, but the inventor in an unpredictable art would have no such recourse.

Inventions in the predictable arts, such as mechanical or electrical inventions, can enable broad claims that capture after-arising technology under literal infringement. Because predictable arts can enable broad literal claims, they may be able to cover the after-arising technology through literal infringement.\textsuperscript{216} For example, in \textit{Hughes}, the after-arising technology could have been covered literally by a broad claim.\textsuperscript{217} The accused device performed the same function, using the same operation as the patented invention.\textsuperscript{218} The only difference was the physical location where the function was being performed.\textsuperscript{219} The patent drafter could have drafted broad claims that included computing the positioning figures by a ground station or a microprocessor on board the satellite.\textsuperscript{220} Because electronics is a predictable art, the claims likely would have been enabling.\textsuperscript{221} Had the patent in \textit{Hughes} contained broad literal claims, the patentee may not have needed the doctrine of equivalents to achieve a finding of infringement against the after-arising technology.

A lack of protection against after-arising technology in the unpredictable arts would decrease the value of these patents and discourage innovation. Research into unpredictable arts, such as biotechnology and pharmaceuticals, is expensive and time consuming and there is no guarantee of success.\textsuperscript{222} Companies will be less willing to spend money on projects involving the unpredictable arts due to an

\begin{itemize}
\item \textsuperscript{215} ADELMAN, \textit{supra} note 3, § 3.4, at 3-40.18.
\item \textsuperscript{216} ADELMAN, \textit{supra} note 3, § 3.4, at 3-40.18.
\item \textsuperscript{217} Adelman & Francione, \textit{supra} note 55, at 712–14.
\item \textsuperscript{218} Hughes Aircraft Co. v. United States, 140 F.3d 1470, 1475 (Fed. Cir. 1998).
\item \textsuperscript{219} Id.
\item \textsuperscript{220} Adelman & Francione, \textit{supra} note 55, at 713–14.
\item \textsuperscript{221} Adelman & Francione, \textit{supra} note 55, at 713–14.
\end{itemize}
inability to protect their investments against after-arising technology. The Festo “complete bar” rule permits competitors in the unpredictable arts to make an insubstantial substitution to a claim element using after-arising technology without fear of infringement litigation.

D. The Festo Decision Conflicts with the Policy of Rewarding Pioneer Inventions

Festo frustrates the patent law policy of rewarding inventors for pioneering inventions. Pioneer inventions gain the benefit of broader claims because the claims can be drafted without being limited by prior art.\(^\text{223}\) Also, during the litigation of pioneer inventions, the doctrine of equivalents will be limited neither by prior art nor by prosecution history estoppel based on prior art rejections.\(^\text{224}\) The Federal Circuit recognizes that pioneer inventions are justly rewarded for their contribution to science by broad claim coverage and no limits to the doctrine of equivalents.\(^\text{225}\)

Festo may deny an inventor the reward of broad claim interpretation for a pioneering invention. Under Festo, narrowing claim amendments made in response to prior art rejections and Section 112 rejections will preclude the doctrine of equivalents in future litigation. By definition, pioneer inventions do not encounter PTO rejections based on prior art, but may make amendments to overcome patentability rejections under Section 112 because of the difficulty in adequately describing a new area of science.\(^\text{226}\) Thus, under Festo, an inventor who needs to make a narrowing amendment due to the inherent difficulty of drafting claims for pioneering inventions may not be able to use the doctrine of equivalents in future litigation. Festo’s bright-line rule unfairly restricts the rights of pioneer inventors to literal infringement only. However, because nearly all infringement of pioneering inventions will be from after-arising technology,\(^\text{227}\) pioneer inventors will rely on the doctrine of equivalents during infringement litigation. Thus, under Festo, an inventor of a pioneering invention may have less protection than

\(^{223}\) Augustine Med., Inc. v. Gaymar Indus., Inc., 181 F.3d 1291, 1301 (Fed. Cir. 1999).
\(^{224}\) Id. at 1301–02.
\(^{225}\) Id.
\(^{227}\) ADELMAN, supra note 3, § 3.4[1], at 3-42.30(22).
someone who makes only an incremental improvement in a well-developed field of science. Such an outcome contradicts the Federal Circuit’s policy that pioneering patent claims deserve a broader scope as a reward for venturing into new areas of science.

E. Festo Should Have Permitted the Flexible Approach to Prosecution History Estoppel in Cases of Infringement by After-Arising Technology

The Supreme Court should overrule Festo and apply the flexible approach to prosecution history estoppel when faced with accused infringement by after-arising technology. The purpose of the patent system is to encourage technological innovation by protecting patentees’ inventions. In cases of after-arising technology, patentees cannot rely on literal infringement for protection; the claims may not literally cover the after-arising technology at the time the claims were drafted. Therefore, patentees need the doctrine of equivalents if they are to protect themselves against after-arising technology. The flexible approach gives a court the opportunity to determine if the patent still covers the after-arising technology even after the patentee has made a narrowing amendment for reasons related to patentability. The fact that a claim was amended during the application process should not mindlessly prohibit the doctrine of equivalents against infringing equivalents that gain the benefit of new technology.

The Festo majority’s decision to limit infringement protection against after-arising technology provided by the doctrine of equivalents is troublesome. While promoting certainty in the patent system and emphasizing the notice function of claims is important, these goals should not be accomplished by completely sacrificing patentees’ protection against after-arising technology. Also, the majority completely failed to consider the effect its decision would have in the context of after-arising technology. A flexible approach would maintain the proper balance between the notice function of claims and the fair protection of patentees. Instead, the majority precluded application of the doctrine of equivalents.


229. See ADELMAN, supra note 3, § 3.5, at 3–42.56(14)–(16) (advocating scope of equivalents in cases of after-arising technology); see also Phillips, supra note 111, at 180–81 (arguing that doctrine of equivalents should be reserved only for equivalents that arise after filing of the patent application).
equivalents in many cases and, consequently, left patentees exposed to copyists clever enough to use after-arising technology to avoid infringement.

VI. CONCLUSION

The *Festo* decision runs contrary to the major function of the doctrine of equivalents—protecting patent holders from after-arising technology. Supreme Court and Federal Circuit precedent expressly advocate maintaining the doctrine of equivalents so that patentees can take action against accused infringers using new technologies. The *Festo* majority created a “complete bar” rule to promote certainty in the patent system and to emphasize the notice function of patent claims. The court failed to consider important language in *Warner-Jenkinson*—as well as Federal Circuit precedent interpreting *Warner-Jenkinson*—that permits a scope of equivalents after a narrowing claim amendment for reasons related to patentability.

The *Festo* court failed to consider the impact of its decision in patent suits against accused devices using after-arising equivalents. The *Festo* decision prejudices patents in the unpredictable arts, which cannot enable broad claims as easily as those in the predictable arts. The application of the complete bar rule also runs contrary to the policy that pioneer inventions should be rewarded with broad claim coverage.

The Supreme Court should overrule *Festo*’s complete bar rule and allow the flexible approach to prosecution history estoppel in all cases of infringement by after-arising technology. The Court needs to correct the imbalance that *Festo* created between the notice function of claims and the fair protection of patented inventions.
HOLDING TORTFEASORS ACCOUNTABLE:
APPORTIONMENT OF ENHANCED INJURIES UNDER
WASHINGTON’S COMPARATIVE FAULT SCHEME

Ryan P. Harkins

Abstract: The enhanced-injury doctrine imposes a negligence-based duty to reasonably minimize the foreseeable risk of injury enhancement in the event of primary accidents, regardless of their cause. When apportioning responsibility for enhanced injuries under principles of comparative fault, a majority of courts outside of Washington use a plaintiff’s fault in causing the primary accident to reduce recovery for enhanced injuries. A minority of courts, however, rule that because the enhanced-injury doctrine presupposes the occurrence of primary accidents, primary fault is legally irrelevant to apportionment of enhanced injuries. Washington courts have not addressed this issue. This Comment argues that Washington courts should not consider primary fault when apportioning responsibility for enhanced injuries. First, primary fault is not “fault” with respect to enhanced injuries and, therefore, should not reduce a plaintiff’s recovery for enhanced injuries. Moreover, Washington’s comparative fault scheme and law regarding proximate cause require that responsibility for enhanced injuries be apportioned using enhanced injury fault rather than primary fault. Finally, using primary fault to reduce an enhanced injury recovery will result in decreased product safety.

Mary did not see the stop sign.1 In the back seat, her two children would not stop making noise. Tired and irritated, she turned around to quiet them as she approached the four-way stop. At fifteen miles per hour, her automobile passed into the intersection without stopping and was struck from the side by an automobile traveling at ten miles per hour. The collision dented the two vehicles and all of the passengers suffered minor bumps and bruises. Mary’s negligence was the sole cause of the collision and, for that reason, she is liable for the dented vehicles and the passengers’ minor injuries.

Now assume that due to a defect in the gasoline tank, Mary’s vehicle exploded when struck by the other automobile. The fire engulfed Mary’s vehicle and severely burned all inside. Instead of suffering only minor bumps and bruises, Mary suffered disfiguring and disabling burns and her two children were killed. Assume further that the fire would not have occurred if the gasoline tank had not been defective. Hence, while the gasoline tank defect did not cause the primary accident, it did lead to the

---

1. Hypothetical created by the author.
fire. The gasoline tank defect enhanced the passengers’ injuries from mere bumps and bruises to severe burns.\textsuperscript{2}

Under the doctrine of enhanced injuries, the automobile manufacturer will be liable for the enhanced injuries—the burns. The doctrine imposes on product manufacturers a negligence-based duty to use reasonable care in the design and construction of their products to minimize foreseeable risks of harm.\textsuperscript{3} As in negligence cases generally, the elements of an enhanced-injury cause of action are: (1) a duty to reasonably minimize foreseeable risks of harm, and (2) a breach of that duty (3) which proximately causes the plaintiff (4) to suffer identifiable damages.\textsuperscript{4}

In the present example, it is foreseeable that automobiles will be involved in accidents with or without the fault of the plaintiff-driver. Hence, Mary’s automobile manufacturer owed her a duty to reasonably design and construct her vehicle to minimize risks of harm in such accidents. Production of the defective gasoline tank design constituted a breach of that duty and proximately caused the enhanced injuries. Under the doctrine of enhanced injuries, the manufacturer is liable only for the burns, however, and not for the bumps and bruises, because the bumps and bruises were proximately caused solely by Mary’s negligence.

Another layer of analysis is necessary in Washington because Washington courts apply comparative fault, allowing proportionate reduction of a plaintiff’s recovery according to the percentage of plaintiff

\textsuperscript{2} See infra Part I. This Comment employs several terms of art. The initial foreseeable accident—the side-impact automobile collision resulting from Mary’s negligence—is termed the primary accident. The fault attributable to the primary accident—Mary running the stop sign—is primary fault. The second accident—the gasoline tank explosion resulting from the defect in the gasoline tank—is the enhanced-injury accident. The fault attributable to the enhanced-injury accident—the manufacturer’s fault in producing a defective gasoline tank—is enhanced-injury fault. Other terms, such as “second collision,” or “crashworthiness,” have been used by courts and commentators to describe these same concepts. See, e.g., Trust Corp. of Mont. v. Piper Aircraft Corp., 506 F. Supp. 1093, 1094 (D. Mont. 1981) (second collision); Reed v. Chrysler Corp., 494 N.W.2d 224, 225–30 (Iowa 1992) (crashworthiness). As one commentator has pointed out, however, those terms inaccurately imply that enhanced-injury doctrine applies only to automobile collisions. See Thomas V. Harris, Enhanced-Injury Theory: An Analytic Framework, 62 N.C. L. Rev. 643, 647–51 & n.22 (1984). While the hypothetical example and most cases discussed in this Comment involve automobile collisions, the enhanced-injury doctrine applies to product liability cases generally. See Larsen v. Gen. Motors Corp., 391 F.2d 495, 504 (8th Cir. 1968) (enhanced injury liability applies “to all manufacturers”); Baumgardner v. Am. Motors Corp., 83 Wash. 2d 751, 758, 522 P.2d 829, 833 (1974) (holding that “manufacturers” can be liable for product defects that proximately cause enhanced injuries); Harris, supra, at 647–51.

\textsuperscript{3} See Larsen, 391 F.2d at 499–506; Baumgardner, 83 Wash. 2d at 756–57, 522 P.2d at 832–33.

\textsuperscript{4} Baumgardner, 83 Wash. 2d at 758, 522 P.2d at 833; Harris, supra note 2, at 651.
fault that proximately caused the plaintiff’s damages. Using comparative fault, the manufacturer will argue that Mary’s fault in causing the primary accident should proportionately reduce its monetary responsibility for the burns suffered by Mary and her children. This raises the question whether a Washington court should reduce the damages awarded by the percentage of fault attributed to Mary in causing the primary accident or find that Mary’s fault in causing the primary accident is legally irrelevant to the issue of damages resulting from the enhanced-injury burns.

This Comment focuses solely on whether Washington courts should consider primary fault when apportioning responsibility for enhanced injuries under Washington’s scheme of comparative fault. Prior to addressing this question, the trier-of-fact in an enhanced-injury case must first find that all elements of an enhanced-injury claim are met. This Comment argues that Washington courts should not use primary fault to reduce a plaintiff’s recovery for enhanced injuries. Part I describes the adoption of the enhanced-injury doctrine in Washington State. Part II discusses the development of Washington’s comparative fault scheme. Part III explains the conflicting resolutions of this issue outside of Washington. Part IV argues that reducing a plaintiff’s enhanced-injury recovery by primary fault is inconsistent with the enhanced-injury doctrine itself, is not required by Washington’s comparative fault scheme, and would lead to decreased product safety.

I. THE ENHANCED-INJURY DOCTRINE IN WASHINGTON

In 1974, the Supreme Court of Washington adopted the enhanced-injury doctrine. In *Baumgardner v. American Motors Corp.*, the plaintiff sued American Motors after his wife was killed in an automobile accident. Upon collision, her seat broke loose, crushing her

5. See infra Section II.
6. The seminal case recognizing the enhanced-injury doctrine was *Larsen v. General Motors Corp.*, 391 F.2d 495 (8th Cir. 1968). Prior to *Larsen*, courts were reluctant to permit an enhanced-injury cause of action. See, e.g., Evans v. Gen. Motors Corp., 359 F.2d 822, 825 (7th Cir. 1966); Campo v. Scofield, 95 N.E.2d 802, 805–06 (N.Y. 1950); Yaun v. Allis-Chalmers Mfg. Co., 34 N.W.2d 853, 858–59 (Wis. 1948). However, the *Larsen* position has been adopted by virtually all American courts. See, e.g., Blankenship v. General Motors Corp., 406 S.E.2d 781, 782 (W.Va. 1991) (noting that West Virginia became the last state to adopt the enhanced injury doctrine).
8. *Id.* at 752, 522 P.2d at 830.
between the seat and seat belt. Furthermore, her seat belt could not be unlatched after the accident, severely exacerbating her injuries. The plaintiff alleged that defects in his wife’s seat and seatbelt buckle, while not contributing to the cause of the original collision, caused or enhanced his wife’s injuries, thereby causing her death. The trial court granted summary judgment in favor of American Motors and the plaintiff appealed. The Supreme Court of Washington reversed, holding that a manufacturer can be liable for product defects that proximately cause enhanced injuries even though the defects do not cause the accident.

The Baumgardner court held that the enhanced-injury doctrine rests on common law negligence principles. Generally, the elements of a negligence cause of action are duty, breach, causation, and damages. Therefore, the elements in an enhanced-injury cause of action are: (1) a duty to reasonably prevent foreseeable risks of enhanced injuries, and (2) a breach of that duty (3) that proximately caused (4) distinct and identifiable enhanced injuries.

Because the enhanced-injury doctrine is based on negligence principles, the Baumgardner court stated that the primary determination in an enhanced-injury claim is the imposition of a duty to minimize the risk of injury enhancement. General negligence principles impose a duty to use reasonable care to protect others from foreseeable risk of harm. The imposition of a duty turns on foreseeability. Therefore, the imposition of an enhanced-injury duty turns on foreseeability. If a

9. Id.
10. Id.
11. Id. at 752–53, 522 P.2d at 830.
12. Id. at 753, 522 P.2d at 830.
13. Id. at 760, 522 P.2d at 834.
14. Id. at 758, 522 P.2d at 833.
15. Id. at 757–58, 522 P.2d at 833. In addition, Baumgardner held that liability for enhanced injuries may also be established under strict liability. Id. at 759, 522 P.2d at 834. Proving an enhanced injury claim under strict liability is beyond the scope of this Comment.
17. Baumgardner, 83 Wash. 2d at 758, 522 P.2d at 833; see also Degel, 129 Wash. 2d at 48, 914 P.2d at 731; KEETON ET AL., supra note 16, § 30, at 164–65.
18. Baumgardner, 83 Wash. 2d at 756–57, 522 P.2d at 832.
20. See Baumgardner, 83 Wash. 2d at 754–57, 522 P.2d at 831–33; see also KEETON ET AL., supra note 16, § 43, at 280–81.
manufacturer should foresee that its product creates a risk of enhancing the injuries suffered in a primary accident, the manufacturer has a duty to reasonably minimize that risk.21

Applying this foreseeability analysis, the Baumgardner court concluded that it is clearly foreseeable to the manufacturer that automobile collisions will occur22 and that the magnitude of injury to persons involved in such collisions will often depend on the design and construction of the automobile.23 Therefore, the manufacturer has a duty to design and construct its automobiles to reasonably minimize the risk of enhanced injuries in such collisions.24 This duty is imposed despite the fact that automobiles are not made for the purpose of colliding with other objects, because people cannot use automobiles without encountering the foreseeable risk of injury-producing collisions.25

The Baumgardner court recognized that the enhanced-injury duty is not unlimited. It emphasized that a manufacturer’s duty is to reasonably minimize foreseeable risks of harm.26 It also emphasized that, under basic negligence principles, a manufacturer is liable only for injuries or enhancement of injuries proximately caused27 by product defects.28 Therefore, a breach of an enhanced-injury duty makes a manufacturer liable only for the injury enhancement—i.e., that portion of the plaintiff’s damages that would not have occurred had the manufacturer exercised reasonable care.29

Since Baumgardner, Washington courts have done little to further define the parameters of the enhanced-injury doctrine and its interaction with proximate cause.25

---

21. See Baumgardner, 83 Wash. 2d at 754–57, 522 P.2d at 831–33; see also Keeton et al., supra note 16, § 43, at 280–81.
22. Baumgardner, 83 Wash. 2d at 755, 522 P.2d at 831–32 (quoting Larsen v. Gen. Motors Corp., 391 F.2d 495, 501–02 (8th Cir. 1968)).
23. Id. at 755, 522 P.2d at 832 (quoting Mickle v. Blackmon, 166 S.E.2d 173, 185 (S.C. 1969)).
24. The question faced by the Baumgardner court was whether to adopt the enhanced-injury doctrine in automobile collision cases. However, the court phrased its holding in much broader language, implying that the enhanced-injury doctrine applies to product liability cases generally, as recognized by Larsen. Specifically, the Baumgardner court held that “a manufacturer can be held liable in negligence for design or manufacture defects which proximately cause enhanced injuries due to such defects.” Baumgardner, 83 Wash. 2d at 758, 522 P.2d at 833.
25. Id. at 754–55, 522 P.2d at 831–32 (quoting Larsen, 391 F.2d at 501–02).
26. Id. at 756–57, 522 P.2d at 832–33.
27. For a discussion of proximate cause, see infra Part III.B.
28. Baumgardner, 83 Wash. 2d at 758, 522 P.2d at 833.
29. Id.
with other tort principles.\textsuperscript{30} In \textit{Bernal v. American Honda Motor Co.},\textsuperscript{31} the Supreme Court of Washington recognized that \textit{Baumgardner} established an enhanced-injury cause of action.\textsuperscript{32} Furthermore, in \textit{Couch v. Mine Safety Appliances Co.},\textsuperscript{33} the court reiterated that, as in negligence cases generally, the plaintiff must prove the elements of an enhanced-injury claim, including the nature and extent of the injury enhancement.\textsuperscript{34} However, Washington case law does not provide guidance regarding the role of comparative fault in an enhanced-injury claim. While \textit{Baumgardner} established the negligence-based framework for the enhanced-injury doctrine generally,\textsuperscript{35} no Washington court has explicitly addressed whether a plaintiff’s fault in causing the primary accident should be considered for purposes of reducing recovery for an enhanced injury.\textsuperscript{36}

II. TORT LAW CONCEPTS IN WASHINGTON

Because the enhanced injury doctrine is based on negligence principles, several tort law concepts affect the application of the doctrine to individual cases. The applicable tort concepts include comparative fault, the determination of duty, and proximate causation. This Section provides background knowledge necessary to understand how those concepts shape the application of the enhanced-injury doctrine to a particular case.

\begin{itemize}
\item[31.] 87 Wash. 2d 406, 553 P.2d 107 (1976).
\item[32.] \textit{Id.} at 411, 415, 553 P.2d at 110, 112.
\item[33.] 107 Wash. 2d 232, 728 P.2d 585 (1986).
\item[34.] \textit{Id.} at 243, 728 P.2d at 591; see also \textit{Baumgardner}, 83 Wash. 2d at 758, 522 P.2d at 833.
\item[35.] \textit{Baumgardner}, 83 Wash. 2d at 757–58, 522 P.2d at 833.
\item[36.] See \textit{Couch}, 107 Wash. 2d at 246, 728 P.2d at 592–93 (declining to rule on whether plaintiff’s fault in causing primary collision was relevant in determining cause of enhanced injury because jury found plaintiff’s comparative fault was not a proximate cause of his enhanced injuries); Amend v. Bell, 89 Wash. 2d 124, 130–34, 570 P.2d 138, 142–44 (1977) (holding that evidence of seatbelt use is inadmissible in Washington and, consequently, not addressing issue of whether, under comparative fault, a plaintiff’s primary fault can reduce recovery for enhanced injuries).
\end{itemize}
Apportionment of Enhanced Injuries

A. Washington’s Scheme of Comparative Fault

Under Washington’s scheme of comparative fault, tort defendants can claim that the plaintiff was also at fault in order to eliminate or reduce the defendant’s responsibility for damages. Proving plaintiff fault involves the same elements of proof required to prove a defendant’s fault. The defendant must prove that (1) the plaintiff owed herself a duty, (2) breached that duty, (3) and proximately caused (4) damages to herself. Over the past thirty years, the Washington State Legislature has repeatedly changed the legal effect that a finding of plaintiff fault has on a plaintiff’s claim.

Prior to 1973, a plaintiff’s contributory negligence completely barred recovery for damages in a negligence cause of action in Washington. In 1973, the Washington State Legislature abolished this complete bar to recovery by adopting a system of pure comparative fault. Thus, the effect of plaintiff negligence changed from a complete bar to a reduction of recovery in proportion to the plaintiff’s percentage of negligence. However, the 1973 act did not affect all tort causes of action; the act referred only to negligence claims and, therefore, did not apply to all product liability causes of action.

---

37 See Wash. Rev. Code § 4.22.015 (2000) ("Fault includes acts or omissions . . . that are in any measure negligent or reckless toward the person or property of the actor or others . . . ."); see also Geschwind v. Flanagan, 121 Wash. 2d 833, 838, 854 P.2d 1061, 1064 (1993) (quoting Seattle-First Nat’l Bank v. Shoreline Concrete Co., 91 Wash.2d 230, 238, 588 P.2d 1308, 1314 (1978) (“A plaintiff’s negligence relates to a failure to use due care for his own protection . . . .”)); Keeton et al., supra note 16, § 65, at 451 ("Contributory negligence is conduct on the part of the plaintiff, contributing as a legal cause to the harm he has suffered, which falls below the standard to which he is required to conform for his own protection.").


40 Ch. 137, §1, 1973 Wash Laws 1st Ex. Sess. 949 (codified at Wash. Rev. Code § 4.22.010 (1973)) (repealed 1981) ("Contributory negligence shall not bar recovery in an action . . . but any damages allowed shall be diminished in proportion to the percentage of negligence attributable to the party recovering.").

41 Godfrey, 84 Wash. 2d at 965, 530 P.2d at 633.

In 1981 the Washington Legislature altered Washington’s comparative fault landscape again when it passed the Product Liability and Tort Reform Act.\(^3\) The 1981 Act codified Washington product liability law, providing that there shall be a single “product liability claim.”\(^4\) In addition, the Act recodified Washington’s scheme of comparative fault, providing that the scheme would apply to all causes of action based on “fault”\(^5\) and defining “fault” to include both “misuse of a product” and conduct subjecting a party to liability under a product liability cause of action.\(^6\) Therefore, unlike its 1973 predecessor, the 1981 Act explicitly applied Washington’s comparative fault scheme to all product liability causes of action.\(^7\)

The 1981 Act provided that a claimant’s fault proportionately diminishes recovery of “compensatory damages for an injury attributable to the claimant’s . . . fault.”\(^8\) In addition, it provides that legal causation requirements\(^9\) apply when considering fault for both plaintiffs and defendants under Washington’s comparative fault scheme.\(^10\) Further, when apportioning fault, the court must consider the nature of each

---


43. Ch. 27, 1981 Wash. Laws 112 (codified at scattered sections of WASH. REV. CODE §§ 4.22, 7.72 (2000)).

44. WASH. REV. CODE § 7.72.010(4). The statute provides, among other things, that a manufacturer’s failure to implement a safeguard on a product is negligent where “at the time of manufacturer, the likelihood [P] that the product would cause the claimant’s harm or similar harms, and the seriousness [L] of those harms outweighed the burden [B] on the manufacturer” to implement the safeguard. Id. § 7.72.030. In other words, a manufacturer’s failure to implement a safeguard is negligent where B<PL. Thus, the statute defines negligence by using the formula devised by Learned Hand to discuss the incentives to exercise care created by negligence law. See United States v. Carroll Towing Co., 159 F.2d 169, 173 (2d Cir. 1947). This formula is also relevant to the incentives to exercise care created by the enhanced-injury doctrine. See infra Section IV.C.2.

45. WASH. REV. CODE § 4.22.005.

46. Id. § 4.22.015.

47. See id. §§ 7.72.010(4), 4.22.005–.015; Lundberg v. All-Pure Chem. Co., 55 Wash. App. 181, 186, 777 P.2d 15, 19 (1989) (“Thus, the Legislature has determined that the comparative fault doctrine shall apply to all actions based on ‘fault,’ including strict liability and product liability claims.”).

48. WASH. REV. CODE § 4.22.005.

49. For conduct to be a proximate cause of a plaintiff’s damages, it must be both: (1) a cause-in-fact and (2) a legal cause of the damages. See infra notes 78–84 and accompanying text.

50. WASH. REV. CODE § 4.22.015.
party’s conduct and the extent of the causal relation between that conduct and the resulting damages.\textsuperscript{51}

In 1986, the Washington State Legislature supplemented the 1981 Product Liability and Tort Reform Act with the 1986 Tort Reform Act.\textsuperscript{52} Previously in Washington, a plaintiff could recover, under joint and several liability, the full amount of her damages from any defendant who contributed to those damages.\textsuperscript{53} However, the 1986 Act abolished joint and several liability, except in cases where the plaintiff is not at fault.\textsuperscript{54} In making this change, the Act provided that, when apportioning responsibility for a claimant’s damages, the fault of every entity that “caused” the claimant’s damages shall be considered.\textsuperscript{55} The entities whose fault will be considered include, among others, “the claimant or person suffering personal injury or . . . property damage . . . .”\textsuperscript{56} Thus, by forcing courts to apportion percentages of fault among various parties, the legislature intended to hold each party accountable for only that portion of damages caused by its own fault.

B. \textit{Determining the Existence of Duty}

Enhanced-injury claims require the imposition of a negligence-based duty.\textsuperscript{57} The imposition of a duty turns on foreseeability—whether the risk of a harm occurring is foreseeable from the alleged tortfeasor’s perspective.\textsuperscript{58} A duty to reasonably minimize enhanced injuries is imposed because the risk of enhanced injuries is foreseeable.\textsuperscript{59} Therefore, foreseeability plays a central role in determining whether certain conduct constitutes a breach of an enhanced-injury duty and,

\textsuperscript{51} Id.
\textsuperscript{53} WASH. REV. CODE § 4.22.070.
\textsuperscript{54} Id.
\textsuperscript{55} Id.
\textsuperscript{56} Id. § 4.22.070(1).
\textsuperscript{57} See \textit{supra} notes 18–21 and accompanying text.
thus, whether that conduct may be deemed fault with respect to an enhanced injury. Only conduct that constitutes fault with respect to a particular harm can be used to apportion responsibility for that harm under Washington comparative fault law.60

The imposition of a negligence duty turns on the foreseeability of the consequences of an act. As Justice Cardozo stated in *Palsgraf v. Long Island R. Co.*,61 “[n]egligence, like risk, is . . . a term of relation.”62 A “bad act,” by itself, does not make a party liable for harm. Rather, a party can be liable for a particular harm only if, from the party’s perspective, the occurrence of that harm was a foreseeable consequence of the “bad act.”63 Hence, the party has no duty to prevent a particular harm and, thus, cannot be liable for the occurrence of that harm unless, from the party’s perspective, the harm was a foreseeable consequence of the act.64 Moreover, the exact manner in which the harm occurs need not be foreseeable as long as the occurrence of the harm itself was a foreseeable consequence of the act.65

The question of whether a party has a duty to reasonably minimize a particular harm is entirely separate from whether the party has another duty to reasonably minimize a different harm.66 Each distinct harm corresponds to a distinct potential duty.67 A particular harm imposes a duty to reasonably minimize the risk of that harm only if, from the party’s perspective, that harm is a foreseeable consequence of the party’s behavior.68 For example, our hypothetical plaintiff, Mary, has a duty, D1, to reasonably minimize the risk of the harm, H1, only if, from Mary’s perspective, the risk of H1 is foreseeable. Failing to reasonably minimize the risk of H1, for example by committing an act, A1, would constitute a

60. See WASH. REV. CODE §§ 4.22.005, .015, .070.
61. 162 N.E. 99 (N.Y. 1928). Washington courts have consistently adopted the reasoning of Justice Cardozo’s *Palsgraf* opinion. See *King*, 84 Wash. 2d at 248, 525 P.2d at 234; *Wells*, 77 Wash. 2d at 802–03, 467 P.2d at 294–95.
63. *Id.* (“Negligence in the abstract, apart from things related, is surely not a tort, if indeed it is understandable at all. . . . Negligence is not a tort unless it results in the commission of a wrong, and the commission of a wrong imports the violation of a right.”).
64. *Id.*
65. *King*, 84 Wash. 2d at 248, 525 P.2d at 234.
67. *Id.*
68. *Id.*
breach of D₁ and, therefore, make Mary liable for H₁. Further, if it is foreseeable from Mary’s perspective that another harm, H₂, could occur, then Mary has a separate and distinct duty, D₂, to reasonably minimize the risk of H₂. Failing to reasonably minimize the risk of H₂, for example by engaging in certain conduct, designated A₂, would constitute a breach of D₂ and, therefore, make Mary liable for H₂.

The fact that Mary has a duty, D₁, to reasonably minimize the risk of the harm, H₁, is irrelevant to the question of whether Mary has a duty, D₂, to reasonably minimize the risk of the harm, H₂. As Lord Simonds wrote in the famous *Wagon Mound* case, each duty “rests on its own bottom . . . .” The risk of each separate harm must be foreseeable from Mary’s perspective in order to impose each separate duty.

Correspondingly, the mere fact that Mary has breached a duty, D₁, for example by committing an act, A₁, is often irrelevant to whether Mary has breached a duty, D₂. A₁ can constitute a breach of D₂, thereby making Mary liable for H₂, only if it is foreseeable from Mary’s perspective that A₁ creates an unreasonable risk of H₂. Therefore, if it is foreseeable from Mary’s perspective that A₁ creates an unreasonable risk of H₁, but it is not foreseeable that A₁ creates an unreasonable risk of H₂, then committing A₁ would make Mary liable for H₁, but not for H₂.

C. Determining Proximate Cause

In Washington, an actor’s conduct must also constitute a proximate cause of a harm in order to make the actor liable for the harm. Proximate cause consists of two elements: cause-in-fact and legal causation. To satisfy the cause-in-fact element, a breach must produce the harm

---

69. Id.
70. Id.
71. Id.
72. Id.
73. Id.
74. Id.
75. Id.
76. Id.
77. Id.
complained of in a direct, unbroken sequence. This is also referred to as “but for” causation, as the harm would not have occurred “but for” the actor’s breach.

Legal causation involves a determination of whether, as a matter of policy, the connection between the defendant’s breach and the resultant harm is too remote or insubstantial to impose liability. Legal causation and duty are intertwined and virtually identical because some of the policy considerations analyzed in determining whether a duty was owed by a party are also analyzed when determining whether a party’s breach constituted a legal cause of the harm at issue. However, the questions of duty and legal cause are not identical. Rather, even where a harm is foreseeable and, thus, a duty is imposed, a party’s breach of that duty may not be a legal cause of the resulting harm if sound policy reasons require that the party not be liable for the harm.

In *Hartley v. Washington*, for example, the Supreme Court of Washington found a lack of legal causation in a case where a decedent was killed by a drunk driver. The drunk driver had previously been arrested numerous times for driving while intoxicated and, therefore, was subject to a revocation of his driver’s license by the State. The decedent’s estate alleged that the State’s failure to revoke the drunk driver’s license negligently caused the decedent’s death. The court held that the State’s failure to revoke the drunk driver’s license formed too attenuated a causal connection with—and therefore was not a legal cause of—the decedent’s death. The court reasoned that (1) revocation of the license would not have prevented the driver from driving, (2) a license

79. *Schooley*, 134 Wash. 2d at 478, 951 P.2d at 754; *Taggart*, 118 Wash. 2d at 226, 822 P.2d at 258; *Hartley*, 103 Wash. 2d at 778, 698 P.2d at 83.
80. *Schooley*, 134 Wash. 2d at 478, 951 P.2d at 754; *Taggart*, 118 Wash. 2d at 226, 822 P.2d at 258; *Hartley*, 103 Wash. 2d at 778, 698 P.2d at 83.
81. *Schooley*, 134 Wash. 2d at 478–79, 951 P.2d at 754; *Taggart*, 118 Wash. 2d at 226, 822 P.2d at 258; *Hartley*, 103 Wash. 2d at 779, 698 P.2d at 83.
82. *Schooley*, 134 Wash. 2d at 479, 951 P.2d at 755; *Taggart*, 118 Wash. 2d at 226, 822 P.2d at 258; *Hartley*, 103 Wash. 2d at 779–80, 698 P.2d at 83–84.
84. *Id.*
85. 103 Wash. 2d 768, 698 P.2d 77 (1985).
86. *Id.* at 770, 698 P.2d at 79.
87. *Id.*
88. *Id.* at 770, 698 P.2d at 78.
89. *Id.* at 784–85, 698 P.2d at 86.
Apportionment of Enhanced Injuries

does not grant authority to disobey the law, and (3) otherwise, the State would be open to unlimited liability.  

Schooley v. Pinch’s Deli Market, Inc. provides another example of the determination of legal causation. In that case, the Supreme Court of Washington held that a store’s illegal sale of alcohol could be a legal cause of an intoxicated minor’s injuries. In Schooley, a vendor sold alcohol to a minor who then furnished it to the plaintiff, who was also a minor. After consuming the alcohol and becoming intoxicated, the plaintiff dove into a partially empty swimming pool. The impact fractured her spinal cord, leaving her a quadriplegic. The vendor argued that the sale of alcohol was not a legal cause of the plaintiff’s injuries; it reasoned that extending the legal consequences of the initial sale of alcohol to the plaintiff’s injuries—which the vendor contended were remote and due to the subsequent transfer of the alcohol—would expose the vendor to unlimited liability. The court rejected the vendor’s argument, however, and held that the sale of alcohol could constitute a legal cause of the plaintiff’s injuries. The court reasoned that the vendor could prevent liability by refusing to sell alcohol to a minor or forcing a suspicious potential buyer to fill out and sign a certification card. Furthermore, the court noted that a minor who consumes alcohol could be found contributorily negligent, and the vendor’s liability would be limited by both foreseeability and the doctrine of superseding causes.

An intervening act can constitute a break in the chain of causation and, thus, constitute a superseding cause of a harm, relieving a party of liability for that harm. An intervening act is a superseding cause when

90. Id. at 785, 698 P.2d at 86–87.
92. Id. at 472, 951 P.2d at 751.
93. Id.
94. Id. at 474, 951 P.2d at 752.
95. Id. at 483, 951 P.2d at 757.
96. Id. at 481, 951 P.2d at 755–56.
97. Id.
it was not reasonably foreseeable. This typically means that an intervening act constitutes a superseding cause where the act (1) brings about a different type of harm than otherwise would have resulted from the actor’s conduct; or (2) operates independently of the situation created by the defendant’s conduct. In *McCoy v. American Suzuki Motor Corp.*, the plaintiff was struck by a vehicle while stopping to help a motorist whose car had overturned, allegedly due to a defect in the car. The plaintiff sued the car’s manufacturer, Suzuki, for his injuries. Suzuki argued that its negligence did not constitute a proximate cause of the plaintiff’s damages because it was unforeseeable that a rescuer would be injured by a third vehicle. The court held, however, that whether the plaintiff’s rescue attempt and the third driver’s negligence constituted superseding causes of the plaintiff’s harm was a question for the jury.

III. OTHER COURTS’ APPROACHES TO THE APPLICATION OF COMPARATIVE FAULT IN ENHANCED-INJURY CLAIMS

Some federal and state courts outside of Washington have addressed whether, under a scheme of comparative fault, primary fault should be compared with enhanced-injury fault when apportioning damages in enhanced-injury cases. A majority of courts favor comparing primary fault with enhanced-injury fault, thereby reducing the plaintiff’s recovery for enhanced injuries in proportion to the plaintiff’s primary fault. These courts favor such a comparison for various reasons: some courts fail to distinguish between primary injuries and the enhancement of injuries, others assume that fault for the primary accident is a proximate cause of enhanced injuries, and still other courts state that whether superseding causes should be analyzed under cause-in-fact or legal cause is beyond the scope of this Comment.


102. *Id.* at 358, 961 P.2d at 957.

103. *Id.*

104. As used in this Comment, the terms “compare” and “comparing” mean that the court uses both the primary fault and enhanced-injury fault to apportion responsibility for a plaintiff’s enhanced injuries.
Apportionment of Enhanced Injuries

primary fault should be compared with enhanced-injury fault. A minority of courts, however, use only enhanced-injury fault to apportion responsibility for enhanced injuries. The minority courts distinguish between primary injuries and injury enhancement, reasoning that only breach of an enhanced-injury duty may make a party responsible for enhanced injuries.

A. The Majority Approach: Reduction of a Plaintiff’s Enhanced Injury Recovery by the Plaintiff’s Primary Fault

1. Divisibility of Primary and Enhanced Injury Claims

Many of the majority courts fail to distinguish between injury enhancement and primary injuries when apportioning responsibility for damages. Instead, these courts treat the plaintiff’s enhanced injuries and primary injuries as inseparable. Therefore, the plaintiff’s primary fault is compared with the manufacturer’s enhanced-injury fault and responsibility is apportioned for all of the plaintiff’s injuries without differentiating between the primary injuries and the enhancement of the primary injuries. For example, in Trust Corp. of Montana v. Piper Aircraft Corp., the decedent negligently caused his plane to crash but allegedly died as a result of the defendant’s failure to provide a shoulder harness restraint system in the aircraft. The plaintiff argued that the primary-accident-causing factors and enhanced-injury-causing factors were qualitatively different and must be considered separately. Therefore, according to the plaintiff, for purposes of apportionment of the enhanced injuries, the court should exclude evidence of the cause of the crash and focus solely on evidence of the plaintiff’s injury enhancement. However, the court ruled that, under Montana law, the plaintiff’s contributory negligence in causing the initial accident should be compared with the defendant’s failure to provide a shoulder harness

107 Id. at 1094.
108 Id.
restraint system. The court stated that, while the view opposing such a comparison has some merit, courts should consider all of the factors that contributed to the event that caused the plaintiff’s injuries. Consequently, the court declined to draw a clear line between the primary accident and the injury enhancement.

Other courts have offered additional reasons for refusing to analyze injury enhancement as separate and distinct from primary injuries. One court has stated that refusing to compare primary fault with enhanced-injury fault would make it too difficult to instruct the jury on apportionment. Furthermore, some commentators have argued that, in practice, it is very difficult to distinguish between collision-causing and injury-causing fault. Under the laws of physics, the enhancement of injuries correlates to the severity of the collision—the more severe the collision, the more enhanced the injuries. For example, not all injury enhancements can be as clearly defined as burns resulting from an explosion. It is possible that injury enhancement will merely increase the severity of the injuries incurred in the primary accident. Hence, these commentators argue, any conduct influencing the severity of a collision is a proximate cause of enhanced injury and should be compared with a manufacturer’s fault.

2. **Primary Fault as a Proximate Cause of Enhanced Injuries**

Some courts assume that fault for the primary accident is a proximate cause of enhanced injuries. Consequently, these courts compare primary fault with enhanced-injury fault when apportioning responsibility for enhanced injuries. For example, in *Meekins v. Ford Motor Co.*, the driver of an automobile was involved in an accident in which the

---

109. Id. at 1098.

110. Id.


113. Id.

114. Id. at 439–40.


Apportionment of Enhanced Injuries

The driver sued Ford, alleging that a defect in the air bag crushed the fingers in his left hand against the steering wheel when the airbag deployed. Ford raised the defense of comparative negligence, claiming that the cause of the driver’s injuries was his failure to stop at a stop sign. The driver countered that, even if he had caused the accident, his negligence should not be compared with Ford’s negligence in defectively designing the airbag that caused his fingers to be crushed. However, the Meekins court ruled that, under comparative fault principles, the driver’s negligence in causing the initial accident should be compared with Ford’s negligence in defectively designing the airbag. The court reasoned that “[i]t is obvious that the negligence of a plaintiff who causes the initial collision is one of the proximate causes of all of the injuries he sustained,” regardless of whether the injuries stem from the initial collision or a product defect.

Other courts reduce plaintiff recovery by analogizing to the doctrine of subsequent tortfeasors. Under this doctrine, negligence by an original tortfeasor exposes that tortfeasor to liability for increased harm caused by the subsequent negligence of other tortfeasors. For example, a party who is initially at fault for causing an accident can be held liable for additional injuries incurred by the victim through subsequent negligent medical treatment. Courts that rely on this reasoning state that it is just as foreseeable to a primary accident-tortfeasor that equipment in a car may be defective as it is that a doctor may negligently treat the plaintiff’s injuries. Because the primary tortfeasor is liable in the latter case for the additional injuries caused by the negligent medical treatment, these courts reason that the plaintiff in an enhanced-injury case should be held at least partially responsible for causing the enhanced injuries.

117. Id. at 340.
118. Id.
119. Id.
120. Id.
121. Id. at 346.
122. Id.
124. KEETON ET AL., supra note 16, § 44, at 301.
125. KEETON ET AL., supra note 16, § 44, at 310.
126. See, e.g., Farnsworth, 965 P.2d at 1217–18; Moore, 596 So. 2d at 238.
127. See, e.g., Farnsworth, 965 P.2d at 1217–18; Moore, 596 So. 2d at 238.
3. **Policy Concerns**

Other courts conclude, with little analysis, that primary fault should be compared with enhanced-injury fault. Some of these courts state that comparative fault applies to strict liability, product liability, or enhanced-injury claims. Then, with little analysis of the manner in which comparative fault should operate in enhanced-injury claims, or the types of “fault” that should be compared in an enhanced-injury case, these courts compare primary fault with enhanced-injury fault. For example, in *Kidron, Inc. v. Carmona*, the driver of a pickup truck was killed when he negligently drove into the rear of a delivery truck. The force of the impact shoved the smaller pickup truck under the delivery truck’s rear assembly, which cut through the passenger compartment of the pickup, killing the driver. The decedent’s estate brought suit against the manufacturer of the delivery truck, alleging negligence and strict liability in assembling the truck without a rear under-ride guard, which allegedly would have prevented the pickup truck from being forced underneath the delivery truck’s bed during the collision. Kidron raised the defense of comparative negligence, arguing that the decedent’s fault in causing the initial accident should reduce any recovery for the enhanced injuries that caused his death. The trial court denied Kidron’s defense, but the Florida Court of Appeals reversed, holding that the manufacturer’s defense should have gone to the jury. The Court of Appeals characterized the issue as “whether the rules of comparative negligence should apply in a claim for strict liability in the context of [an enhanced-injury claim] . . . .” The court then concluded “that principles of comparative negligence should be

---


129. *Zuern*, 937 P.2d at 681–82; *Daly*, 575 P.2d at 1175; *Kidron*, 665 So. 2d at 292–93; *Day*, 345 N.W.2d at 358; *Whitehead*, 897 S.W.2d at 694.


131. *Id.* at 290.

132. *Id.*

133. *Id.*

134. *Id.*

135. *Id.*

136. *Id.* at 291.

137. *Id.* at 292.
Apportionment of Enhanced Injuries

applied in the same manner in a strict liability suit, regardless of whether the injury at issue has resulted from the primary or secondary collision.138

Other courts state that fairness and good reason mandate a comparison of primary fault with enhanced-injury fault.139 In addition, some commentators have argued that public policy dictates comparing primary fault with enhanced-injury fault in automobile cases in order to deter driver misconduct.140 Otherwise, they contend, drivers will have insufficient incentives to abstain from negligent driving.141

B. The Minority Approach: Distinguishing Between the Primary Injuries and the Enhanced Injuries

A minority of courts do not compare primary fault with enhanced-injury fault when apportioning responsibility for enhanced injuries.142 These courts focus solely on the enhancement of injuries because the enhanced-injury doctrine presupposes the occurrence of primary accidents, regardless of their cause.143 They note that an enhanced-injury duty requires reasonable steps to prevent only the enhancement of injuries in such accidents.144 Furthermore, because an enhanced-injury duty focuses solely on enhancement of injuries, these courts distinguish between proximate cause of the primary accident and proximate cause of enhanced injuries.145 They further state that only breach of a duty to

138. Id.
139. See, e.g., Montag v. Honda Motor Co., 75 F.3d 1414, 1419 (10th Cir. 1996); Trust Corp. of Mont. v. Piper Aircraft Corp., 506 F. Supp. 1093, 1098 (D. Mont. 1981). Some commentators have argued that refusing to compare all of a plaintiff’s conduct places “extraordinary hardships” on manufacturers who are singled out for “discriminatory application” of proximate cause and comparative fault principles. Vickles & Oldham, supra note 112, at 439.
140. Vickles & Oldham, supra note 112, at 440.
141. Vickles & Oldham, supra note 112, at 440.
143. See, e.g., Katsugeris, 973 F.2d at 1344–45; Jimenez, 74 F. Supp. 2d at 566; Cota, 684 P.2d at 894–95; Reed, 494 N.W.2d at 230, Andrews, 796 P.2d at 1095–96.
144. See, e.g., Katsugeris, 973 F.2d at 1344–45; Jimenez, 74 F. Supp. 2d at 566; Cota, 684 P.2d at 895–96; Reed, 494 N.W.2d at 230, Andrews, 796 P.2d at 1095.
145. See, e.g., Katsugeris, 973 F.2d at 1344–45; Jimenez, 74 F. Supp. 2d at 566; Cota, 684 P.2d at 895–96; Reed, 494 N.W.2d at 230, Andrews, 796 P.2d at 1095.
prevent enhanced injuries can constitute a proximate cause of such injuries. Moreover, they reason that the cause of the primary accident, while relevant to the cause of the primary injuries, is legally irrelevant to the cause of the enhanced injuries.

The Iowa Supreme Court’s decision in Reed v. Chrysler Corp. provides an example of the minority approach. In Reed, the plaintiff was a passenger in a Jeep whose driver, intoxicated and speeding, negligently drove off the road. Once off the road, the Jeep rolled onto its fiberglass top, breaking it, and continued to slide 300 feet in an upside-down position. During the vehicle’s slide, the plaintiff’s arm became pinched between the Jeep’s roll bar and the highway, causing severe fractures. The plaintiff alleged that Chrysler negligently designed its hardtop with fiberglass instead of steel, and that this defect caused his arm injury.

Chrysler sought to introduce evidence that the plaintiff and driver were intoxicated prior to the primary collision. The Iowa Supreme Court held, however, that evidence of the driver’s and plaintiff’s intoxication was irrelevant to liability for the plaintiff’s enhanced injuries. The court reasoned that the enhanced-injury theory presupposes the occurrence of primary collisions, whatever their cause. The court should focus solely on the enhancement of the plaintiff’s injuries and, for that reason, ought to concentrate only on the manufacturer’s fault in enhancing the plaintiff’s injuries. Any part that the plaintiff played in causing the primary collision was irrelevant.

Another court has held that primary collision fault should not be compared with enhanced-injury fault, not only because primary

146. See, e.g., Katsugeras, 973 F.2d at 1344–45; Jimenez, 74 F. Supp. 2d at 566; Cota, 684 P.2d at 895–96; Reed, 494 N.W.2d at 230; Andrews, 796 P.2d at 1095–96.

147. See, e.g., Katsugeras, 973 F.2d at 1344–45; Jimenez, 74 F. Supp. 2d at 566; Cota, 684 P.2d at 895–96; Reed, 494 N.W.2d at 230; Andrews, 796 P.2d at 1095–96.


149. Id. at 225–26.

150. Id. at 226.

151. Id.

152. Id. at 227–28.


154. Id. at 230.

155. Id.

156. Id.

157. Id.
Apportionment of Enhanced Injuries

collision fault is legally irrelevant to the cause of enhanced injuries, but also because the very concept of enhanced injury fairly apports fault and damages on a comparative basis. The defendant-manufacturer is liable only for that portion of the plaintiff’s injuries proximately caused by its breach—the enhanced injuries. The plaintiff is responsible only for that portion of her injuries proximately caused by her breach—the primary injuries.

IV. WASHINGTON COURTS SHOULD NOT USE A PLAINTIFF’S PRIMARY FAULT TO REDUCE THE PLAINTIFF’S RECOVERY FOR ENHANCED INJURIES

Washington courts should not use a plaintiff’s primary fault to reduce a plaintiff’s recovery for enhanced injuries. Under the enhanced-injury doctrine, the primary accident and injury enhancement are two separate torts, each consisting of its own distinct negligence elements. Therefore, the mere fact that a plaintiff has breached a primary duty—and is thus guilty of primary fault—does not mean that the plaintiff has breached an enhanced-injury duty—and, thus, is not necessarily guilty of enhanced injury fault. Furthermore, under Washington’s scheme of comparative fault and law of proximate cause, a court cannot reduce a plaintiff’s recovery for enhanced injuries if the plaintiff is guilty of only primary fault (and not enhanced injury fault). Finally, using primary fault to reduce a plaintiff’s recovery for enhanced injuries will result in decreased product safety, windfalls for manufacturers, and punitive consequences for plaintiffs.

A. Under the Enhanced-Injury Doctrine, Fault for Enhanced Injuries Is Separate and Distinct from Fault for Primary Injuries

Under the enhanced-injury doctrine, the primary accident and injury enhancement are two distinct torts. An enhanced-injury duty is a duty

159. Id.
160. Id.
161. See Baumgardner v. Am. Motors Corp., 83 Wash. 2d 751, 756–58, 522 P.2d 829, 832–33; see also supra notes 15–25 and accompanying text; supra Section II.B. Enhanced injuries are defined both by the type of harm and by the severity of the harm. Hence, an enhanced injury may be an additional type of harm that occurs—for example burns are a different type of harm than bumps and bruises. However, the enhanced injury may also be an increase in the severity of the same type.
to reasonably prevent only the enhancement of primary injuries; a separate duty governs the prevention of un-enhanced primary injuries.\textsuperscript{162} Moreover, an enhanced injury is, by definition, separate and distinct from a primary injury; it is an enhancement of a primary injury due to some conduct—for example, production of a defective product—that did not cause the primary accident but enhanced the injury suffered in the primary accident.\textsuperscript{163} Proof of injury enhancement is essential for the survival of an enhanced-injury claim; before addressing the question of apportioning fault for an enhanced injury, the trier-of-fact must first find an injury enhancement that is distinct and separate from the primary injuries.\textsuperscript{164} Therefore, courts must analyze the negligence elements of the injury enhancement separately from the elements of the primary accident.

Some courts refuse to separate the primary accident from the injury enhancement.\textsuperscript{165} For example, in \textit{Trust Corp. of Montana v. Piper Aircraft Corp.},\textsuperscript{166} the decedent negligently caused his plane to crash but allegedly died as a result of the defendant’s failure to provide a shoulder harness restraint system.\textsuperscript{167} The court ruled that the plaintiff’s contributory negligence in causing the initial accident should be compared with the defendant’s failure to provide a shoulder harness of harm—for example, severe whiplash, bumps, and bruises are different than minor whiplash, bumps, and bruises.

\textsuperscript{162} See Baumgardner, 83 Wash. 2d at 756–58, 522 P.2d at 832–33; see also supra notes 28–29 and accompanying text; supra Section II.B.

\textsuperscript{163} See Baumgardner, 83 Wash. 2d at 752, 522 P.2d 829 at 830 (stating that a court must decide "whether the manufacturer of an automobile involved in a collision is liable for injuries caused or enhanced because of a defect . . . even though the defect did not cause or contribute to the collision itself."); see also Harris, supra note 2, at 649 ("The [enhanced injury] theory should be applied to any situation in which an object or conduct does not cause contact, but wrongfully causes the damage from the contact to be greater than it would have been had a deficiency in the object or conduct not existed.").

\textsuperscript{164} See Baumgardner, 83 Wash. 2d at 758, 522 P.2d at 833 ("[T]he plaintiff has the usual burdens of proof as in any negligence action including proof of the nature and extent of the injuries proximately caused or enhanced by the defect."); see also KEETON ET AL., supra note 16, § 30, at 165 ("[P]roof of damage [is] an essential part of the plaintiff’s case.").


\textsuperscript{167} Id. at 1094.
restraint system because courts should consider all of the factors that contributed to the event that caused the plaintiff’s injuries.\textsuperscript{168}

Other courts consider the primary accident and the injury enhancement as two separate events and, therefore, two separate torts.\textsuperscript{169} For example, in Reed v. Chrysler Corp.,\textsuperscript{170} an intoxicated plaintiff negligently drove his Jeep off the road, causing a rollover accident.\textsuperscript{171} Due to an allegedly defective top, the plaintiff’s arm became pinched between the Jeep’s roll bar and the highway, causing severe fractures.\textsuperscript{172} The Iowa Supreme Court held that the initial rollover accident was a separate tort from the arm fractures, reasoning that the enhanced-injury theory presupposes the occurrence of primary accidents, whatever their cause, and therefore focuses solely on the enhancement of injuries.\textsuperscript{173} Unlike the Piper Aircraft court, the Reed court considered the injury enhancement separately from the primary accident and, consequently, considered the elements of fault for the primary accident separately from the elements of fault for the injury enhancement. The Reed approach is consistent with the enhanced-injury doctrine.

Courts that treat the primary accident and injury enhancement as a single, indivisible event gloss over the analytical distinctions between primary accidents and injury enhancement required by the enhanced-injury doctrine. Because only injury enhancement is at issue in an enhanced-injury claim, only enhanced-injury fault is relevant when apportioning fault for enhanced injuries. Many courts that fail to make the distinction between the primary accident and injury enhancement may actually be faced with cases that should be dealt with at the damages stage, prior to the issue of apportionment. For example, in Piper Aircraft, it may not have been possible to distinguish between the injuries caused by the plane crash and the allegedly separate injuries that resulted from the defective restraint system.\textsuperscript{174} Therefore, there was

\begin{itemize}
  \item \textsuperscript{168} Id. at 1098.
  \item \textsuperscript{170} 494 N.W.2d 224 (Iowa 1993).
  \item \textsuperscript{171} Id. at 225–26.
  \item \textsuperscript{172} Id. at 226.
  \item \textsuperscript{173} Id. at 230.
\end{itemize}
likely no separate injury enhancement and, consequently, the issue of apportionment would not have arisen.

Once it is established that there is a distinction between primary injuries and injury enhancement, a court must consider the elements of fault for each harm separately. In negligence law generally, each distinct harm corresponds to an equally distinct duty to reasonably minimize that harm.\textsuperscript{175} A party’s duty to prevent one harm does not necessarily mean that the party has a duty to prevent another harm.\textsuperscript{176} Therefore, the duty to prevent an injury enhancement, which by definition is a separate harm from a primary injury, is distinct from the duty to prevent a primary injury. In an enhanced-injury claim, a court must determine which parties had a duty to prevent injury enhancement, which is necessarily a separate inquiry from whether those same parties had a duty to prevent primary injuries.\textsuperscript{177}

Moreover, breach of a duty to prevent enhanced injury is not necessarily the same as breach of a duty to prevent primary injuries.\textsuperscript{178} Conduct constituting breach of an enhanced injury operates independently from conduct constituting breach of a primary duty and brings about a different type of harm.\textsuperscript{179} Breach of the primary duty, for example Mary’s failure to stop at the stop sign, leads to primary injuries—the bumps and bruises. Breach of the enhanced-injury duty—the gasoline tank defect—brings about an enhancement of the primary injuries—the severe burns. Therefore, a party’s conduct must be compared to that party’s distinct duty to determine whether the party has breached that duty.

\textsuperscript{175} See Overseas Tankship (U.K.) Ltd. v. Motts Dock & Eng’g Co. Ltd. (the “Wagon Mound #1”), [1961] App. Cas. 388, 426 (P.C. 1961); see also supra notes 66–71 and accompanying text.

\textsuperscript{176} See Overseas Tankship, [1961] App. Cas. at 426; see also supra notes 72–74 and accompanying text.


\textsuperscript{178} See Overseas Tankship, [1961] App. Cas. at 426; see also supra notes 75–77 and accompanying text.

\textsuperscript{179} See Overseas Tankship, [1961] App. Cas. at 426; see also supra notes 75–77 and accompanying text.
Apportionment of Enhanced Injuries

B. Under Washington’s Comparative Fault Scheme, a Court Cannot Hold a Plaintiff Responsible for Enhanced Injuries if the Plaintiff Has Breached Only a Duty To Prevent the Primary Accident

In order to show that the plaintiff was also partially at fault for the enhanced injuries, the defendant must show that the plaintiff had a duty to avoid injury enhancement, the plaintiff breached that duty, and the plaintiff’s act was a proximate cause of the enhanced injury. Frequently, a plaintiff will not have a duty to prevent injury enhancement because the enhanced injuries are not foreseeable from the plaintiff’s perspective. Furthermore, conduct that constitutes breach of a primary duty cannot constitute breach of an enhanced-injury duty. Moreover, a plaintiff’s primary fault is not a proximate cause of enhanced injuries under Washington law. Washington’s statutory comparative fault scheme supports this analysis. Finally, plaintiffs can be apportioned responsibility for enhanced injuries only if they are guilty of enhanced-injury fault. Therefore, Washington courts should not reduce a plaintiff’s recovery for enhanced injuries based on primary accident fault.

180. An enhanced-injury claim consists of the basic elements of negligence. Baumgardner, 83 Wash. 2d at 758, 522 P.2d at 833. The elements of negligence are duty, breach, causation, and damages. Degel v. Majestic Mobile Manor, Inc., 129 Wash. 2d 43, 48, 914 P.2d 728, 731 (1996); KEETON ET AL., supra note 16, § 30, at 164–65. A finding of comparative fault involves the same considerations as a finding of negligence. See WASH. REV. CODE § 4.22.015 (2000) (“Fault includes acts or omissions . . . that are in any measure negligent or reckless toward the person or property of the actor or others . . . .”); see also Geschwind v. Flanagan, 121 Wash. 2d 833, 838, 854 P.2d 1061, 1064 (1993) (quoting Seattle-First Nat’l Bank v. Shoreline Concrete Co., 91 Wash. 2d 230, 238, 588 P.2d 1308, 1314 (1978) (“A plaintiff’s negligence relates to a failure to use due care for his own protection . . . .”)); KEETON ET AL., supra note 16, § 65, at 451 (“Contributory negligence is conduct on the part of the plaintiff, contributing as a legal cause to the harm he has suffered, which falls below the standard to which he is required to conform for his own protection.”). Therefore, to find an enhanced injury plaintiff guilty of comparative fault, a court must find (1) that the plaintiff had a duty to herself to reasonably minimize a foreseeable risk of enhanced injury, (2) that the plaintiff breached her duty, and (3) that the plaintiff’s breach proximately caused (4) an enhanced injury.

181. See infra Part IV.B.1.
182. See infra Part IV.B.2.
183. See infra Part IV.B.2.
184. See infra Part IV.B.4.
185. See infra Part IV.B.4.
1. A Plaintiff’s Breach of a Primary Duty Cannot Constitute Breach of a Duty To Avoid Injury Enhancement

To find that a plaintiff has breached an enhanced-injury duty, a court must first determine that the plaintiff had an enhanced-injury duty. Duty depends on foreseeability of the harm. To impose a duty on a plaintiff to avoid enhanced injuries, it must be foreseeable from the plaintiff’s perspective that her conduct raises an unreasonable risk of an enhanced injury. Therefore, a plaintiff cannot have a duty to avoid injury enhancement unless she can foresee the injury enhancement.

Frequently, an enhanced injury is simply not foreseeable to a plaintiff and, therefore, the plaintiff will not have a duty to prevent it. For example, it is disingenuous to claim that Mary, our hypothetical plaintiff, should foresee the risk of disabling burns if she runs a stop sign at fifteen miles per hour. Certainly, her conduct raises a foreseeable risk of primary injuries—bumps and bruises. However, Mary can have a duty to prevent the burns only if the risk of burns is foreseeable to her. Requiring Mary to foresee the burns would require her to foresee the hidden defect in the gasoline tank; this is inconsistent with imposing an enhanced injury duty on the manufacturer: the manufacturer is liable for enhanced injuries resulting from such defects, but is relieved from such liability due to Mary’s failure to foresee the manufacturer’s breach.

Some commentators have expressed concern that the severity of the primary collision is intrinsically linked to the severity of the resulting enhanced injuries. However, this concern pertains to the scope of an enhanced-injury duty, not to the apportionment of enhanced injuries between parties. An enhanced-injury duty necessarily recognizes the foreseeability of not just the occurrence of primary accidents, but also

186. Baumgardner, 83 Wash. 2d at 756–57, 522 P.2d at 832 (stating that primary determination with enhanced injury doctrine is whether a duty was owed).


188. Baumgardner, 83 Wash. 2d at 757, 522 P.2d at 832 (stating that key in determining whether an enhanced injury duty is owed is whether risk of enhanced injury is reasonably foreseeable); see also King, 84 Wash. 2d at 248, 525 P.2d at 234; Wells, 77 Wash. 2d at 802–03, 467 P.2d at 294–95; Palsgraf v. Long Island R. Co., 162 N.E. 99, 101 (N.Y. 1928).

189. See Baumgardner, 83 Wash. 2d at 757, 522 P.2d at 832 (stating that imposition of an enhanced injury duty turns on foreseeability of the risk of injury enhancement).

190. See Baumgardner, 83 Wash. 2d at 757–58, 522 P.2d at 832.

the severity of such accidents. Therefore, an enhanced-injury duty requires reasonable minimization of the risk of injury enhancement in accidents of a foreseeable severity. For example, it is foreseeable that Mary will cause a side impact collision in which the vehicles are traveling anywhere from one to Y miles per hour. The fact that a primary accident at Y miles per hour will result in greater injury enhancement than a primary accident at Y–1 miles per hour does not change the fact that the Y miles per hour accident was foreseeable and, thus, the manufacturer had a duty to reasonably minimize the risk of injury enhancement in that accident. In addition, the manufacturer need only take reasonable steps to prevent injury enhancement. Hence, if the manufacturer takes reasonable steps, but the plaintiff causes a primary accident at Y+1 miles per hour and, thus, suffers injury enhancement anyway, the manufacturer is not liable for the enhanced injuries because they could not have been prevented by reasonable precautions.

Because the primary and enhanced-injury duties are distinct and separate, a plaintiff’s conduct that constitutes breach of a primary duty does not necessarily constitute breach of an enhanced-injury duty. Conduct constitutes breach of an enhanced-injury duty when it is foreseeable that the conduct raises the risk of the injury enhancement. Therefore, conduct constituting breach of a primary duty could also constitute breach of an enhanced injury duty only if it is foreseeable that...

---

192. This proposition is implicit in the concept of foreseeing circumstances that give rise to a risk of injury enhancement. A side-impact automobile collision at thirty miles per hour might raise a foreseeable risk of injury enhancement, while the same collision at two miles per hour might not. See, e.g., Larsen v. Gen. Motors Corp., 391 F.2d 495, 502-03 (8th Cir. 1968):

The intended use and purpose of an automobile is to travel on the streets and highways, which travel more often than not in close proximity to other vehicles and at speeds that carry the possibility, probability, and potential of injury-producing impacts. Therefore, the manufacturer should... be held to a reasonable duty of care... to minimize the effect of accidents.

193. See id.

194. See Baumgardner, 83 Wash. 2d at 756, 522 P.2d at 832 (stating that, as with negligence defendants generally, “a manufacturer is not expected to produce an accident-free product, it is not an insurer of the users of its product and it need not adopt every possible safety device”).

195. See id. At 756-57, 522 P.2d at 823-33. In this hypothetical example, the manufacturer could also seemingly avoid liability because the Y+1 miles per hour collision was not foreseeable.

196. See id. at 756–58, 522 P.2d at 832–33; see also supra notes 161–164 and accompanying text.

the conduct in question also raises the risk of injury enhancement. Many courts that compare primary fault with enhanced-injury fault skip this foreseeable analysis. Without an additional duty to prevent enhanced injuries, a plaintiff cannot breach a duty to prevent enhanced injuries. 198

Even if an enhanced injury is foreseeable to a plaintiff, however, conduct constituting breach of a duty to prevent primary injuries, by definition, cannot constitute breach of a duty to prevent enhancement of primary injuries because it is the breach of a primary duty that gives rise to the imposition of an enhanced-injury duty. An enhanced-injury duty presupposes the occurrence of primary accidents; an enhanced-injury duty requires reasonable minimization of the risk of injury enhancement because it is foreseeable that a primary accident will occur, regardless of its cause. Therefore, the risk of enhanced injuries—the only harm at issue in an enhanced-injury claim—cannot arise until a primary accident and primary injuries have already occurred. 199 Given that primary injuries may occur, a duty to prevent injury enhancement requires additional conduct to prevent enhancement of the primary injuries. Therefore, in order for there to be a breach of an enhanced-injury duty, there must be some conduct other than that constituting breach of a primary duty.

2. Plaintiff’s Primary Fault Cannot Be a Proximate Cause of Injury Enhancement

The statutory changes made to Washington’s comparative fault scheme support the separation of primary fault from enhanced-injury fault required under the enhanced-injury doctrine. Washington’s comparative fault statutes indicate that a plaintiff is responsible only for injuries that she actually caused. 200 The statutes make no reference to the enhanced-injury doctrine. 201 However, the 1981 Tort Reform Act provides that a claimant’s contributory fault proportionately diminishes a claimant’s recovery of damages only “for an injury attributable to the claimant’s contributory fault.” 202 Furthermore, the 1981 Tort Reform

---

198 See Baumgardner, 83 Wash. 2d at 758, 522 P.2d at 833; Overseas Tankship, [1961] A.C. at 426; see also supra notes 29, 57–60, 66–77 and accompanying text.

199 See Baumgardner, 83 Wash. 2d at 757, 522 P.2d 833; see also supra notes 15–25 and accompanying text.


201 See supra Part II.A.

202 WASH. REV. CODE § 4.22.005 (emphasis added).
Apportionment of Enhanced Injuries

Act adds that, when apportioning fault under Washington’s comparative fault scheme, a court must consider the extent of the causal relation between each party’s conduct and the resulting damages. Finally, under the 1986 Tort Reform Act, when apportioning responsibility for a plaintiff’s damages, only the fault of entities that “caused” the plaintiff’s damages is relevant when apportioning responsibility for the damages. By definition, primary fault is not “fault” with respect to enhanced injuries because primary fault constitutes the breach of a duty to prevent primary injuries, not the breach of a duty to prevent enhanced injuries. Therefore, under Washington’s comparative fault scheme, a plaintiff’s primary fault cannot be considered during apportionment of enhanced injuries; only breach of an enhanced-injury duty—enhanced-injury fault—may be considered during enhanced-injury apportionment.

Some courts have ruled that a plaintiff’s primary fault is a proximate cause of enhanced injuries, thereby reducing a plaintiff’s enhanced-injury recovery. Many of these courts assume that primary fault is a proximate cause of enhanced injuries. Other courts analogize to the doctrine of subsequent tortfeasors. These courts reason that it is just as foreseeable to a tortfeasor who causes a primary automobile accident that equipment in a car may be defective as it is that a doctor may negligently treat the plaintiff’s injuries, thereby causing additional injuries. Because the primary tortfeasor is liable in the latter case for the additional injuries caused by negligent medical treatment, these courts reason that the plaintiff in an enhanced-injury claim should be held at least partially responsible for causing the enhanced injuries.

However, even if a plaintiff has a duty to prevent injury enhancement, primary fault is not a proximate cause of enhanced injuries under Washington law because enhanced-injury fault constitutes a superseding

203. Id. § 4.22.015.
204. Id. § 4.22.070(1).
205. See id. §§ 4.22.005, .015, .070; see also supra Section II.B.
207. See, e.g., Hinkamp, 735 F. Supp. at 178; Zuern, 937 P.2d at 681–82; Meekins, 699 A.2d at 246; Kidron, 665 So. 2d at 292; Whitehead, 897 S.W.2d at 694.
209. See, e.g., Farnsworth, 965 P.2d at 1218; Moore, 596 So. 2d at 238.
210. See, e.g., Farnsworth, 965 P.2d at 1218; Moore, 596 So. 2d at 238.
cause of enhanced injuries.\textsuperscript{211} An intervening act is a superseding cause when the act is not reasonably foreseeable.\textsuperscript{212} Typically, an act is a superseding cause when (1) the act brings about a different type of harm than otherwise would have resulted from the actor’s conduct; or (2) the act operates independently of the situation created by the defendant’s conduct.\textsuperscript{213} By definition, enhanced injuries are distinct from primary injuries; enhanced injuries are an enhancement of primary injuries.\textsuperscript{214} Furthermore, breach of an enhanced-injury duty operates independently of the cause of a primary accident; an enhanced-injury duty presupposes the occurrence of a primary accident.

Even if a breach of an enhanced-injury duty was not a superseding cause of an enhanced injury, primary fault should not constitute a legal cause of an enhanced injury under Washington law. In Washington, legal causation involves a determination of whether, as a matter of policy, the connection between the defendant’s breach and the resultant harm is too remote or insubstantial to impose liability.\textsuperscript{215} This question is similar, but not identical, to the question of duty.\textsuperscript{216} By imposing an enhanced-injury duty, the Supreme Court of Washington has said that plaintiffs have the right to be free from unreasonable risks of enhanced injuries in primary accidents.\textsuperscript{217} Because plaintiffs maintain this right, and because it is foreseeable that primary accidents will occur, manufacturers must take reasonable care to prevent further injuries.\textsuperscript{218} Therefore, to penalize the plaintiff for the manufacturer’s failure to reasonably minimize a foreseeable risk of injury enhancement is


\textsuperscript{212} McCoy, 136 Wash. 2d at 358, 961 P.2d at 957; Ballard, 91 Wash. App. at 758–59, 959 P.2d at 1127; Cramer, 73 Wash. App. at 520–21, 870 P.2d at 1001; Anderson, 48 Wash. App. at 442–43, 739 P.2d at 1184.

\textsuperscript{213} Campbell v. ITE Imperial Corp., 107 Wash. 2d 807, 813, 733 P.2d 969, 973 (1987); Anderson, 48 Wash. App. at 444, 739 P.2d at 1185.

\textsuperscript{214} See Baumgardner v. Am. Motors Corp., 83 Wash. 2d 751, 752, 522 P.2d 829, 830; Harris, supra note 2, at 649; see also supra note 163 and accompanying text.


\textsuperscript{216} Schooley, 134 Wash. 2d at 479, 951 P.2d at 755; see also supra notes 82–84 and accompanying text.

\textsuperscript{217} See Baumgardner, 83 Wash. 2d at 758, 822 P.2d at 833.

\textsuperscript{218} Id.
Apportionment of Enhanced Injuries

inconsistent with imposing an enhanced-injury duty in the first place. In addition, because it is foreseeable that primary accidents will inevitably occur, the manufacturer is almost always in the better position to prevent enhancement of the primary injuries sustained in primary accidents. Therefore, the manufacturer should be responsible for minimizing the risk of injury enhancement in such accidents.

3. Courts That Use Primary Fault To Reduce Recovery for Enhanced Injuries Often Mischaracterize the Relevance of Primary Fault During Apportionment of Enhanced Injuries

Courts that compare primary and enhanced-injury fault under apportionment for enhanced injuries mischaracterize the primary issue presented when applying comparative fault to enhanced-injury claims: what conduct may be considered fault with respect to enhanced injuries. Some courts incorrectly frame the issue as whether comparative fault should apply to product liability, strict liability, or enhanced-injury claims. In Washington, comparative fault is the law and, therefore, should apply to enhanced-injury claims. The real issue is the relevance of primary fault when comparative fault is applied to an enhanced-injury claim. Put another way, the issue is whether primary fault may be considered “fault” with respect to enhanced injuries. Prior to reducing a plaintiff’s recovery for enhanced injuries, a court must, at minimum, address this question. Consequently, the court must address the elements required to prove “fault”—duty, breach, causation and damages—with respect to the enhanced injuries.

219. In some cases, a plaintiff is also in a good position to reduce the risk of injury enhancement. For example, a motorcyclist can reduce the risk of enhancement of injuries sustained in a motorcycle accident by wearing a helmet. However, product users are typically unable to prevent injury enhancement once primary accidents occur. For example, Mary could do little about the defect in the gasoline tank that caused severe burns.


221. Baumgardner, 83 Wash. 2d at 758, 522 P.2d at 833 (stating that elements of an enhanced injury claim are the negligence elements of duty, breach, causation, and damages); Geschwind v. Flanagan, 121 Wash. 2d 833, 838, 854 P.2d 1061, 1064 (1993) (noting that a plaintiff’s negligence is failure to exercise due care toward herself); KEETON ET AL., supra note 16, § 65, at 451 (“Contributory negligence is conduct on the part of the plaintiff, contributing as a legal cause to the harm he has suffered, which falls below the standard to which he is required to conform for his own
4. **Plaintiffs Can Be Apportioned Responsibility for Enhanced Injuries, but Only If Guilty of Enhanced-Injury Fault**

At least one commentator has argued that a plaintiff’s enhanced-injury recovery should be reduced by the plaintiff’s primary fault based on the erroneous notion that, otherwise, plaintiffs are not deterred from misconduct.\(^{222}\) However, the fact that a plaintiff’s primary fault cannot be used to reduce a plaintiff’s recovery for enhanced injuries does not mean that a plaintiff can never be apportioned responsibility for her enhanced injuries. If a primary accident is foreseeable to a plaintiff, and it is foreseeable that the accident will give rise to a risk of injury enhancement, then the plaintiff must take reasonable steps to minimize the risk of injury enhancement in the event that the primary accident occurs.\(^{223}\) If the plaintiff fails to take such steps, her recovery for enhanced injuries should be proportionately reduced by such failure. For example, if a plaintiff sues the manufacturer of a motorcycle helmet, alleging a defect in the helmet caused an enhancement of her injuries in a motorcycle accident, the manufacturer should be permitted to introduce evidence, if any, that even though the helmet was defective, the plaintiff wore it incorrectly, thereby contributing to the enhancement of her injuries. Moreover, plaintiffs are deterred from misconduct by the legal consequences of their primary fault; a plaintiff’s recovery for primary injuries is reduced proportionately by the plaintiff’s primary fault.\(^{224}\)

---

\(^{222}\) See Vickles & Oldham, supra note 112, at 440.

\(^{223}\) See Baumgardner, 83 Wash. 2d at 754–57, 522 P.2d at 831–33; Palsgraf v. Long Island R. Co., 162 N.E. 99, 101 (N.Y. 1928); KEETON ET AL., supra note 16, § 43, at 280–81; see also supra notes 18–21, 61–65 and accompanying text.

\(^{224}\) See WASH. REV. CODE §§ 4.22.005, 4.22.015, 4.22.070 (detailing reduction of a plaintiff’s recovery under Washington’s comparative fault scheme); Degel v. Majestic Mobile Manor, Inc., 129 Wash. 2d 43, 48, 914 P.2d 728, 731 (1996) (stating elements of negligence are duty, breach, causation and damages); Geschwind, 121 Wash. 2d at 838, 854 P.2d at 1064 (noting that plaintiff’s negligence relates to failure to use due care for her own protection); KEETON ET AL., supra note 16, § 30, at 164–65 (stating elements of negligence).
C. Reducing a Plaintiff’s Enhanced-Injury Recovery by Primary Fault Will Decrease Product Safety

1. The Purpose of the Enhanced-Injury Doctrine Is To Require Tortfeasors (Not Victims) To Pay for Enhanced Injuries

By adopting the enhanced-injury doctrine, the Washington Supreme Court has provided that product users are entitled to be free from unreasonable risks of injury enhancement in the event of primary accidents. Courts that compare primary fault with enhanced-injury fault when apportioning responsibility for enhanced injuries undermine the enhanced-injury duty by contradicting this bedrock principle of the enhanced-injury doctrine. The purpose of the enhanced-injury doctrine is to hold parties responsible for failing to take reasonable steps to prevent foreseeable risks of enhanced harm. A duty to reasonably minimize enhanced harm is imposed because it is foreseeable that Mary will be involved in a side-impact collision. Therefore, a court that uses a plaintiff’s primary fault to reduce her enhanced-injury recovery uses the very reason for imposing an enhanced-injury duty to reduce the penalty for breach of that duty. Such a reduction is logically inconsistent with the imposition of an enhanced-injury duty. The enhanced-injury tortfeasor (manufacturer) escapes liability due to conduct for which the law has already determined it should be liable, and the reason that the tortfeasor escapes liability is the same reason for which the law imposed liability.

---

225. See Baumgardner, 83 Wash. 2d at 758, 522 P.2d at 833. This is merely the correlative of duty. By imposing a duty of care on one party to reasonably minimize the risk of harm to another party, a court is effectively granting the latter party the right to be free from the unreasonable risk of that harm. See Wesley Newcomb Hohfeld, Some Fundamental Legal Conceptions As Applied in Judicial Reasoning, 23 Yale L.J. 16, 28–32 (1913); Arthur L. Corbin, Legal Analysis and Terminology, 29 Yale L.J. 163 (1919).

226. See Baumgardner, 83 Wash. 2d at 756–58, 522 P.2d at 832–33.

227. See Baumgardner, 83 Wash. 2d at 757, 522 P.2d at 833.

228. This is akin to the following scenario: a duty is imposed to stop at a red light (D) because of the foreseeable risk of collision (FR) if drivers do not stop at red lights. However, the existence of the foreseeable risk of collision (FR) is then used to reduce the punishment for breaching the duty to stop at the red light (D). In other words, D is imposed because of the existence of FR, but the penalty for breaching D is reduced because of the existence of FR.
2. **Permitting Reduction of an Enhanced-Injury Recovery by a Plaintiff’s Primary Fault Reduces Incentives for a Manufacturer To Create Safer Products**

Reducing a plaintiff's enhanced-injury recovery by a plaintiff's primary fault will decrease manufacturers' incentives to implement safety precautions in products and, therefore, lead to decreased product safety. Washington's product liability cause of action imposes liability on a manufacturer where the burden of implementing a safeguard (B) is less than the probability that injury will occur absent the safeguard (P) multiplied by the magnitude of injury that will occur (L).\(^{229}\) This creates an incentive for a manufacturer to undergo the burden of designing and constructing a safety precaution into its product where \(B < PL\).\(^{230}\) Therefore, when \(B < PL\), the result is increased product safety. If a plaintiff's enhanced-injury recovery is reduced by a plaintiff's primary fault, however, the amount of damages that a manufacturer is required to pay for causing enhanced injuries is correspondingly reduced. Instead of the formula reading \(B < PL\), it reads \(B < PL - X\), where \(X\) is the amount of reduction.\(^{231}\) Hence, the manufacturer's incentive to implement the safety precaution is reduced by the factor \(X\), making the incentive \(PL - X\) instead of \(PL\). In cases where the value of \(B\) is between \(PL\) and \(PL - X\), the “\(X\) factor” is determinative. The manufacturer will suffer no liability for failing to implement the safety precaution because, while \(B\) is less than \(PL\), \(B\) is greater than \(PL - X\). Therefore, the manufacturer lacks the incentive to implement the safety precaution and the result is a more dangerous product.

Some commentators have argued that reducing a plaintiff's enhanced-injury recovery by a plaintiff's primary fault is necessary to encourage safety on the part of product users.\(^{232}\) However, there is no need to add this extra layer of incentive for product users to use reasonable care. Product users already have sufficient incentives to use reasonable care, as they are already held accountable for the damages foreseeably caused


\(^{230}\) See generally United States v. Carroll Towing Co., 159 F.2d 169, 173 (2d Cir. 1947).

\(^{231}\) “\(X\)” is the amount of damages for which the tortfeasor is not legally responsible. It is calculated in the following manner: take the percentage of primary fault erroneously attributed to the proximate cause of the enhanced injuries (Cf) and divide it by 100. Then multiply the answer by P. Finally, multiply this second answer by L. In other words, \(X = (Cf/100)PL\).

\(^{232}\) Vickles & Oldham, supra note 112, at 440.
by their fault. 233 For example, Mary is fully responsible for the primary damages resulting from her failure to stop at the stop sign—the passengers’ bumps and bruises, and the dented automobiles. It is extremely undesirable not only to suffer primary damages, but also to be responsible for those primary damages. For that reason, product users will use reasonable care to avoid primary accidents, and an additional layer of incentive is unnecessary.

3. Precluding the Use of Primary Fault To Reduce a Plaintiff’s Recovery for Enhanced Injuries Would Avoid Windfalls for Defendants and Punitive Consequences for Plaintiffs

Using primary fault to reduce the amount that an enhanced-injury tortfeasor must pay for causing enhanced injuries gives the tortfeasor a windfall. Such a reduction relieves the tortfeasor from paying for all of the damages caused by its conduct and, therefore, fails to hold the tortfeasor fully accountable for the consequences of its actions. On the other hand, separating primary fault from enhanced-injury fault not only fails to result in a windfall for plaintiffs, but effectively forces plaintiffs to pay punitive damages. A plaintiff who has breached a duty to prevent primary injuries, but not a duty to prevent injury enhancement, is legally responsible for the primary injuries, but not the injury enhancement. Therefore, reducing the plaintiff’s recovery for enhanced injuries by her primary fault forces her to pay for harm for which she is not responsible and effectively punishes her for being a victim of enhanced injuries.

As one court has recognized, if properly applied, the enhanced-injury doctrine itself prevents windfalls because it inherently apportions fault on a comparative basis. 234 The primary accident tortfeasors are liable only for that portion of the plaintiff’s injuries proximately caused by their breach—the primary injuries—while the enhanced-injury tortfeasors are liable solely for the portion of the plaintiff’s injuries proximately caused by their breach—the enhanced injuries. Thus, the enhanced-injury doctrine implicitly incorporates comparative fault principles and thereby prevents windfalls to both enhanced-injury victims and enhanced-injury tortfeasors.

233. This is basic negligence law, which holds a party responsible for harm when (1) a party had a duty to reasonably minimize a foreseeable risk of harm, (2) the party breached that duty, and (3) proximately caused (4) harm. KEETON ET AL., supra note 16, § 38, at 239.

V. CONCLUSION

Washington courts should not use primary fault to reduce a plaintiff’s recovery for enhanced injuries. The enhanced-injury doctrine is merely an application of basic negligence law; parties are liable for failing to reasonably minimize foreseeable risks of harm. Hence, the adoption of the enhanced-injury doctrine recognizes that manufacturers should be accountable for failing to minimize foreseeable risks of injury enhancement. Since its adoption, however, many courts have undermined the enhanced-injury doctrine by improperly applying comparative fault schemes to enhanced-injury claims. In Washington, comparative fault law requires courts to hold parties responsible for only the consequences of their actions. Thus, when applying comparative fault to enhanced injury claims, Washington courts must take care to determine what conduct constitutes fault with respect to enhanced injuries. Only breach of a duty to minimize injury enhancement—enhanced injury fault—can be considered when apportioning responsibility for enhanced injuries. Permitting use of primary fault to reduce a plaintiff’s recovery for enhanced injuries constitutes a violation of established tort law principles (foreseeability, proximate causation and comparative fault), will result in decreased product safety, and undermines a fundamental purpose of the enhanced injury doctrine and tort law generally: holding tortfeasors responsible for the consequences of their action.
FIRE SALE! THE ADMISSION OF EVIDENCE OF ENVIRONMENTAL CONTAMINATION TO DETERMINE JUST COMPENSATION IN WASHINGTON EMINENT DOMAIN PROCEEDINGS

Paul W. Moomaw

Abstract: Jurisdictions across the United States are split on the issue of whether evidence of environmental contamination should be admissible to determine just compensation in an eminent domain proceeding. Jurisdictions that admit this evidence reason that environmental contamination is a property characteristic that necessarily affects the value of the property. Those that exclude the evidence cite procedural due process concerns and the risk of extra liability for the landowner. Washington’s Model Toxics Control Act (MTCA) establishes a system of assigning liability and recovering cleanup costs for environmental contamination. No Washington court has addressed whether evidence of environmental contamination should be admissible to determine just compensation in an eminent domain proceeding. This Comment argues that, under MTCA and Washington eminent domain law, the evidence should not be admitted, because its admission (1) would violate the prohibition in Washington eminent domain law against speculative evidence, (2) would infringe upon the procedural due process rights of landowners under MTCA, and (3) may result in additional liability on the part of the landowners and extra recovery on the part of the condemning authority.

The Smith family owns a small, independent service station alongside a Washington highway, selling gasoline and services to motorists who pass by.1 Recently, motorists using the highway have increased in number, and the Smiths are delighted to see business booming. However, the increase in traffic has also put pressure on the highway system, and it is clear that the two-lane, winding country road is no longer sufficient. State officials have determined that the only solution is to widen the highway to four lanes. Unfortunately, the Smiths’ service station is in the path of the planned highway expansion. The Smith family soon receives notice that their property is needed for the highway project, but that they will receive fair market value for the land.

Unbeknownst to the Smiths, the land beneath the service station has become a small environmental catastrophe. As appraisers for the State investigate the land to determine its fair market value, they discover that the underground fuel storage tanks are leaking and have been doing so for years. Accordingly, the appraisers come up with a market value for the land that is significantly less than it would be without the newly discovered contamination. Meanwhile, officials from Washington’s

1. Hypothetical created by the author.
Department of Ecology (DOE) have begun an investigation of their own, with the ultimate intention of holding the Smiths liable for the costs of cleanup under Washington’s environmental cleanup statute.

Two legal forces have collided on this Washington highway at the intersection between modern environmental regulations and traditional eminent domain law, leaving the Smiths and their family business as the unwitting victims. Under traditional eminent domain principles, an entity that takes land through the exercise of eminent domain must pay a landowner “just compensation,”\(^2\) which is generally deemed to be the fair market value of the land, based upon all of the “elements reasonably affecting value.”\(^3\) Certainly, environmental contamination is a characteristic that affects the market value of land. On the other hand, Washington’s environmental cleanup statute, the Model Toxics Control Act (MTCA),\(^4\) makes liability strict, joint, and several,\(^5\) and holds past and current landowners alike liable for cleanup.\(^6\) Hence, the Smiths find themselves in the position of receiving less value for their land, while concurrently being held responsible for the cleanup of the contamination.

Jurisdictions across the United States are split on the issue of whether evidence of environmental contamination should be admissible to determine just compensation in an eminent domain action. Some courts have determined that environmental contamination is a property characteristic that bears upon the land’s market value.\(^7\) Therefore, those courts deem evidence of contamination admissible.\(^8\) Others have excluded the evidence, concluding that adjudicating the issue of environmental contamination in an eminent domain proceeding raises various troubling issues, including procedural due process violations and double liability for the landowner.\(^9\) Therefore, these courts hold that the evidence should be excluded.\(^10\) Thus far, no Washington court decision has addressed this issue.

\(^3\) In re Town of Issaquah, 31 Wash. 2d 556, 564, 197 P.2d 1018, 1022 (1948).
\(^4\) WASH. REV. CODE § 70.105D (2000).
\(^5\) Id. § 70.105D.040(2).
\(^6\) Id. § 70.105D.040(1); see also infra note 62 and accompanying text.
\(^7\) See infra Part III.B.
\(^8\) See infra Part III.B.
\(^9\) See infra Part III.A.
\(^10\) See infra Part III.A.
This Comment argues that, under Washington law, evidence of environmental contamination should not be admissible to determine just compensation in an eminent domain proceeding. Part I provides an overview of eminent domain law in Washington, covering its statutory basis, procedural aspects, and judicial interpretation, with particular attention to what constitutes just compensation. Part II discusses MTCA, and gives a synopsis of the relevant regulatory procedures the DOE has enacted. Part III outlines the current state of jurisprudence on the issue of admitting evidence of environmental contamination in eminent domain proceedings in jurisdictions across the United States. Finally, Part IV argues that, given the current state of the law in Washington, evidence of environmental contamination is inappropriate in an eminent domain proceeding because (1) it would violate the prohibition in Washington eminent domain law against speculative evidence, (2) it would infringe upon the procedural due process rights of landowners under MTCA, and (3) it may result in extra liability on the part of the landowner and extra recovery on the part of the condemning authority.

I. EMINENT DOMAIN LAW IN WASHINGTON STATE

Eminent domain is the inherent power of a government to take private property for public use. The power of governmental entities to take property through the exercise of eminent domain is limited by the federal and state constitutions. Every eminent domain proceeding includes a determination that the property will be used for public purposes, and an assessment of just compensation to the owner. Just compensation is defined as the fair market value of the property, or what a willing buyer would pay a willing seller for the property in an open-market transaction. Washington courts rely on numerous valuation methodologies to determine fair market value.
A. Sources of Authority for Eminent Domain

The State of Washington inherently possesses the authority to take private land through the exercise of eminent domain as an attribute of state sovereignty. However, under the United States and Washington constitutions, property taken pursuant to eminent domain must be necessary for “public use,” and the governmental entity must pay the landowner “just compensation.” The state government may delegate the power to local government entities, which do not possess the inherent power of eminent domain.

The Washington Legislature has enacted legislation governing the state’s own eminent domain power and delegating the power to various other entities. The legislation requires governmental entities to perform certain procedures before condemning property under eminent domain. For example, the condemnor must provide the landowner with adequate notice. Moreover, the legislation requires a hearing to determine whether the property is necessary for “public use” and an opportunity for a jury trial to assess “just compensation.”

16. See City of Tacoma v. Welcker, 65 Wash. 2d 677, 683, 399 P.2d 330, 334 (1965) (“The power of eminent domain is an attribute of sovereignty. It is an inherent power of the state, not derived from, but limited by, the fundamental principles of the constitution.”); see also STOEBUCK, supra note 11, § 9.3.

17. The Washington State Constitution provides that “[n]o private property shall be taken or damaged for public or private use without just compensation having first been made.” WASH. CONST. art. I, § 16. Similarly, the Fifth Amendment to the United States Constitution states that “nor shall private property be taken for public use, without just compensation.” U.S. CONST. amend. V. The question of what properly constitutes “public use” is beyond the scope of this Comment, which focuses on the “just compensation” requirement.

18. See Welcker, 65 Wash. 2d at 683, 399 P.2d at 334–35 (stating that a municipal corporation, unlike the state government, does not have the inherent power to take property through eminent domain, but “may exercise such power only when it is expressly so authorized by the state legislature”); see also STOEBUCK, supra note 11, § 9.3.

19. WASH. REV. CODE §§ 8.04.010–191 (state), 8.08.010–150 (counties), 8.12.010–580 (cities), 8.16.010–160 (school districts), 8.20.010–180 (corporations) (2000). This Comment focuses on the issue from the broad perspective of the state’s power of eminent domain; any difference between it and the more limited powers of other entities is insignificant for the purposes of this Comment.

20. Id. § 8.04.020.

21. Id. §§ 8.04.070–080. Of course, the parties may forego judicial proceedings if they settle upon an amount of compensation that is satisfactory to all. Washington’s eminent domain statute provides that “[e]very reasonable effort shall be made to acquire expeditiously real property by negotiation.” Id. § 8.26.180.
B. The Exercise of Eminent Domain

A Washington eminent domain proceeding consists of three discrete phases: a finding that the property to be condemned is necessary for public use, an assessment of just compensation, and an order to transfer title. The condemning authority initiates an eminent domain action by submitting a petition for appropriation to the superior court of the county in which the property sits and serving notice upon all interested parties. Next, a hearing is held to ensure that the condemning authority has notified all parties with an interest in the subject property and that the property is truly necessary for public use. If the government successfully shows that the land is necessary for public use, the court will issue an order adjudicating that the contemplated use of the property is truly a public use, as well as an order to determine damages for the taking of the property.

If the court issues an order for the determination of damages, a jury trial is held to determine the proper amount of just compensation for the property. This portion of the proceeding tends to be the most hotly debated: the great majority of eminent domain cases focus upon the issue of the proper measure of just compensation. Both the landowner and

23. WASH. REV. CODE § 8.04.010. The petition must contain the names of any party that has an interest in the property to be condemned, a description of the property, a description of the interests to be condemned, the purposes for which the land is to be condemned, the authority for the condemnation, a request that a jury determine the proper amount of just compensation, and a prayer that the court award the property interest to the condemning authority. STOEBUCK, supra note 11, § 9.27.
24. WASH. REV. CODE § 8.04.020. Generally, the notice must contain a statement that the condemnor seeks to condemn the property, a description of the property to be condemned, and a statement of the place and time at which the petition is to be submitted to the court. STOEBUCK, supra note 11, § 9.27.
25. WASH. REV. CODE § 8.04.070. “[W]hether the contemplated use be really public shall be a judicial question, and determined as such, without regard to any legislative assertion that the use is public.” WASH. CONST. art. I, § 16. This language explains the requirement that there be a hearing to determine whether the land has actually been taken for a public use. STOEBUCK, supra note 11, § 9.20. To meet the requirement that the property be necessary for public use, the court must make a determination “(1) that the use is really public, (2) that the public interests require it, and (3) that the property appropriated is necessary for the purpose.” In re City of Seattle, 96 Wash. 2d 616, 625, 638 P.2d 549, 555 (1981).
27. Id. § 8.04.080.
28. Id.
29. STOEBUCK, supra note 11, § 9.1.
the condemning authority may present evidence bearing upon just compensation, as neither party has the burden of proof with respect to valuation.30

C. Just Compensation

An entity that condemns private property under eminent domain must pay the landowner “just compensation.”31 Just compensation is based upon the fair market value of the property, as defined by the amount a willing buyer would pay a willing seller on the open market.32 The driving principle behind the constitutional just compensation requirement is that the property owner is entitled to be placed in the same monetary position as he or she would have been had the property not been taken.33 Consequently, the condemnee must receive full value for the property.34

I. Valuation

Washington courts accept numerous valuation methodologies for purposes of measuring the fair market value of property. The most common valuation methods used in Washington are the capitalized rental value method, the depreciated replacement cost method, and the comparable sales method.35 Under the capitalized rental value approach, the value of the land is measured according to the income it produces.36 Under the depreciated replacement cost method, an appraiser determines the replacement cost of improvements, adjusts for depreciation, then adds the market value of the land itself.37 Finally, under the comparable sales method, a land’s value is determined by comparing it to properties

31. WASH. CONST. art. I, § 16.
33. State v. Trask, 98 Wash. App. 690, 697, 990 P.2d 976, 980 (2000) (concluding that pre- and post-judgment interest are to be included in just compensation).
34. Id.
35. STOEBUCK, supra note 11, § 9.30.
36. See 4 JULIUS SACKMAN, NICHOLS ON EMINENT DOMAIN § 12.02[1], at 12-72 (3d ed. Rev. 2000). At its most basic, this method would involve valuing the property according to how much rental income it could generate, capitalized at a reasonable discount rate.
37. See id. Of course, the market value of the underlying land would have to be determined using one of the other methodologies.
that have been sold within a reasonable period of time, and that are similar in location, use, improvements, and other qualities. 38

Parties may present testimony regarding the value of the property through expert witnesses, the property owner, or neighboring property owners. Although experts typically testify about the value of property, 39 “[a]n owner of property may testify as to its value (without qualifying as an expert), upon the assumption that he is particularly familiar with it and, because of his ownership, knows of the uses for which it is particularly adaptable.” 40 Neighbors may also be particularly knowledgeable with respect to the value of the property, and may be allowed to testify thereto. 41 For example, the Washington Supreme Court has upheld admission of testimony of a neighbor who had only seen the subject property from the road, reasoning that the decision to admit the testimony was within the trial judge’s sound discretion, and that such testimony would go to the weight rather than the admissibility of the evidence. 42

2. Speculative Evidence

While Washington courts allow the trier of fact to consider a wide array of factors that may affect the fair market value of property, 43 these factors must meet two basic requirements: (1) they must actually affect the property’s fair market value, and (2) they must be established by the evidence. 44 Hence, evidence that is overly speculative is not admissible to determine fair market value. 45 For example, in In re City of Medina, 46 the court held that the fair market value of unimproved property could not be determined by comparing it with town lots that were fully platted and developed. 47 Basing a determination of the property’s fair market

38. See id.
42. Id.
43. See In re Town of Issaquah, 31 Wash. 2d 556, 564, 197 P.2d 1018, 1022 (1948).
45. Id.; see also In re City of Medina, 69 Wash. 2d 574, 578, 418 P.2d 1020, 1022–23 (1966).
46. 69 Wash. 2d 574, 418 P.2d 1020 (1966).
47. Id. at 578, 418 P.2d at 1022–23.
value on such evidence, the court held, would involve “pure conjecture.”

In State v. Mottman Mercantile Co., the Washington Supreme Court stated that land containing mineral content may not be valued by multiplying the number of cubic yards of the mineral in the land by a unit price of the mineral as extracted. Mottman involved a piece of property to be condemned that had potential value as a gravel pit. The court stated that evidence of the present value of the mineral content in its natural state was admissible on a cubic yard basis. However, evidence of the gravel’s market value as extracted would be inadmissible for determining the land’s fair market value, because such evidence would require speculation about the cost of extraction, the extent and duration of market demand for the minerals, marketing costs, and other variable factors.

Other decisions indicate that Washington courts consider the fear of potentially dangerous conditions upon land to be overly speculative and, therefore, inadmissible to determine just compensation in an eminent domain proceeding. In Pacific Northwest Pipeline Corp. v. Myers, a corporation condemned an easement for the installation of a gas pipeline. The court approved a jury instruction directing the jury to ignore a witness’s testimony regarding the effect that fear of gas transmission lines had on the market value of the subject property. Additionally, in State v. Evans, the Washington Supreme Court reversed a lower court’s decision that had admitted evidence of fear. The lower court had declared that “[t]he psychological effect of an adverse condition, real or

50. Id. at 724–25, 321 P.2d at 914.
51. Id. at 725, 321 P.2d at 914.
54. 50 Wash. 2d 288, 311 P.2d 655 (1957).
55. Id. at 290, 311 P.2d at 656. Although the court was unwilling to reverse on the grounds that the testimony was given, its refusal to do so was predicated on the fact that the limiting instruction had been given. Id.
57. Id. at 128.
imagined, on a potential buyer may [materially influence] the market value of property.” 58 Although the Washington Supreme Court did not directly address the lower court’s assertion that “real or imagined” psychological effects may be considered in assessing fair market value, it expressed concern that the jury was permitted to consider speculative elements. 59 The court reaffirmed Washington’s standard that the only elements that a jury should consider “are those which will actually affect the fair market value of the property and which are established by the evidence.” 60

II. THE MODEL TOXICS CONTROL ACT

The Model Toxics Control Act (MTCA) 61 is Washington’s counterpart to the federal environmental cleanup statute, the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA). 62 MTCA provides a statutory framework to allocate liability for environmental contamination, to raise funds for the cleanup of contamination, and to prevent future contamination. 63 MTCA makes owners of contaminated property and other parties that are responsible for the contamination liable for its cleanup. 64 It also creates certain limited defenses to liability. 65 Furthermore, MTCA requires the prioritization of particularly contaminated sites 66 and compels remedial action upon those sites that are deemed to warrant investigation and

58. State v. Evans, 26 Wash. App. 251, 261, 612 P.2d 442, 449 (1980) (“These effects and their impact on the market value have been recognized in cases involving the inherent fear of electricity and gas transmission lines.”), rev’d, 96 Wash. 2d 119, 634 P.2d 845, opinion amended, 649 P.2d 633 (1982).
59. Evans, 96 Wash. 2d at 127, 634 P.2d at 849.
60. Id.
63. WASH. REV. CODE § 70.105D.010. A “contaminant” is “any hazardous substance that does not occur naturally or occurs at greater than natural background levels.” WASH. ADMIN. CODE § 173-340-200 (2001). “‘Environment’ means any plant, animal, natural resource, surface water (including underlying sediments), ground water, drinking water supply, land surface (including tidelands and shorelands) or subsurface strata, or ambient air within the state of Washington or under the jurisdiction of the state of Washington.” Id.
64. WASH. REV. CODE § 70.105D.040(1).
65. Id. § 70.105D.040(3).
66. Id. § 70.105D.030(2)(b).
cleanup. MTCA provides due process to landowners by giving them the right to assert defenses to liability.

A. Liability and Defenses Under MTCA

Under MTCA, any current owner or operator of a contaminated site, and any past owner who contributed to the contamination, is a potentially liable party (PLP) for the purpose of remediating the contamination. Liable parties are subject to strict, joint, and several liability for all remedial action costs and any damages to natural resources that occur as a result of the release “or threatened release” of hazardous substances. MTCA requires PLPs to conduct remedial actions or pay the state to conduct those remedial actions. Washington’s Attorney General may recover all remedial action costs from the liable parties at the request of the Department of Ecology (DOE).

MTCA provides certain affirmative defenses that may absolve the PLP of liability. These defenses are primarily limited to those situations in which the PLP is innocent, oblivious, or both, with respect to environmental contamination. If the PLP can establish that the contamination was caused by an “act of God,” an “act of war,” or “the

67. See infra Part II.B.
68. See infra Part II.D.
69. WASH. REV. CODE § 70.105D.040.
70. Strict liability is “[l]iability that does not depend on actual negligence or intent to harm, but that is based on the breach of an absolute duty to make something safe.” BLACK’S LAW DICTIONARY 926 (7th ed. 1999). Joint and several liability is:
71. Id.
72. See infra notes 127–29 and accompanying text.
73. WASH. REV. CODE § 70.105D.040(2).
74. Id. §§ 70.105D.030(1)(b), .040(2).
75. Id. § 70.105D.040(2).
76. Id. § 70.105D.040(3).
77. Id. § 70.105D.040(3)(a)(i). MTCA does not define “act of God.”
78. Id. § 70.105D.040(3)(a)(ii). MTCA does not define “act of war.”
act or omission of a third party,”79 the PLP is not liable under MTCA.80
Moreover, if the PLP can show by a preponderance of the evidence that,
on upon acquiring the property, the PLP neither knew nor had reason to
know that any hazardous substance had been released or disposed of on
the property, the PLP is not liable under MTCA.81

MTCA absolves governmental entities from liability in certain
circumstances. The statute states that the term “owner/operator,” for
the purposes of MTCA liability, does not include governmental agencies
“which acquired ownership or control involuntarily.”82 It is not clear
whether MTCA’s involuntary acquisition exception to a government’s
MTCA liability includes the exercise of eminent domain. However, some
commentators have indicated that a similar provision in CERCLA83 may
include circumstances in which the government acquires the property
through eminent domain.84 Because Washington courts often look to
CERCLA for guidance in interpreting MTCA,85 it is possible that such an
interpretation would extend to MTCA as well. Thus, it is unclear whether
a governmental entity that acquires contaminated property through
eminent domain is liable for remediation under MTCA.

---

79. Id. § 70.105D.040(3)(a)(iii). This defense does not apply if the potentially liable party had a
direct or indirect contractual relationship with the third party. Id.
80. Id. § 70.105D.040(3).
81. Id. § 70.105D.040(3)(b). This is subject to certain restrictions. For example, the property
owner must have made reasonable inquiries into prior uses of the property to show that he or she had
no reason to know of the contamination. Id. § 70.105D.040(3)(b)(i). Moreover, the court will take
into consideration the specialized knowledge of the buyer, the purchase price for the property, the
obviousness of the condition, and commonly known or reasonably ascertainable information about
the property. Id.
82. Id. § 70.105D.020(12)(b)(i). Involuntary acquisition, for the purposes of this subsection,
means acquiring the property “through bankruptcy, tax delinquency, abandonment, or circumstances
in which the government involuntarily acquires title.” Id.
84. See, e.g., Lawrence P. Schnapf, CERCLA and the Substantial Continuity Test: A Unifying
(stating that 42 U.S.C. § 9601(20)(D) “applies to state or local governments that acquired title to a
contaminated facility involuntarily through . . . eminent domain”); cf. Robert I. McMurry & David
H. Pierce, Environmental Remediation and Eminent Domain, C709 ALI-ABA 105, 117 (1992)
(stating that, at least in cases in which the government condemns property for cleanup of
contaminated property, it is “involuntarily” acquiring property to protect public health and safety).
B. Assessing Contamination and Determining Liability Under MTCA

Washington’s legislature has delegated the authority to enforce and administer MTCA to the DOE, which has promulgated regulations governing the procedural aspects of MTCA. The regulations provide for the investigation and assessment of potentially hazardous sites, and outline the various actions that may be taken for the remediation of those sites. They also establish criteria for determining the proper cleanup level for a given piece of property.

1. Investigation and Assessment

Both MTCA and the DOE regulations outline methods by which the DOE can identify contaminated sites. The DOE regulations provide that any owner or operator of a site who discovers the release of a hazardous substance that poses a potential threat to human health must notify the DOE within ninety days of the discovery. MTCA provides that if the DOE has a reasonable basis to believe that a hazardous substance has been or may be released, it may enter upon the property to conduct further investigation, upon notice to the landowner. Furthermore, the DOE regulations provide that the DOE may take any other actions consistent with MTCA to identify potentially contaminated sites.

Upon discovering a potentially hazardous site, the DOE will investigate the site to decide what action, if any, to take with respect to the site. Within ninety days of discovering the release of a hazardous substance, the DOE will conduct an initial site investigation consisting of, at a minimum, a visit to the site to document the conditions observed at the site. Based upon the results of the investigation, the DOE will

88. See infra Part II.B.1.
89. See infra notes 101–03 and accompanying text.
93. Id. § 173-340-310.
94. Id. § 173-340-310(2). The DOE may defer to another governmental body or independent contractor for the purposes of conducting the investigation, as long as the other entity did not contribute to the hazardous condition at the site. Id. § 173-340-310(3).
Environmental Contamination and Eminent Domain

decide to undertake either (1) a site hazard assessment,95 (2) emergency remedial action,96 (3) interim action,97 or (4) no further action.98 The DOE maintains a list of sites requiring remedial action and sets priorities for remedial action based upon the results of the site hazard assessment.99 Once the DOE has credible evidence of liability and is prepared to proceed with remedial action, it notifies PLPs of their potential liability.100

The requisite level of cleanup of a piece of property depends largely upon the nature of the site.101 The proper level of cleanup is dictated by the specific hazardous substances found at a site and the specific media, such as soil, air, or water, by which humans and the environment could become exposed to the hazardous substances.102 To determine the proper method to set cleanup levels, the DOE must examine the nature of the contamination, the current and potential pathways of human and environmental exposure to the contamination, the current and potential receptors of the contamination, and the current and potential use of the land.103

95. Id. § 173-340-310(5)(a). "A site hazard assessment is an early study to provide preliminary data regarding the relative potential hazard of the site." Id. § 173-340-320(4).
96. Id. § 173-340-310(5)(b). If a site requires emergency remedial action, the DOE must notify the potentially affected area of the threat. Id. § 173-340-310(6)(a). "Emergency remedial action" is not defined by the regulations.
97. Id. § 173-340-310(5)(c). An interim action may include an action necessary to (1) reduce the threat to human health by eliminating or reducing "pathways for exposure" to a hazardous substance, (2) correct a problem that may become worse or more expensive if action is delayed, or (3) complete a site hazard assessment, remedial investigation, or feasibility study, or provide for the design of a cleanup action. Id. § 173-340-430(1).
98. Id. § 173-340-310(5)(d). A decision to undertake no further action may be predicated upon a determination that no hazardous substance has been released, no threat to human health exists, or any further action is more appropriately undertaken under a different authority. Id.
99. Id. §§ 173-340-120(3), -330. A site’s appearance on the hazardous sites list is not an implication that the parties associated with the site are subject to MTCA liability. Id. § 173-340-330(5).
100. Id. § 173-340-500; see also WASH. REV. CODE § 70.105D.040 (2000).
102. Id. § 173-340-700(4)(a).
103. Id. § 173-340-700(5). The Washington Administrative Code outlines three basic methods to determine the proper level of remediation. Id. "Method A" is used for sites for which cleanup is routine, or which involve relatively few hazardous substances. Id. § 173-340-700(5)(a). "Method B" is the standard method used for all sites, unless one of the conditions applicable to Methods A or C exists. Id. §§ 173-340-700(5)(b), -705(1). "Method C" is a conditional method, applicable if compliance with Methods A or B is impossible or will cause greater environmental harm. Id. § 173-340-700(5)(c). Method C is also applicable to certain industrial properties. Id.; see also id. § 173-340-745 (setting forth soil cleanup standards for industrial properties).
2. Remedial Actions

A remedial action is any action intended to identify and either eliminate or minimize a threat to human health or the environment caused by the release of hazardous substances. The DOE may initiate a remedial action by sending a negotiation letter or enforcement order to a PLP or by requesting an “agreed order.” The negotiation letter informs the PLP that the DOE wishes to negotiate toward a consent decree. The letter typically explains the nature of the DOE’s conclusions about a contaminated site, requests a written statement of the PLP’s willingness to negotiate, and asks for the names of other PLPs. An enforcement order requires a PLP to take action on its own. Under an agreed order, the PLP agrees to undertake remediation of the site. In return, the DOE will not take action against the PLP as long as the PLP complies with the order. However, the agreed order is not a settlement and, therefore, will not contain covenants not to sue or protection from claims of contribution. Finally, the DOE may undertake remedial action entirely on its own if necessary, as in the case of an emergency.

A PLP may also initiate a remedial action by demanding a settlement under a consent decree, by requesting an agreed order, or simply by taking action itself. In a settlement under consent decree, the PLP accepts responsibility for the contamination and proposes a remedial action plan for the site. In requesting a consent decree, the PLP must also identify other PLPs and provide information regarding the history

104. WASH. REV. CODE § 70.105D.020(21).
105. Id.
107. Id. § 173-340-520(2).
108. Id. § 173-340-520(2)(b).
109. Id. § 173-340-540. The DOE generally issues the enforcement order either after it has sent a negotiation letter or else concurrently, in the case of an emergency. Id.
110. Id. § 173-340-530(1).
111. Id.
112. Id.
113. Id. § 173-340-510(4).
114. Id. §§ 173-340-510(2), 515.
115. Id. § 173-340-520; see also WASH. REV. CODE § 70.105D.040(4) (2000).
Environmental Contamination and Eminent Domain

and use of the site. The Attorney General may accept the settlement if it appears that the proposed plan is in compliance with cleanup standards under MTCA and would expedite the cleanup process. Finally, an independent remedial action is a remedial action undertaken by a PLP outside of an order or decree and without the approval or oversight of the DOE. The DOE may, however, provide a limited amount of informal assistance. Within ninety days of completion, a PLP that initiates such an action must submit to the DOE a report containing a description of its investigation, remediation, and monitoring conducted on the property. The DOE must inform the PLP whether any further action is necessary within ninety days of receiving the report.

C. Recovery of Remedial Action Costs

Under MTCA, the Washington Attorney General can file an action to recover remedial action costs that the DOE has spent to clean a contaminated site. The Attorney General is authorized to file an action if necessary to recover all costs incurred, including the costs of undertaking any investigative and remedial actions. Remedial action costs are any costs “reasonably attributable” to the site, and include costs of direct remedial activities, support costs, and interest charges. The Attorney General may also bring an action against a PLP that has failed to comply with an enforcement order or an agreed order issued by the

117. WASH. REV. CODE § 70.105D.040(4)(a); WASH. ADMIN. CODE § 173-340-520(f).
118. WASH. ADMIN. CODE § 173-340-515.
119. Id.
120. Id. § 173-340-515(5); see also 24 TIMOTHY H. BUTLER, ENVIRONMENTAL LAW AND PRACTICE, in WASHINGTON PRACTICE, § 15.26 (1997).
121. WASH. ADMIN. CODE § 173-340-300(4).
122. Id. §§ 173-340-310(4), -515(4).
123. WASH. REV. CODE § 70.105D.050(3) (2000).
124. Costs of direct remedial activities include, for example, the payment of staff for work on the site, the cost of any travel related to the site, costs incurred publishing documents with respect to the site, the purchase or rental of any equipment necessary for remediation of the site, and the cost of work on the site that must be contracted out. WASH. ADMIN. CODE § 173-340-550(2)(a).
125. Support costs include, for example, costs of any facilities, personnel, and administrative support that are indirectly related to the remediation of the site. Id. § 173-340-550(2)(b)–(c).
126. Id. § 173-340-550(2).
DOE. In such an action, the Attorney General may recover up to three times the amount that the state must incur to undertake remediation activities, as well as a civil penalty of up to $25,000 for each day that the PLP refuses to comply.

MTCA provides that any “person” who undertakes a remedial action on its own may recover cleanup costs and litigation expenses from PLPs. This private right of action encourages private parties to undertake remedial actions independently. Moreover, the DOE and other governmental entities may also bring a private right of action, as MTCA’s definition of “person” includes “state government agenc[ies] [and] unit[s] of local government.” Hence, any authority that may exercise the power of eminent domain may also bring a private right of action under MTCA.

A party seeking recovery under the private right of action may recover only remedial action costs that are substantially equivalent to those that the DOE would have undertaken. To facilitate private rights of action, the DOE has enumerated a number of elements that a remedial action should contain in order to meet the substantial equivalent requirement. For example, information on the site and its remediation must be reported to the DOE, the DOE must not object to the action, and the public at large must be notified of the remediation.

D. Procedural Due Process Under MTCA

MTCA provides due process to PLPs by giving them the right to assert defenses to liability. The right to due process of law is

127. WASH. REV. CODE § 70.105D.050(1).
128. Id.
129. Id. § 70.105D.080.
130. Id. § 70.105D.020 (14). “Person” is defined as an “individual, firm, corporation, association, partnership, consortium, joint venture, commercial entity, state government agency, unit of local government, federal government agency, or Indian tribe.” Id.
131. Id. § 70.105D.080.
133. Id. § 173-340-545(2)(c)(i).
134. Id. § 173-340-545(2)(c)(ii).
135. Id. § 173-340-545(2)(c)(iii). The remedial action may also be “substantially equivalent” if it has actually been conducted by the DOE itself, or if it is being conducted under an order or decree, and the requirements of the decree have been fulfilled. Id. § 173-340-545(2)(a)–(b).
136. See supra notes 75–81 and accompanying text.
Environmental Contamination and Eminent Domain

embedded in the Washington State and United States Constitutions. The Washington State Constitution provides that “[n]o person shall be deprived of life, liberty, or property, without due process of law.” Similarly, the Fourteenth Amendment to the United States Constitution states that a state may not “deprive any person of life, liberty, or property, without due process of law.” In Washington, procedural due process requires that parties whose rights are affected by a governmental proceeding have a meaningful opportunity to be heard, and notice calculated to advise the parties of the proceeding to allow them the opportunity to present defenses. If any “significant property interest” is at stake, the safeguards afforded by due process are applicable.

Encompassed within the constitutional right to due process is the “guarantee of fair procedure,” or procedural due process. Procedural due process imposes limitations upon governmental action that deprives individuals of life, liberty, or property. In a procedural due process claim, the question is not whether the deprivation of “life, liberty, or property” is itself unconstitutional, but whether the deprivation took place without the procedural guarantees envisioned by the Constitution. The rules of procedural due process “minimize substantively unfair or mistaken deprivations of” life, liberty, or property by enabling persons to contest the basis upon which a State proposes to deprive them of protected interests.

137. WASH. CONST. art. I, § 3.
140. Id. at 428, 511 P.2d at 1008. In determining the level of process that is due, the court will balance the interest to be protected, the risk of deprivation of that interest by the government’s acts, and the government’s interest in maintaining the procedure. Silver Firs Town Homes, Inc. v. Silver Lake Water Dist., 103 Wash. App. 411, 425, 12 P.3d 1022, 1029 (2000).
143. Zinermon, 494 U.S. at 125.
144. Carey v. Piphus, 435 U.S. 247, 259–60 (1978) (quoting Fuentes v. Shevin, 407 U.S. 67, 81 (1972)). Whether procedural due process protections are implicated in a given case depends upon whether the interest in question falls within the ambit of “life, liberty, or property.” Morrissey v. Brewer, 408 U.S. 471, 481 (1972). Neither the Washington State nor the United States Constitutions define exactly what that phrase encompasses. The United States Supreme Court has made it clear, however, that the scope of “property interests” is broader than simply the ownership of money, real estate, or chattel. Bd. of Regents of State Colls. v. Roth, 408 U.S. 564, 571–72 (1972). Similarly, “liberty” means something more than simply freedom from physical confinement. Id. at 572 n.11.
In *Mathews v. Eldridge*, the United States Supreme Court articulated a three-part balancing test for identifying the requirements of procedural due process. Under this test, a court should attempt to balance (1) the private interests that are affected by a governmental action, (2) the risk that the procedures employed by a governmental actor will result in the mistaken deprivation of those interests, and (3) the government’s own interests, including the economic and administrative burden that the procedural protection would entail. In implementing this test, the Court has generally held that some minimal opportunity to be heard is required before the government deprives an individual of liberty or property.

One of the fundamental requirements of procedural due process is the opportunity to be heard “at a meaningful time and in a meaningful manner.” For the opportunity to be meaningful, it must be appropriate to the nature of the case. For example, *Bell v. Burson* involved a due process challenge to a statute that divested uninsured motorists who were

Moreover, for purposes of procedural due process, property interests “are not created by the constitution but are reasonable expectations of entitlement derived from independent sources such as state law.” *Mission Springs, Inc. v. City of Spokane*, 134 Wash. 2d 947, 962 n.15, 954 P.2d 250, 257 n.15 (1998) (citing Roth, 408 U.S. at 577). Thus, a property interest may be created by a state statute or statutory scheme. See *Mathews*, 424 U.S. at 332; *Mission Springs*, 134 Wash. 2d at 963, 954 P.2d at 257 (citing Bateson v. Geisse, 857 F.2d 1300, 1304–05 (9th Cir. 1988)).

146. Id. at 335 (citing Goldberg v. Kelly, 397 U.S. 254, 263–71 (1970)).
147. Id.

149. Goldberg, 397 U.S. at 267 (citing Armstrong v. Manzo, 380 U.S. 545, 552 (1965)).
150. *Bell* at 542.
involved in accidents of their driver’s licenses unless they posted security sufficient to cover the damage alleged by the other party to the accident.\textsuperscript{152} The United States Supreme Court held that, under the statute, motorists were entitled to a hearing wherein they were afforded an opportunity to present defenses to liability.\textsuperscript{153} Further, \textit{Goldberg v. Kelly}\textsuperscript{154} involved a procedural due process claim for the wrongful termination of welfare benefits. The Supreme Court similarly held that the “meaningful” requirement dictated that individuals who were being deprived of their public assistance should have notice of the reasons for the termination and an opportunity to defend themselves by confronting adverse witnesses and presenting arguments and evidence.\textsuperscript{155}

In Washington, procedural due process requires that parties whose rights are affected by a governmental proceeding have, at a minimum, a meaningful opportunity to be heard and notice calculated to advise them of the proceeding, to allow the party to present defenses.\textsuperscript{156} Moreover, Washington courts also hold that the opportunity must be presented “at a meaningful time and in a meaningful manner.”\textsuperscript{157} If any “significant property interest” is at stake, the safeguards afforded by due process are applicable.\textsuperscript{158} In \textit{Olympic Forest Products, Inc. v. Chaussee Corp.},\textsuperscript{159} the Washington Supreme Court articulated the specific procedures required by due process. It determined that a court must consider the nature of the affected interest, the manner in which it is affected, the government’s reasons for acting as it did, what procedural alternatives are available, the amount of protection that ought to be given to the governmental actor, and the balance between the benefit accomplished and the detriment suffered.\textsuperscript{160}

\begin{itemize}
  \item \textsuperscript{152} Id. at 535–36.
  \item \textsuperscript{153} Id. at 541–42.
  \item \textsuperscript{154} 397 U.S. 254 (1970).
  \item \textsuperscript{155} Id. at 267–68.
  \item \textsuperscript{156} Olympic Forest Prods., Inc. v. Chaussee Corp., 82 Wash. 2d 418, 422, 511 P.2d 1002, 1005 (1973) (citing Armstrong v. Manzo, 380 U.S. 545, 552 (1965); Mullane v. Cent. Hanover Bank & Trust Co., 339 U.S. 306, 314 (1950); Grannis v. Ordean, 234 U.S. 385, 394 (1914)). In Washington, the court will balance the interest to be protected, the risk of deprivation of that interest by the government’s acts, and the government’s interest in maintaining the procedure. \textit{See, e.g.}, Silver Firs Town Homes, Inc. v. Silver Lake Water Dist., 103 Wash. App. 411, 425, 12 P.3d 1022, 1029 (2000).
  \item \textsuperscript{157} Olympic Forest Prods., 82 Wash. 2d at 422, 511 P.2d at 1005.
  \item \textsuperscript{158} Id. at 428, 511 P.2d at 1008.
  \item \textsuperscript{159} 82 Wash. 2d 418, 511 P.2d 1002 (1973).
  \item \textsuperscript{160} Id. at 423–24, 511 P.2d at 1006 (citing Joint Anti-Fascist Refugee Comm. v. McGrath, 341 U.S. 123, 163 (1951) (Frankfurter, J., concurring)).
\end{itemize}
III. JURISDICTIONS ARE SPLIT ON THE ADMISSIBILITY OF EVIDENCE OF ENVIRONMENTAL CONTAMINATION IN EMINENT DOMAIN PROCEEDINGS

There is a split of authority across the United States on the issue of whether evidence of environmental contamination should be admissible to determine just compensation in eminent domain proceedings.\textsuperscript{161} Jurisdictions that refuse to admit such evidence have expressed concern that it would circumvent procedures already established by environmental laws and regulations for determining liability for contamination and result in additional liability for the condemnee. Jurisdictions that admit the evidence reason that contamination is a property characteristic—which must necessarily be taken into account to determine fair market value—and that environmental contamination creates a “stigma” upon property.

A. Excluding the Evidence

1. Considerations of Additional Liability

At least one court has refused to admit evidence of environmental contamination in an eminent domain proceeding in order to prevent the potential for extra liability for the landowners.\textsuperscript{162} In \textit{Aladdin, Inc. v. Black Hawk County},\textsuperscript{163} a county condemned a contaminated laundry facility in preparation for building a new jail\textsuperscript{164} and sought to reduce the compensation it paid for the property on account of the environmental contamination.\textsuperscript{165} Because Iowa law provided for administrative procedures to remedy environmental contamination, the Iowa Supreme


\textsuperscript{162} \textit{Aladdin}, 562 N.W.2d at 615.

\textsuperscript{163} 562 N.W.2d 608 (Iowa 1997).

\textsuperscript{164} \textit{Id. at 610.}

\textsuperscript{165} \textit{Id.}
Environmental Contamination and Eminent Domain

Court expressed concern that the landowner would incur additional liability if evidence of the contamination were admitted. The court reasoned that if the proper administrative procedures were not followed with respect to the environmental remediation, and the contamination were figured into the amount of compensation for the property, “the landowner will not receive just compensation because the award will be less than full value. In addition, the property owner will have the same legal liability for cleanup cost as before.”

2. Procedural Due Process Concerns

The Aladdin court was also concerned that admitting contamination evidence would compromise the landowner’s due process rights. In Iowa, as in Washington, the environmental cleanup statute contained procedural safeguards for the benefit of the landowner. The Aladdin court held that admitting evidence of environmental contamination would deprive the landowner of just compensation. In making its determination, the Aladdin court cited procedural due process concerns, expressing the fear that the scope of procedural guarantees available in an environmental liability adjudication would not be available in the context of eminent domain. The court reasoned that a landowner has the right to have environmental cleanup liability adjudicated in a proceeding in which there is an opportunity to show that the landowner is not responsible for the contamination. Therefore, establishing the landowner’s liability in a condemnation proceeding would violate the landowner’s procedural due process rights. Furthermore, if another party is able to prove that the landowner is legally responsible for the contamination, the remediation costs can be recovered in an environmental action after completion of the eminent domain proceeding.

166. Id. at 615.
167. Id. at 615–16.
168. See id. at 615.
169. Id.
170. Id. at 615–16.
171. Id. at 615.
172. Id.
173. Id. The Aladdin court was also concerned that valuation of contaminated property would be problematic due to the difficulty of locating comparable contaminated property, and that the fact-
In *Department of Transportation v. Parr*,\(^{174}\) the Illinois Court of Appeals also excluded evidence of environmental contamination in an eminent domain proceeding due to procedural due process concerns.\(^{175}\) In *Parr*, the Illinois Department of Transportation condemned a landowner’s contaminated property to construct a bridge.\(^{176}\) The Illinois environmental statute, like MTCA, set forth a comprehensive process for remediating environmental contamination and adjudicating liability.\(^{177}\) Among other things, the statute provided for certain procedural safeguards to protect landowners when adjudicating liability for environmental contamination.\(^{178}\) The court held that admitting evidence of environmental contamination would violate the procedural due process rights of the condemnees, because it would allow the condemning authority to circumvent the procedural safeguards implemented to protect the rights of landowners under the statutory scheme.\(^{179}\) The court noted that procedural due process requires that “orderly proceedings” advance according to rules that do not violate the fundamental rights of the parties involved.\(^{180}\) Therefore, depriving landowners of the rights and defenses afforded by the Illinois environmental statute would constitute a violation of procedural due process.\(^{181}\)

### B. Admitting the Evidence: Environmental Contamination as a Property Characteristic

Some courts admit evidence of environmental contamination in eminent domain proceedings on the basis that it is a property characteristic and necessarily affects the fair market value of property.\(^{182}\)

\(^{175}\) Id. at 22.
\(^{176}\) Id. at 20.
\(^{177}\) Id. at 22–23.
\(^{178}\) Id. at 22.
\(^{179}\) Id.
\(^{180}\) Id. at 23.
\(^{181}\) Id. at 22.
For example, in *State v. Hughes*, an Oregon court of appeals held that the evidence of contamination is admissible in an eminent domain action. In *Hughes*, the State of Oregon condemned contaminated property for a highway improvement project. After condemning the property, the condemnor discovered that petroleum contamination was present in the groundwater beneath the site. Citing Oregon’s evidentiary rule for relevance, the court reasoned that, although the contamination was discovered subsequent to the condemnation, the evidence could have been discovered on the date the action was commenced and would bear on the fair market value of the property at the time. Therefore, the evidence would “easily pass the threshold for relevance” and should be admitted in a proceeding to determine just compensation.

In *Finkelstein v. Department of Transportation*, the Florida Supreme Court admitted evidence of environmental contamination to determine just compensation in an eminent domain proceeding on the basis that there is a “stigma” associated with contaminated property that affects its market value. In *Finkelstein*, the Florida Department of Transportation condemned a piece of property in connection with the construction of an interstate highway. The Florida Supreme Court held that, as long as the facts show that environmental contamination actually has an effect on the market value of the property to be condemned, evidence of environmental contamination is admissible to determine just

---

184. *Id.* at 703.
185. *Id.* at 701.
186. *Id.* at 702. The court did not discuss any applicable environmental remediation statute in Oregon.
187. “Relevant evidence” means “evidence having any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence.” *Id.* at 703.
188. *Id.*
189. *Id.; see also Redevelopment Agency v. Thrifty Oil Co.*, 5 Cal. Rptr. 2d 687, 689 n.9 (Cal. Ct. App. 1992) (stating that remediation issue was properly before jury as a property characteristic that affected value of land); *City of Olathe v. Stott*, 861 P.2d 1287, 1290 (Kan. 1993) (stating that evidence of contamination should be admissible in eminent domain proceeding because a purpose of an eminent domain proceeding is to determine fair market value of subject property and contamination affects that value).
190. 656 So.2d 921 (Fla. 1995).
191. *Id.* at 924.
compensation. The court supported its conclusion with an earlier decision holding that the public fear of power lines is relevant to the market value of property. The Finkelstein court listed several reasons why the stigma associated with contaminated property would affect its market value. For example, a buyer on the open market would consider the cost of remediating the contamination and the buyer would be subject to strict, joint, and several liability for contamination under environmental statutes. The contamination would subject the buyer to liability to the community. Finally, lenders would hesitate to finance the acquisition or improvement of contaminated property, particularly if the financing arrangement could subject the lender to liability for the contamination.

Recently, in Northeast Connecticut Economic Alliance, Inc. v. ATC Partnership, a municipal government condemned a contaminated parcel of property as part of a regional redevelopment plan, seeking to pay $1 in compensation due to the contaminated state of the property.

193. Finkelstein v. Dep’t of Transp., 656 So.2d 921, 922 (Fla. 1995).
194. See id. at 924:
Holding contamination to be relevant to the market value of property in an eminent domain valuation proceeding is consistent with our decision in Florida Power & Light Co. v. Jennings, 518 So.2d 895 (Fla. 1987), in which we held that “any factor including public fear which impacts on the market value of land taken for a public purpose may be considered to explain the basis for an expert’s valuation opinion.”


195. Finkelstein, 656 So.2d at 924. The court defined “stigma” as “the reduction in value caused by contamination resulting from the increased risk associated with the contaminated property. In sum, many prospective buyers are afraid of the financial risk associated with contaminated or even previously contaminated properties and would therefore pay less for the property.” Id. (internal quotation marks omitted).

196. Id.
197. Id.
198. Id. Ultimately, however, the Finkelstein court concluded that the evidence of contamination was not admissible in the case before it, as there was not a sufficient factual basis for the state expert’s valuation opinion. Id. at 925. Because there was no factual evidence in the record upon which to base the conclusion that the property’s value had actually been compromised by the alleged environmental contamination, the landowner was entitled to the fair market value of the property without respect to contamination. Id.

199. 776 A.2d 1068 (Conn. 2001).
200. Id. at 1072–73. The trial court had held that the contamination evidence was not admissible. N.E. Econ. Alliance, Inc. v. ATC P’ship, No. CV 940049248S, 1998 WL 197632, *15 (Conn. Super. April 16, 1998). During the pendency of the parties’ appeal, the Connecticut legislature amended its eminent domain statute to provide that evidence of environmental contamination is admissible to determine just compensation. ATC P’ship, 776 A.2d at 1076 (citing Public Acts 2000, No. 00-89,
Environmental Contamination and Eminent Domain

There, the court held that “evidence of environmental contamination and remediation costs is relevant to the valuation of real property taken by eminent domain and admissible in a condemnation proceeding to show the effect, if any, that those factors had on the fair market value of the property on the date of the taking.”201 The court reasoned that failing to admit the evidence would result in a “fictional” property value, because a purchaser of property on the open market would base the price it was willing to pay on a variety of factors, including (1) potential liability for the contamination, (2) “stigma” related to the property even after it has been remediated, (3) increased financing costs charged by lending institutions, and (4) the potential for increased regulation.202

IV. EVIDENCE OF ENVIRONMENTAL CONTAMINATION SHOULD NOT BE ADMISSIBLE TO DETERMINE JUST COMPENSATION IN WASHINGTON EMINENT DOMAIN PROCEEDINGS

Washington courts should not admit evidence of environmental contamination to determine just compensation in eminent domain proceedings. First, evidence of environmental contamination falls short of Washington’s rule prohibiting overly speculative evidence. Second, admitting evidence of environmental contamination in eminent domain proceedings deprives landowners of defenses and procedural safeguards under MTCA, thereby constituting a violation of procedural due process. Finally, admitting the evidence may effectively result in additional liability for the landowner and a windfall to the condemnor.

A. Evidence of Environmental Contamination Should Not Be Admitted To Determine Just Compensation Because It Contradicts Washington’s Prohibition Against Speculative Evidence in Eminent Domain Proceedings

Admitting evidence of environmental contamination for the purpose of valuing property in an eminent domain proceeding would require the

---

201. ATC P’ship, 776 A.2d at 1080.
202. Id. at 1081.
jury to engage in impermissible speculation. In determining just compensation in Washington eminent domain proceedings, the jury may not consider evidence that “is remote, imaginary, or speculative.” Juries may only consider those elements that have an actual effect on the property’s fair market value and are established by evidence. For example, the market value of a parcel of land with mineral content may not be determined by using an estimated market price for the mineral as extracted. This would entail speculation about, among other things, the costs of extracting and marketing the mineral and the nature and extent of demand for the mineral.

Similarly, devaluing property based on contamination would require the jury to speculate about the extent of contamination, the necessary responses, and the response costs. Under MTCA, liability for cleanup is based upon all costs which are “reasonably attributable” to remediation of the property. By definition, costs may not be attributed to remediation of the contamination until cleanup is complete, as the DOE expects to receive payment of remediation costs as they are incurred. In addition, cleanup may take years to complete and, consequently, may not be completed while an eminent domain proceeding is taking place. The full extent of contamination and necessary expenses may not be known until the cleanup is complete. Therefore, attempting to estimate the costs of cleanup before remediation is complete and the full extent of the contamination is known would involve a great deal of conjecture. Moreover, even if accurate evidence regarding the extent of the contamination is admitted, the jury would need to speculate to determine the effect of the contamination on the property’s value. As at least one commentator has noted, the potential range of devaluation due to

205. Id.
206. See supra notes 49–52 and accompanying text.
207. See id.
208. See supra notes 101–28 and accompanying text.
209. BUTLER, supra note 120, § 15.57.
210. See Robert I. McMurry, Treatment of Environmental Contamination in Eminent Domain Cases, 975 ALJ-ABA 237, 247 (1995) (stating that “consultants will concede that environmental investigations are more than a little like handicapping horse races: there is some science involved and some things we can put into statistics, but a big part of the equation is uncertain, unpredictable, incalculable, and perhaps even unknowable”).
contamination can be vast. Moreover, the average jury may not be equipped to handle the complicated technical information involved with environmental contamination and its effect on market value.

Use of the accepted valuation methodologies in Washington eminent domain proceedings does not make determining environmental contamination and its effect on market value in an eminent domain proceeding any less speculative. For example, under the income capitalization approach, the value of a plot of land is measured by projecting the amount of income it is likely to produce, subtracting future expenses, and capitalizing at an appropriate discount rate. Environmental contamination may theoretically be considered a future expense which is deducted from the projected income stream. However, even under this formula, the jury would be required to speculate with respect to the extent of environmental contamination, the necessary remediation measures, and the costs of cleanup.

The comparable sales approach is of little help. Under that approach, the value of a given piece of property is determined by comparing it with similar properties that have been sold within a reasonable period of time. In the case of contaminated property, it would be necessary to find another piece of property that is similarly contaminated and similar in other respects, such as size, location, and zoning. Since the character and extent of contamination tends to be unique to a given piece of property, it would be exceedingly difficult to locate similar property upon which to draw a comparison. Moreover, comparing contaminated property to uncontaminated property that is similar in other respects would certainly be speculative. As one Washington court has noted, even a comparison of undeveloped land with fully developed property to

---

211. 7A JULIUS SACKMAN, NICHOLS ON EMINENT DOMAIN § 13B.03[2][ii][iii] (3d Ed. Rev. 2000).
212. SACKMAN, supra note 211, § 13B.03[2][ii][iii]; Amy Grigham Boulris, Dealing with Contaminated Land from the Condemnee’s Perspective, C975 ALI-ABA 197, 203 (1995).
213. See SACKMAN, supra note 211, § 13B.04[2].
214. See SACKMAN, supra note 211, § 13B.04[2].
215. See SACKMAN, supra note 36, § 12.02[1], at 12-72.
216. See Aladdin, Inc. v. Black Hawk County, 562 N.W.2d 608, 616 (Iowa 1997) (“Properties with contamination are hard to compare because they involve multiple varieties of contamination of varying concentrations and require assorted methods of cleanup. The commission, judge, or jury required to determine ‘just compensation’ would likely be compelled to speculate as to the damages.”). See also Robert I. McMurry & David H. Pierce, Environmental Contamination and Its Effect on Eminent Domain, C791 ALI-ABA 133, 162. Although locating similar property may not, in all cases, be impossible, considerations of judicial economy suggest it would be inefficient to always attempt to do so.
estimate market value leads to impermissible speculation and conjecture.\textsuperscript{217} As that court declared, a court “cannot be too careful in excluding evidence of this character.”\textsuperscript{218}

The depreciated replacement cost method is similarly unavailing. Under that approach, just compensation is measured by determining the replacement costs of improvements on the property, subtracting depreciation, and adding the market value of the underlying property.\textsuperscript{219} Such an approach simply begs the question: once the cost of improvements and the amount of depreciation have been established, the market value of the property itself must still be determined. Presumably, this will require recourse to the methodologies previously discussed.\textsuperscript{220} Thus, the depreciated replacement cost method falls short, for the same reasons that the income capitalization and comparable sales approaches fail.\textsuperscript{221}

Some jurisdictions that admit evidence of environmental contamination to determine just compensation have much more liberal standards than Washington for admitting speculative evidence.\textsuperscript{222} Therefore, they do not provide persuasive precedent for admitting such evidence in Washington. For example, Florida courts will consider evidence of the public’s fear of a condition on land to determine its market value.\textsuperscript{223} Based in part upon that reasoning, the \textit{Finkelstein} court was willing to admit evidence of environmental contamination in an eminent domain proceeding, stating that doing so is consistent with the notion that “any factor including public fear which impacts on the market value of land taken for a public purpose may be considered . . . .”\textsuperscript{224} However, Washington has eschewed the admissibility of speculative

\textsuperscript{217} \textit{In re City of Medina}, 69 Wash. 2d 574, 578, 418 P.2d 1020, 1022–23 (1966).
\textsuperscript{218} \textit{Id}.
\textsuperscript{219} See \textit{Sackman}, supra note 36, § 12.02[1].
\textsuperscript{220} See supra notes 213–18 and accompanying text.
\textsuperscript{221} For a discussion of novel valuation methodologies that may alleviate the problem to some degree, see McMurry & Pierce, supra note 216, at 165–69 (suggesting that these novel alternatives are still somewhat unsatisfactory); see also \textit{Sackman}, supra note 211, § 13B.04[2] (listing newer valuation methods but noting that “[t]he development of techniques for valuing contaminated properties is still in its infancy”) (internal quotation marks omitted).
\textsuperscript{222} See, e.g., \textit{Finkelstein v. Dep’t of Transp.}, 656 So.2d 921, 924 (Fla. 1995) (citing Fla. Power & Light Co. v. Jennings, 518 So.2d 895 (Fla. 1987)).
\textsuperscript{223} See supra notes 193–94 and accompanying text.
\textsuperscript{224} See supra notes 193–94 and accompanying text.
Environmental Contamination and Eminent Domain

evidence like public fear. Thus, one of the primary rationales relied upon by jurisdictions admitting evidence of environmental contamination in eminent domain proceedings is inapplicable in Washington.

B. Admitting Evidence of Environmental Contamination To Determine Just Compensation in an Eminent Domain Proceeding Violates the Procedural Due Process Rights of Landowners

Admitting evidence of environmental contamination to determine just compensation would circumvent the procedural safeguards established by MTCA and, therefore, constitute a violation of procedural due process. To determine whether governmental action constitutes a violation of procedural due process, a court must first determine whether a party has been deprived of a property interest. Upon finding such a deprivation, the court must then determine the level of process that is due. Applying the balancing test set forth in *Mathews v. Eldridge*, admission of evidence of environmental contamination to determine just compensation will result in the mistaken deprivation of the landowner’s property interests with no significant concomitant governmental benefit and is therefore a violation of the landowner’s right to procedural due process.

Reducing just compensation in an eminent domain proceeding due to environmental contamination deprives a landowner of an important property interest. One of the fundamental principles behind just compensation in eminent domain is the requirement that the property owner be put in the same monetary position as she would have occupied had the property not been taken. Payment of anything less than full compensation divests the landowner of that important interest.

Under the *Mathews* balancing test, a court must balance (1) the private interests that are affected by a governmental action, (2) the risk that the procedures employed by the governmental actor will result in the mistaken deprivation of those interests, and (3) the government’s own

---


226. *See supra* note 144 and accompanying text.

227. *See supra* note 148 and accompanying text.


229. *See supra* note 33 and accompanying text.
interests, including the economic and administrative burden that the procedural protection would entail. Reducing just compensation due to environmental contamination in an eminent domain proceeding deprives the landowner of full compensation, without affording the landowner the procedural safeguards outlined in MTCA. MTCA provides a thorough procedural framework for investigating the existence of contamination and assigning responsibility for any contamination discovered. It provides the landowner with numerous options for mitigating the contamination, offers certain defenses to liability, affords the right to seek contribution from other PLPs, and enables the landowner to bring a private right of action by remediating the property according to DOE standards. An eminent domain proceeding provides the landowner with none of these procedural safeguards and, thus, may not be a level of process that is appropriate to the nature of the interests being divested.

At the same time, preventing the governmental entity from paying reduced compensation on account of contamination does not deprive it of its interests. On the contrary, whatever interests the government has—for example, remediating contaminated property and holding landowners responsible for the contamination—is amply addressed by MTCA. In addition, any financial loss sustained by the condemnor in paying full value for contaminated property can be amply redressed by forcing the PLPs to remediate the contamination under MTCA. If the proper procedures are employed under MTCA, the government wins by recovering cleanup expenses in an action against the PLP, and the landowner wins by maintaining her procedural guarantees.

Courts in other jurisdictions have come to similar conclusions. For example, Department of Transportation v. Parr involved an environmental statute similar to MTCA in that it assigned liability for environmental contamination while providing landowners with certain procedural safeguards. There, the court concluded that adjudicating

230. See supra notes 145–47 and accompanying text.
231. See supra Part II.B.
232. See supra Part II.A.
233. See supra note 70 and accompanying text.
234. See supra notes 129–35 and accompanying text.
235. See supra notes 149–55 and accompanying text.
238. Id. at 22–23.
liability for environmental contamination in an eminent domain proceeding would constitute a procedural due process violation, because it would circumvent the rights, defenses, and procedural safeguards established by the statute.\footnote{Id.} Because the same holds true under Washington law, Washington should adopt the \textit{Parr} court’s reasoning and preclude consideration of environmental contamination to determine just compensation.

\textbf{C. Evidence of Environmental Contamination Should Not Be Admitted in an Eminent Domain Proceeding Because It May Result in a Windfall for the Condemnor and Create Additional Liability for the Landowner}

Where the landowner is a PLP under MTCA, admitting evidence of contamination to determine just compensation potentially subjects the landowner to additional liability. First, the landowner receives a reduced price for the property in an eminent domain proceeding due to a determination that the contamination affects its fair market value. Then, the landowner may be subject to liability in a MTCA action and may consequently be forced to pay remediation costs for the contamination. The landowner is forced to sell the land at a price reduced by environmental contamination and subsequently required to pay for the cleanup of the same contamination. Therefore, the landowner is effectively subject to additional liability in the amount that the just compensation is reduced.

Although additional liability may also result if the landowner simply sells the property on the open market and is later held liable for remedial action under MTCA, there are important distinctions between a transaction on the open market and an eminent domain action. No one is forcing the landowner to sell in a transaction on the open market. The seller may prefer to remediate the property on her own and sell at a higher price, rather than selling at a reduced price due to the contamination and later being held liable under MTCA. Moreover, if no buyer is willing to pay what the landowner considers a fair price, the landowner may find it more economically efficient to simply keep the property and take her chances with future liability.

Furthermore, the landowner can bargain away future MTCA liability if permitted to sell the land on the open market. For example, the
landowner may agree to sell the land at a discounted rate in exchange for the buyer’s promise to indemnify the landowner against any future MTCA liability. In this way, the landowner avoids additional MTCA liability because, although the seller may receive a reduced rate due to the contamination, he or she can avoid future liability under MTCA. Similarly, the parties to a voluntary sale may agree to mitigate potential MTCA liability by taking independent action to remediate environmental contamination. In an eminent domain action, however, the landowner is deprived of the bargaining power to negotiate such a trade.

Similarly, admitting evidence of environmental contamination to determine just compensation may effectively give the condemning authority a windfall. First, the condemning authority acquires the property at a reduced price. Then, once the contamination has been cleaned, the condemnor may subsequently hold the landowner responsible for remediation costs under MTCA. Not only is the state Attorney General entitled to recover its remediation costs from the PLP, but any governmental entity that may condemn property by eminent domain may also bring a private right of action under MTCA. In effect, the condemning authority purchases the property at a price adjusted for contamination, but ultimately receives a piece of property free of contamination. Thus, the condemnor receives a windfall in the amount that the just compensation is reduced for contamination.

V. CONCLUSION

Evidence of environmental contamination should not be admissible to determine just compensation in a Washington eminent domain proceeding. Although it is true that environmental contamination may have an impact on the market value of land, admitting this evidence contravenes Washington’s prohibition against speculative evidence to determine just compensation, violates the procedural due process rights


242. See supra notes 129–35 and accompanying text.
of landowners, and may result in a windfall for the condemnor and additional liability for the landowner. There is no doubt that the cleanup of contaminated property should be a public priority of the highest order, for the well-being of Washington and for the planet as a whole. However, this interest must be balanced against the interest citizens have in receiving just compensation when their land is taken by the government. Washington has established a comprehensive and effective scheme for addressing the problem of environmental contamination, which provides for the landowners’ interests as well as the public interest. The problem of environmental contamination should be addressed under that scheme—condemnors should not have the option of circumventing MTCA in order to obtain property at fire sale prices.
THE RELIGIOUS LAND USE AND INSTITUTIONALIZED PERSONS ACT: AN ANALYSIS UNDER THE COMMERCE CLAUSE

Evan M. Shapiro

Abstract: Congress based the Religious Land Use and Institutionalized Persons Act (RLUIPA) on accumulated evidence suggesting that the land use decisions of local governments unfairly burden religious uses. The RLUIPA is narrower in scope than two previous statutes aimed at protecting religious liberty. The United States Supreme Court held the first of these religious liberty statutes unconstitutional, and Congress failed to enact the other. This Comment examines the constitutionality of the RLUIPA under the Commerce Clause and argues that Congress exceeded its Commerce Clause authority because (1) land use regulation does not constitute “economic activity” as defined by the United States Supreme Court in United States v. Lopez and United States v. Morrison and (2) land use regulation is insufficiently connected to interstate commerce.

Congress has recognized that religious institutions are often overly burdened by local land use decisions. As a remedy, Congress passed the Religious Land Use and Institutionalized Persons Act (RLUIPA) to prevent local governments from burdening religious institutions with their land use decisions absent a compelling governmental interest. For example, a local ordinance might limit the size of houses of worship in commercial or residential zones. Congress determined that such zoning laws often facially discriminate against religion, and that zoning boards may apply neutral zoning laws discriminatorily. The RLUIPA prohibits local governments from substantially burdening religion with their land use determinations unless the local regulation furthers a compelling governmental interest and is the least restrictive means to accomplish that interest.

Congress’s authority to enact legislation is limited and must be based on a constitutionally granted power. Congress based its authority to

1. 42 U.S.C. §§ 2000cc to cc-5 (2000). The RLUIPA also protects institutionalized persons’ right to freely exercise their religion. See id. However, this Comment discusses only the land use portion of the law.
2. See Eric Pryne, Fimia Crafting Own Plan To Limit Rural Churches, SEATTLE TIMES, June 2, 2001, at B1.
4. Id. § 2000cc(a)(1).
enact the RLUIPA on several constitutional provisions, including the Commerce Clause. The Commerce Clause grants Congress the power to regulate commerce among the states. Although the Supreme Court has struck down previous legislation aimed at protecting religious liberty, the RLUIPA is much more limited in scope.

This Comment argues that Congress unconstitutionally exceeded its Commerce Clause power by enacting the land use portion of the RLUIPA. Congress has authority under the Commerce Clause to regulate those activities that are economic in nature and that substantially affect interstate commerce. The regulated activity at issue in the RLUIPA is land use regulation, which does not constitute “economic activity” as recently defined by the Court. When enacting the RLUIPA, Congress additionally failed to satisfy the elements the Court has set forth to ensure that the regulated activity has substantial effects on interstate commerce.

Part I of this Comment discusses the relationship between land use and religion, and explains local governments’ justifications for excluding religious uses from certain zones. Part II details the provisions of the RLUIPA, its legislative history, and previous statutes. Part III describes the U.S. Supreme Court’s past and current interpretations of Congress’s Commerce Clause power. Part IV argues that Congress exceeded its authority in enacting the RLUIPA because land use regulation does not constitute 

7. Id.
8. U.S. CONST. art. I, § 8, cl. 3.
11. Although this Comment contends that Congress did not have authority to enact the RLUIPA under its Commerce Clause power, Congress supported the RLUIPA through other constitutional powers such as the Spending Clause and Section Five of the Fourteenth Amendment. 146 CONG. REC. S7774 (exhibit 1). Thus, even if, as argued in this Comment, Congress unconstitutionally exceeded its Commerce Clause powers, the statute may be deemed constitutional under these other powers. However, such an analysis is beyond the scope of this Comment.
13. See infra Part IV.A.
14. See infra Part IV.B.
Religious Liberty and the Commerce Clause

relate to an economic endeavor or commercial transaction and is insufficiently connected to interstate commerce.

I. LAND USE REGULATION AND RELIGION: THE NATURE OF THE PROBLEM

Local governments historically have based their land use decisions on considerations such as public safety, aesthetics, and economics. Religious uses often conflict with these concerns. In fact, courts have often upheld local governments’ justifications for excluding religious uses in both residential and industrial zones. To prevent exclusion of religious uses, Congress promulgated the RLUIPA to regulate local governments’ land use determinations.

A. The Nature of Land Use Regulation

The growth of land use regulation across the country was spurred by concerns of public safety and aesthetics. In 1922, the United States Department of Commerce promulgated the Standard State Zoning Enabling Act (SZEA). The SZEA authorized municipalities to establish zoning districts in which compatible uses are grouped together and incompatible ones are excluded. Section Three of the SZEA sets out the purposes of zoning: to ensure safety from fire and other dangers, provide adequate air and light, lessen street congestion, promote health and general welfare, regulate land use intensity and population density, and


facilitate public services. Additionally, a local government should regulate with reasonable consideration of the district’s characteristics and with the aim of preserving the value of buildings.

The United States Supreme Court has recognized the beneficial purposes of zoning systems. In Village of Euclid v. Ambler Realty Co., the Court upheld segregation of residential, business, and industrial uses. The Court expressed that the benefits of zoning include an increase in the safety and security of home life, fewer street accidents due to reduced traffic, lower noise levels in residential neighborhoods, and a better environment in which to raise children.

The Court continues to affirm the validity of zoning regulations based on the rationale expressed in Euclid. For example, in Village of Belle Terre v. Boraas, the Court upheld a zoning ordinance limiting the occupancy of single-family houses to traditional families or to groups including only two unrelated persons. The Court stated that government may create zones for the purpose of providing families and youth with secluded, clean areas of sanctuary.

B. Justifications for Land Use Regulation of Religious Uses

Courts have upheld local governments’ decisions to exclude religious uses from both residential and industrial zones. These cases illustrate the nature and purpose of land use determinations. For residential areas, zoning ordinances and decisions are often based on aesthetic, nuisance-

25. See id. at 397; MANDELKER, supra note 15, § 2.40, at 56.
29. Id. at 8–10.
30. Id. at 9.
31. See supra note 16. Exclusion of religious uses from entire zones raises First Amendment issues, a discussion of which is outside the scope of this Comment.
type, and property-value concerns. In commercial and industrial areas, economic considerations may impact zoning ordinances and decisions.

1. Residential Areas

Reasons for excluding religious uses from residential zones mirror some of the purposes of land use regulation expressed in the SZEAs. Perceived detrimental effects of religious uses on municipal services, property values, and traffic congestion often explain local governments’ decisions. In *Corp. of Presiding Bishop of Church of Jesus Christ of Latter-Day Saints v. City of Porterville*, the California Court of Appeals held that a land use ordinance excluding religious uses from residential zones based on factors such as potential noise, traffic, and parking problems constituted a proper exercise of a city’s zoning power. Likewise, the local government in *Seward Chapel, Inc. v. City of Seward* was justified in implementing an ordinance excluding religious schools from residential zones based on concerns of excess traffic, noise, and other nuisances.

Local ordinances dedicated to preserving landmarks with historical or architectural value also may conflict with religious uses. Historical preservation laws prohibit demolition of historic buildings and require a certain level of maintenance. The United States Supreme Court’s recent decisions indicate that historical preservation laws do not violate the Free

---


34. See, e.g., Lakewood Congregation, 699 F.2d at 305, 308; Seward Chapel, 655 P.2d at 1297–99; Porterville, 203 P.2d at 825; Lutheran High School Ass’n, 381 N.W.2d at 419, 421.

35. MANDELKER, supra note 15, § 5.58, at 187.

36. 203 P.2d 823 (6th Cir. 1983).

37. Id. at 825.

38. 655 P.2d 1293 (Alaska 1982).

39. Id. at 1297–99.

40. MANDELKER, supra note 15, § 11.24, at 462.
Exercise Clause. Courts have held that local governments are justified both in excluding new churches from residential zones and in preventing renovations of historical church buildings already located in such zones.

2. Commercial and Industrial Zones

In industrial and commercial areas, safety and aesthetic concerns may be secondary to economic considerations. For example, in *City of Chicago Heights v. Living Word Outreach Full Gospel Church and Ministries, Inc.*, the local government adopted a city ordinance that required religious institutions to apply for a special use permit to locate a house of worship in a commercial zone. An Illinois Court of Appeals accepted the city’s rationale for adopting the ordinance, which was intended to reinvigorate falling revenues and provide a stable tax base, and denying the special use permit. The court stated that the city’s effort to designate certain areas for commercial activity encouraged economic growth and stability.

41. See MANDELKER, supra note 15, § 11.35, at 471–72. Recently, the Court denied certiorari in *Rector, Wardens & Members of Vestry of St. Bartholomew’s Church v. City of New York*, 914 F.2d 348 (2d Cir. 1990), cert. denied, 499 U.S. 905 (1991), where the issue was whether a New York City ordinance could prevent a church from replacing a church-owned building with an office tower. Id. at 350. The court of appeals held that the ordinance did not violate the Free Exercise Clause of the U.S. Constitution. Id. at 353–56. On the same day, the Supreme Court remanded a Washington State Supreme Court decision that held that a historical landmark law interfered with the practice of religion. *First Covenant Church of Seattle v. City of Seattle*, 114 Wash.2d 392, 787 P.2d 1352 (Wash. 1990), vacated and remanded, 499 U.S. 901 (1990).

42. See supra notes 37, 39, 41 and accompanying text.

43. See, e.g., *Cornerstone Bible Church v. City of Hastings*, 740 F. Supp. 654, 661 (D. Minn. 1990); *City of Chi. Heights v. Living Word Outreach Full Gospel Church and Ministries, Inc.*, 707 N.E.2d 53, 59 (Ill. App. 1999), rev’d in part, 749 N.E.2d 916, 932 (Ill. 2001) (holding that denial of special use permit was improper because it conflicted with the zoning ordinance’s legislative history stating that churches are compatible with other uses in the commercial district).


45. A special use permit is necessary in situations in which a zoning ordinance authorizes a particular use in a district only with the approval of a local zoning board or agency. DANIEL R. MANDELKER ET AL., PLANNING AND CONTROL OF LAND DEVELOPMENT 473 (4th ed. 1995).


47. Id. at 59. In addition to the ordinance’s valid purpose of maintaining a stable economy within the industrial zone, the Court held that “public health, safety and morals are served by the ordinance.” Id.

48. Id.
A Federal District Court in Minnesota upheld a similar ordinance in *Cornerstone Bible Church v. City of Hastings.* The court validated an ordinance because its purpose was to allow the city to develop the downtown area commercially and set aside space for industrial uses, while preserving the quality of the residential areas. These cases illustrate that local governments’ purposes for excluding religious uses often include economic considerations as well as nuisance-type concerns.

C. Impact of Land Use Regulation on Religion

Local governments’ land use regulations impact both the construction of a religious facility in a zone and the use of a facility already located in a zone. In addition, religious uses may conflict with lot and building size regulations. This is particularly common when the religious institution desires to build a “megachurch” or construct several buildings in one location.

Land use regulation may result in the outright exclusion of religious uses from entire zones. Although churches that have been in residential communities prior to a zoning change would not be threatened by outright removal, new churches often are unable to locate in residential neighborhoods. In most states, local governments may not exclude churches from residential areas; however, at least one commentator has asserted that the actual effect of land use regulation is much less favorable than this general rule would suggest. Although most land use cases relate to religious uses in residential zones, more recent cases also

50. Id. at 661.
52. Id.
55. See Douglas Laycock, *State RFRAs and Land Use Regulation,* 32 U.C. DAVIS L. REV. 755, 763–64 (1999). “The actual experience of many churches is more in line with the hostile federal cases than with the more encouraging summaries of state zoning doctrine. . . . Commentators writing from the land use perspective share my sense that the climate has changed and that churches now face less sympathetic regulation.” Id. at 764 (internal citations omitted).
56. Id. at 763–64.
have involved religious uses in industrial and commercial zones. 57 It is
often a challenge to locate religious facilities in these zones as well. 58

Local government action that results in the outright exclusion of
religious uses from a zone generally takes one of two forms: (1) the
zoning code on its face excludes religious uses from certain zones or (2)
zoning boards refuse to grant permits for religious uses. 59 One expert
claims that zoning codes, by design, frequently treat religious uses less
favorably than secular uses. 60 For example, a land use attorney’s survey
of Chicago-area land use laws found evidence that Chicago’s zoning
codes facially discriminate against religion. 61 In several zoning codes,
there was no place a church could locate within a suburb without a
special use permit. 62

Furthermore, Congress has found evidence that zoning boards
discriminate in their application of facially neutral zoning laws. 63 Land
use zoning boards have a significant amount of discretion in applying
zoning laws. 64 Congress has explained that discrimination against
religious uses often “lurks behind such vague and universally applicable
reasons as traffic, aesthetics, or ‘[inconsistency] with the city’s land use
plan.” 65 The outright exclusion of religious uses from both residential
and commercial zones often is due to zoning codes that facially exclude
religious uses or a zoning board’s application of neutral laws. 66

57. Id. at 761.
59. See id. at 18–24; Laycock, supra note 55, at 764–65, 768, 773–74, 777–83; 146 CONG. REC.
60. Laycock, supra note 55, at 776.
61. Laycock, supra note 55, at 773–74 (citing Compilation of Zoning Provisions Affecting
Churches in 29 Suburbs of Northern Cook County by John W. Mauck [as] of 7-10-98, Based Upon
1995 Published Standards, attached to statement of John Mauck, partner, Mauck, Ballande, Baker &
62. Laycock, supra note 55, at 773.
63. See H.R. REP. NO. 106-219, at 18; 146 CONG. REC. S7774 (exhibit 1).
64. Laycock, supra note 55, at 764–65.
65. 146 CONG. REC. S7774 (exhibit 1).
66. See supra notes 59–65 and accompanying text.
II. RECENT CONGRESSIONAL ATTEMPTS TO PROTECT RELIGIOUS LIBERTY

Since 1993, Congress has considered three bills that attempted to protect religious liberty. The first was the Religious Freedom and Restoration Act of 1993 (RFRA), which the U.S. Supreme Court held to be unconstitutional in 1997. The second, the Religious Liberty Protection Act of 1999 (RLPA), failed to pass in the Senate in 1999. Finally, in 2000, Congress enacted the RLUIPA, which is narrower in scope than the two previous acts, focusing only on local land use regulation and institutionalized persons.

A. The Religious Freedom Restoration Act: An Initial Attempt To Restore Protection to Religion

Congress passed the RFRA in 1993 in response to the Court’s decision in Employment Division v. Smith, which held that generally applicable laws that burdened religion were not subject to a strict scrutiny standard. Under a strict scrutiny standard, a government would be required to show a “compelling governmental interest” to justify zoning decisions that negatively impact religious uses. The RFRA was enacted to restore protection to religion by requiring that governments refrain from substantially burdening religion absent a compelling state interest. The U.S. Supreme Court in City of Boerne v. Flores struck down the RFRA, at least as it applied to state and local governments, by

72. See supra, note 10.
75. See Employment Div., 494 U.S. at 876–82; Conkle, supra note 74, at 637.
79. The Court did not address the validity of the RFRA with regard to federal laws and practices. Conkle, supra note 74, at 633 n.5.
holding that Congress exceeded its power under the Fourteenth Amendment.80 As a result, the rule in Smith persisted: generally applicable, religion-neutral state or local laws could burden religiously-motivated conduct without triggering heightened judicial scrutiny.81

B. The Religious Liberty Protection Act: An Attempt To Pass Constitutional Muster with Additional Constitutional Authority

In response to the Court’s holding that the RFRA was unconstitutional, some members of Congress attempted to adopt new legislation that would survive a constitutional challenge and protect religious liberty.82 The goal of the new legislation (RLPA) mirrored that of the RFRA,83 but the new act relied on additional constitutional authority.84 The Flores decision led Congress to believe that its ability to enact religious liberty bills was limited to its Spending Clause power, its Commerce Clause power, and its power to remedy states’ constitutional violations under the Fourteenth Amendment.85 The drafters used all of these constitutional powers to justify their authority to enact and implement the RLPA.86

The validity of the drafters’ constitutional justifications was never tested because Congress failed to pass the RLPA, due in part to fears that

80. See Flores, 521 U.S. at 516–36. Section One of the Fourteenth Amendment states, in part, that [n]o State shall make or enforce any law which shall abridge the privileges or immunities of citizens of the United States; nor shall any State deprive any person of life, liberty, or property, without due process of law; nor deny to any person within its jurisdiction the equal protection of the laws.

U.S. Const. amend. XIV, § 1. Section Five of the Fourteenth Amendment states that “Congress shall have power to enforce by appropriate legislation, the provisions of this article.” U.S. Const. amend. XIV, § 5.

81. See Conkle, supra note 74, at 633.

82. During the 105th Congress, the Subcommittee on the Constitution held five hearings on the RLPA. Congress took no further action in 1998, but in the 106th Congress, the Subcommittee on the Constitution held a hearing and two markup sessions on the RLPA. H.R. Rep. No. 106-219, at 5 (1999).


85. Id. at 12.

86. Id. at 12–18.
its protection was too expansive. Many Senators were concerned that the RLPA could violate various citizens' civil rights. For example, one could foresee conflict between a state’s interest in eliminating sexual discrimination and the freedom of religious institutions to adhere to practices of limiting eligibility for ordination into the clergy based on sex. Therefore, although the House of Representatives passed the RLPA overwhelmingly, it was never enacted into law.

C. The Religious Land Use and Institutionalized Persons Act: A Narrower Statute for Religious Protection

After the Senate rejected the RLPA, Congress focused on drafting a new statute that would provide much more limited protection to religion. The RLUIPA prohibits state and local governments from imposing land use regulations that substantially burden the religious exercise of a person or institution unless the government demonstrates that the regulation serves a compelling governmental interest and is the least restrictive means of furthering that interest.

Although the RLUIPA codifies the same strict scrutiny standard found in the RFRA and the RLPA, the RLUIPA has a more narrow scope. Whereas the RFRA and the RLPA applied to any state action that offended the free exercise of religion, the RLUIPA only addresses burdens imposed on religion through land use decisions and through regulation of persons housed in state institutions. Thus, Congress minimized potential conflicts with some civil rights. The American Civil Liberties Union and other groups who had expressed concerns about the RLPA’s impact on civil rights supported the RLUIPA.

89. See id. at 14.
92. Id. § 2000cc(a)(1).
94. Id. §§ 2000cc to cc-5.
1. **RLUIPA’s Purpose: To Protect the Right To Gather and Worship**

   The land use provision of the RLUIPA is intended to protect the right to “gather and worship.” Although the legislative history behind the RLUIPA is limited because it passed both houses without committee action, speeches in both the House and the Senate provide some background. Senators Hatch and Kennedy indicated that the RLUIPA targets land use regulation because local land use decisions frequently burden religious liberty. As justification for the RLUIPA’s land use provision, Senators Hatch and Kennedy referred to evidence showing that zoning codes facially discriminate against churches and that zoning boards frequently apply neutral regulations in a discriminatory manner.

2. **Congress’s Constitutional Authority for Enacting the RLUIPA**

   Legislative history and the express language of the statute reveal Congress’s determination that it had the constitutional authority to enact the RLUIPA based in part on the Commerce Clause. Under the Commerce Clause, Congress can “regulate Commerce with foreign Nations, and among the several States, and with the Indian Tribes.” The RLUIPA states that the statute applies in any case in which “the substantial burden affects, or removal of that substantial burden would affect, commerce with foreign nations, among the several States, or with

---

100. 146 CONG. REC. S7774–95 (exhibit 1).
101. 146 CONG REC. S7774–75 (exhibit 1). Support for this conclusion took the form of anecdotes and statistics from national surveys, summarized in the report of the House Committee on the Judiciary regarding the RLPA. H.R. REP. 106-219, at 18–24 (1999); see also Issues Relating to Religious Liberty Protection, and Focusing on the Constitutionality of a Religious Protection Measure: Hearing Before the Senate Committee on the Judiciary, 106th Cong. 72-101 (1999) (statement of Douglas Laycock, Alice McKean Young Regents Chair in Law, University of Texas School of Law); Laycock, supra note 55, at 769–83.
103. U.S. CONST. art. I, § 8, cl. 3.
Religious Liberty and the Commerce Clause

Indian tribes, even if the burden results from a rule of general applicability.”

Congress determined that land use regulations that burden religion substantially affect interstate commerce, basing this finding primarily on the legislative history of the RLPA. Congress relied on Marc Stern’s testimony to the House Subcommittee on the Constitution regarding the RLPA on June 16, 1998. His comments were aimed at demonstrating that religion, as a whole, is a major economic factor in the United States. In his written statement, he cited statistics regarding property controlled by religious institutions, noting that these institutions spent $6 billion in 1992 on capital improvements and construction. In New York and Wisconsin, the value of property controlled by religious institutions was $17.1 billion and $5 billion, respectively. He also opined, without citing statistics, that land use regulation limiting the ability of religious institutions to build or expand affects interstate movement of goods and services.

After two failed attempts to implement a broader religious freedom statute, Congress adopted the more narrow statute focused on local governments’ land use regulation. One of the primary objectives of the RLUIPA is to prevent local governments from substantially burdening religion with their land use regulations. Congress asserted that it had the authority to adopt the RLUIPA based, in part, on the Commerce Clause because it believed land use regulation substantially affects interstate commerce.

106. Id. Mr. Stern, the Director of the Legal Department of the American Jewish Congress, provided oral testimony and a prepared written statement to the House Subcommittee. See Religious Liberty Protection Act of 1998: Hearing on H.R. 4019 Before the Subcommittee on the Constitution of the Committee on the Judiciary, 105th Cong. 54–65 (1998) [hereinafter 1998 RLPA] (statement of Marc Stern, Director, Legal Department, American Jewish Congress).
110. 1998 RLPA, supra note 106, at 58.
112. Id. § 2000cc(a)(1).
III. THE COMMERCE CLAUSE: A LIMITATION ON
CONGRESS’S AUTHORITY TO REGULATE INTRASTATE
ACTIVITY

Congress must base its authority to regulate local governments’ land
use decisions on a constitutionally granted power.\textsuperscript{114} Since 1824, the
U.S. Supreme Court has limited to varying degrees Congress’s power
under the Commerce Clause.\textsuperscript{115} In 1995, in \textit{United States v. Lopez},\textsuperscript{116} the
Court returned to a narrow interpretation of Congress’s Commerce
Clause power, allowing regulation only of those activities that are
economic in nature and that have a substantial effect on interstate
commerce.\textsuperscript{117} This narrow interpretation was reinforced in 2000, in
\textit{United States v. Morrison}.\textsuperscript{118}

A. \textit{Pre-Lopez} Interpretation of the Commerce Clause

\textit{Gibbons v. Ogden},\textsuperscript{119} decided in 1824, was the first U.S. Supreme
Court case to define “commerce.”\textsuperscript{120} In \textit{Gibbons}, Justice Marshall
expressed a broad notion of commerce, stating that it represents more
than just “buying and selling, or the interchange of commodities.”\textsuperscript{121} The
Court defined commerce as “the commercial intercourse between nations
and parts of nations,”\textsuperscript{122} but also reasoned that, although the commercial
activity must affect commerce between the states, commerce power
could reach within the borders of a state.\textsuperscript{123}

After 1888, the Court began narrowing its interpretation of the
Commerce Clause. At the turn of the century, the Court invoked a
formulistic approach for invalidating federal social and economic

\textsuperscript{114} Congress may exercise only those powers granted to it in the U.S. Constitution. United

\textsuperscript{115} Compare, e.g., Gibbons v. Ogden, 22 U.S. (9 Wheat.) 1 (1824) with NLRB v. Jones &
Laughlin Steel Corp., 301 U.S. 1 (1937).

\textsuperscript{116} 514 U.S. 549 (1995).

\textsuperscript{117} Id. at 559–68.

\textsuperscript{118} 529 U.S. 598, 607–19 (2000).

\textsuperscript{119} 22 U.S. (9 Wheat.) 1 (1824).

\textsuperscript{120} See Lopez, 514 U.S. at 553; Heart of Atlanta Motel, Inc. v. United States, 379 U.S. 241, 253–
54 (1964).

\textsuperscript{121} Gibbons, 22 U.S. (9 Wheat) at 189.

\textsuperscript{122} Id. at 189–90.

\textsuperscript{123} Id. at 194.
Religious Liberty and the Commerce Clause

regulations. It began distinguishing between regulations “directly” and “indirectly” affecting interstate commerce; those regulations only “indirectly” affecting commerce exceeded Congress’s federal powers. For example, the Court held that regulations relating to “manufacturing” were invalid because they only indirectly affected commerce.

The Court’s approach to the Commerce Clause changed dramatically in NLRB v. Jones & Laughlin Steel Corp. Specifically, the Jones & Laughlin Court addressed the constitutionality of the National Labor Relations Act of 1935 (NLRA), which established a comprehensive system for regulating labor relations in all industries affecting interstate commerce. The Court abandoned its previous distinction between “indirect” and “direct” effects on interstate commerce, and instead adopted a “substantial relations” test. The Court stated that although activities—when considered separately—may be intrastate in nature, Congress may regulate the activities “if they have such a close and substantial relation to interstate commerce that their control” is necessary to protect commerce from interference.

The Court continued its broad view of the Commerce Clause in Wickard v. Filburn. Although the activity at issue was an individual’s production of wheat for home consumption, the Court determined that it fell under Congress’s Commerce power. The Court explained that it did not matter that Filburn’s effect on the price of wheat was trivial...
because his contribution to the market “taken together with that of many others similarly situated, [was] far from trivial.”  

As a result, the Court vastly expanded Congress’s Commerce Clause power during the period beginning in 1935 with *Jones & Laughlin* and ending in 1995 with *Lopez*. The Court overwhelmingly deferred to Congress’s judgment regarding which activities fell under its Commerce Clause power. In fact, in the fifty-three years following *Wickard*, every statute Congress enacted under that power passed judicial review.

### B. The Court’s Current Interpretation of the Commerce Clause

In 1995, the Supreme Court substantially narrowed its view of Congress’s Commerce Clause power by holding in *United States v. Lopez* that Congress could only reach economic activities that have a substantial effect on interstate commerce. The defendant in *Lopez* was convicted for possessing a firearm in a school zone in violation of the Gun-Free School Zones Act of 1990 (GFSZA), which made it a federal offense “for any individual knowingly to possess a firearm at a place that the individual knows, or has reasonable cause to believe, is a school zone.” The defendant argued that Congress lacked authority to pass the GFSZA and the Supreme Court agreed.

In *United States v. Morrison*, the Court continued on to narrow Congress’s Commerce Clause power. The Court held, by a five to four vote, that Congress exceeded its Commerce Clause authority by establishing a civil remedy under Section 13,981 of the Violence Against Women Act of 1994 (VAWA). The act stated that “[a]ll persons within the United States shall have the right to be free from

---

134. *Id.* at 127–28 (describing the “aggregate” principle).
138. *Id.* at 559–68.
140. *Id.* § 922(q)(2)(A).
144. *See Morrison*, 529 U.S. at 627.
Religious Liberty and the Commerce Clause

crimes of violence motivated by gender.”\textsuperscript{145} Subsection (c) stated that anyone who commits a crime of violence motivated by gender is liable to the injured party for damages, injunctive, and declaratory relief.\textsuperscript{146} The defendants successfully argued that Congress lacked the authority to pass this act.\textsuperscript{147}

The Court in \textit{Lopez} and \textit{Morrison} established a new framework to evaluate the validity of statutes enacted under Congress’s Commerce Clause power. First, the regulated activity must be economic in nature.\textsuperscript{148} Second, the Court set forth the following three elements to guide lower courts in determining whether the activity has substantial interstate effects: (1) whether Congress accumulated findings indicating that the regulated activity has interstate effects, (2) the strength of the nexus between the regulated activity and interstate commerce, and (3) whether the statute contains a jurisdictional element.\textsuperscript{149}

1. \textit{Definition of “Economic Activity”}

According to the Court in \textit{Lopez} and \textit{Morrison}, the central issue in evaluating the constitutionality of the GFSZA and VAWA was whether the regulated activity was economic in nature in that it must involve economic enterprise and commercial transactions.\textsuperscript{150} The Court in \textit{Lopez} narrowed the substantial effects test established in \textit{Jones & Laughlin} by requiring not only that the activity at issue substantially affect interstate commerce, but also that it be economic in nature.\textsuperscript{151} The Court in \textit{Morrison} confirmed that the constitutionality of a statute enacted under the Commerce Clause depends on the economic nature of the activity being regulated.\textsuperscript{152} The \textit{Morrison} Court stated that a proper interpretation

\textsuperscript{145} 42 U.S.C. § 13,981(b).
\textsuperscript{146} Id. § 13,981(c).
\textsuperscript{147} \textit{Morrison}, 529 U.S. at 627.
\textsuperscript{149} See \textit{Morrison}, 529 U.S. at 611–19; \textit{Lopez}, 514 U.S. at 561–68.
\textsuperscript{150} See infra notes 151–77 and accompanying text.
\textsuperscript{151} See \textit{Lopez}, 514 U.S. at 559–61.
\textsuperscript{152} \textit{Morrison}, 529 U.S. at 610–11. Legal scholars and lower courts have debated about how to interpret the requirement that a federal regulation relate to an economic activity. Many have interpreted this requirement broadly. See, e.g., United States v. Gregg, 226 F.3d 253, 261–63 (3d Cir. 2000) (holding that, although the connection to economic or commercial activity plays a central role in whether a law is valid, economic activity can be understood in broad terms); Gibbs v. Babbitt, 214 F.3d 483, 491 (4th Cir. 2000) (“[E]conomic activity must be understood in broad terms. Indeed, a cramped view of commerce would cripple a foremost federal power and in so doing would
of *Lopez* demonstrates that the “noneconomic, criminal nature” of the act of carrying a gun in a school zone was critical to the Court’s decision rejecting the validity of the GFSZA.\(^{153}\) In fact, the *Lopez* Court distinguished past cases in which it had upheld congressional acts regulating economic activity under the Commerce Clause because the activity related to an economic endeavor or enterprise.\(^{154}\) In past Commerce Clause cases the regulated activity always arose out of, or was connected to, a commercial transaction.\(^{155}\)

The Court in *Lopez* referenced numerous cases to support its reasoning that regulated activities at issue in previous Commerce Clause cases always related to an economic endeavor and commercial transactions.\(^{156}\) For example, the Court cited *Hodel v. Virginia Surface Mining & Reclamation Ass’n*,\(^{157}\) in which the Court upheld, as a valid exercise of Congress’s Commerce Clause power, the Surface Mining Control and Reclamation Act, which regulated coal-mining operations involved in mining and selling coal interstate.\(^{158}\) In addition, the Court referred to *Heart of Atlanta Motel, Inc. v. United States*\(^{159}\) and *Katzenbach v. McClung*,\(^{160}\) which held that Title II of the Civil Rights Act of 1964 was constitutional under the Commerce Clause because it

\(^{153}\) *Morrison*, 529 U.S. at 610.

\(^{154}\) *Lopez*, 514 U.S. at 559–61. The *Morrison* Court also emphasized that the federal regulations the Court has sustained in past cases have involved economic endeavors. *Morrison*, 529 U.S. at 611.

\(^{155}\) See, e.g., *Lopez*, 514 U.S. at 561; *Camps Newfound/Owatonna, Inc. v. Town of Harrison*, 520 U.S. 564, 573 (1997) (holding that, although petitioner’s camp was a non-profit business, it engaged in interstate commerce by purchasing and providing goods and services out of state).


\(^{158}\) *Id.* at 276–84.

\(^{159}\) 379 U.S. 241 (1964).

\(^{160}\) 379 U.S. 294 (1964).
aimed to remove burdens on interstate commerce. The cases reasoned that discriminatory practices by hotels and restaurants burdened interstate commerce because they discouraged interstate travel. Therefore, hotels that catered to interstate guests, and restaurants that necessitated the interstate movement of the food products served to customers, were subject to federal regulation under the Commerce Clause.

In addition, the Lopez Court cited Perez v. United States, which held that the Consumer Protection Act’s prohibition of loan sharking was constitutional under Congress’s Commerce Clause power because an individual’s loan sharking activities substantially affect interstate commerce. The common tie among each of these cited cases is that all of the regulated activities relate to an economic enterprise involved in commercial transactions. Coal mining operations relate to the selling of coal; hotels and restaurants purchase supplies and sell food or accommodations; and loan sharking relates to illegal credit transactions.

Finally, the Lopez Court found that, even in Wickard v. Filburn, the activity subject to regulation was economic in nature. The Lopez Court stated that wheat production is an “essential part of a larger regulation of economic activity, in which the regulatory scheme could be undercut unless the intrastate activity were regulated.” Because interstate commercial transactions, namely the buying and selling of wheat, depended upon wheat production, that production could be regulated under the Commerce Clause, whether or not the landowner planned to sell his wheat in interstate commerce. As a result, the Lopez Court’s recharacterization of federal Commerce Clause power makes a finding of economic activity crucial in determining the validity of a statute under the Commerce Clause.

Under this new emphasis on economic activity, the Court rejected the federal regulations at issue in Lopez and Morrison, holding that neither...
regulated economic activity. The Court held that both regulations therefore exceeded Congress’s Commerce Clause power. The GFSZA was a criminal statute that was unrelated, even in the broadest sense, to commerce or an economic enterprise. The Lopez Court found that the GFSZA could not be sustained under the line of cases upholding regulations that “arise out of or are connected with a commercial transaction, which viewed in the aggregate, substantially [affect] interstate commerce.” Similarly, in Morrison, the Court stated that gender-motivated crimes of violence are not economic activity “in any sense of the phrase.”

The Lopez Court did reaffirm the Wickard “aggregate” principle, but it held that the principle has been used only to aggregate activities that “arise out of or are connected with a commercial transaction.” The Court in Morrison clarified its stance on the aggregation principle, declining to adopt “a categorical rule against aggregating the effects of any noneconomic activity.” However, thus far in the history of Commerce Clause jurisprudence, the Court has only upheld regulations when the activity is economic in nature. Thus, in evaluating a new federal regulation, courts should consider whether the activity in question is economic in nature, in the sense that it involves an economic enterprise and commercial transactions, which—in the aggregate—affect interstate commerce.

2. Interstate Effect Analysis

Even if a court finds that an activity is economic in nature, the activity must still have substantial interstate effects before it can be federally regulated under the Commerce Clause. The Lopez Court set forth for

---

171 Morrison, 529 U.S. 598, 627; Lopez, 514 U.S. at 551.
172 Id.
173 Id. The government argued, inter alia, that the GFSZA was related to interstate commerce because the presence of guns in school areas results in violent crime, which in turn negatively impacts the nation’s economy. Id. at 563–64.
175 Lopez, 514 U.S. at 561.
176 Morrison, 529 U.S. at 613.
177 Id. at 613, 617 (rejecting “the argument that Congress may regulate noneconomic, violent criminal conduct based solely on that conduct’s aggregate effect on interstate commerce”).
178 Lopez, 514 U.S. at 559.
Religious Liberty and the Commerce Clause

the first time three elements to assist courts in determining whether an activity has interstate effects: (1) whether legislative findings indicate that the activity has interstate effects, (2) the strength of the nexus between the activity and interstate commerce, and (3) whether the statute contains a jurisdictional element.179

a. The Presence of Congressional Findings

Congressional findings, while not dispositive, are helpful in enabling courts to evaluate Congress’s judgment that the regulated activity substantially burdens interstate commerce.180 In Lopez, Congress did not support its assertion that possession of a gun in a school zone substantially affects interstate commerce with any legislative or congressional committee findings.181 In contrast, the Court in Morrison stated that Congress provided a significant amount of evidence regarding the effect of gender-motivated crimes on victims and their families.182 Although the Court did not discredit these findings, it stated that the presence of congressional findings alone is insufficient to uphold the constitutionality of the Commerce Clause regulation.183

b. The Nexus Between the Activity and Interstate Commerce

The Court, in both Lopez and Morrison, examined the strength of the connection between the regulated activity and interstate commerce.184 In both cases, this link was too attenuated. For example, in Lopez the government argued that the possession of a gun in a school zone affects interstate commerce because such an activity may result in violent crime.185 Crime affects the national economy because its costs are spread among the population through the price of insurance, and people are less likely to travel to a high crime area.186 In addition, guns in school zones

179. Id. at 561–68.
182. Morrison, 529 U.S. at 614.
183. Id. at 614–15 (failing to invalidate Congress’s judgment, but stating that Congress’s findings were “substantially weakened” because they relied on a method of reasoning the Court rejected); see also infra Part III.B.2.b (discussing Congress’s method of reasoning).
185. Lopez, 514 U.S. at 563–64.
186. Id.
affect the educational process, which might lead to less productive citizens, thereby hurting the nation’s economy. In *Morrison*, the government argued that gender motivated crimes affect interstate commerce by deterring potential victims from engaging in employment and traveling interstate. Furthermore, violence against women decreases national productivity, increases medical costs, and decreases supply and demand for interstate products. Despite these potential connections, the Court in both cases held that the regulated activity did not substantially affect interstate commerce.

The Court in both *Lopez* and *Morrison* emphasized that the government would have the Court follow the "but-for causal chain from the initial occurrence of violent crime to every attenuated effect upon interstate commerce." The *Lopez* Court feared that “piling inference upon inference” would lead to federal retention of a general police power, which is properly reserved to the States. In both *Lopez* and *Morrison*, the Court stated that accepting the government’s logic would mean that Congress could regulate not only all violent crimes, but also those activities that lead to violent crimes, regardless of their relationship to interstate commerce. The *Lopez* Court further indicated that it would be hard to imagine any activity by an individual that would be beyond the scope of Congress’s Commerce power under this reasoning. In *Lopez* and *Morrison*, the Court was particularly concerned about the separation of powers implications resulting from the government’s arguments.

c. The Presence of a Jurisdictional Element in the Statute

The presence of a jurisdictional element in a statute provides support for the proposition that the regulated activity is not merely local in

187. Id.
189. Id.
190. See id. at 614–19; *Lopez*, 514 U.S. at 563–68.
193. Id.
194. See *Morrison*, 529 U.S. at 615–16; *Lopez*, 514 U.S. at 564.
nature, but has some connection to interstate commerce. The Court stressed that the purpose of a jurisdictional element is to ensure through a “case-by-case inquiry” that the statute’s reach is limited to those activities that affect interstate commerce. The statute addressed in United States v. Bass provides an example of such a jurisdictional element. The court addressed 18 U.S.C. App. § 1202(a), which states, in part, that any convicted felon “who receives, possesses, or transports in commerce or affecting commerce . . . any firearm shall be fined not more than $10,000 or imprisoned for not more than two years, or both.” The Court’s interpretation of the language “in commerce or affecting commerce” was critical to its determination that the statute passed constitutional muster. The offenses of receiving, possessing, and transporting must occur in commerce or affect commerce to implicate the statute, thereby creating a jurisdictional element. The government could meet its interstate commerce burden by demonstrating that the gun had previously traveled interstate, was moving interstate, was on an interstate facility at the time of the offense, or that possession of the gun affected commerce.

In contrast, neither of the acts at issue in Lopez and Morrison contained a jurisdictional element. The GFSZA did not limit the statute’s reach to a “discrete set of firearm possessions” that clearly would have impacted interstate commerce. The Morrison Court likewise found that the VAWA contained no jurisdictional element. The Court stated that Congress chose to extend the reach of the VAWA to regulate a body of violent crime occurring within the states.
IV. CONGRESS EXCEEDED ITS COMMERCE CLAUSE POWER IN ENACTING THE RLUIPA

The RLUIPA represents an unconstitutional exercise of Congress’s Commerce Clause power because land use regulation does not constitute an economic enterprise or a commercial transaction and does not substantially affect interstate commerce. Beginning with *Lopez*, the Court emphasized that the regulated activity must be economic in nature; that is, it must relate to an economic enterprise and commercial transactions, for Congress to exercise its Commerce power.209 Land use regulation does not constitute an economic enterprise or a commercial transaction.210 Although economics, in addition to public safety and aesthetic considerations, factor into local governments’ land use decisions regulating religious use, this connection does not bring land use regulation within the realm of economic activity as defined by the Court.211

Even if land use decisions are deemed to constitute economic activity, the RLUIPA is still unconstitutional under the Commerce Clause because Congress has not satisfied the elements the Court set forth to ensure that the regulated activity has substantial interstate effects on commerce. First, Congress accumulated little evidence to support its assertion that the economic impact of land use decisions affecting religion is substantial.212 Second, the nexus between land use regulation and interstate commerce is too tenuous because one has to make several inferences to conclude that a land use regulation has an impact on interstate commerce. Furthermore, the Court will not allow the federal government to infringe on an activity properly reserved by the state.213 Finally, although the RLUIPA contains a jurisdictional element, it is not effective because it does not allow for a case-by-case inquiry to ensure that the land use regulation has interstate effects.214

209. See id. at 610–11; *Lopez*, 514 U.S. at 559–61.
210. See infra Part IV.A.
211. See infra Part IV.B.
212. See infra Part IV.B.1.
213. See infra Part IV.B.2.
214. See infra Part IV.B.3.
Religious Liberty and the Commerce Clause

A. The RLUIPA Does Not Regulate Economic Activity

Land use regulation is the activity a court must evaluate to determine whether the RLUIPA is valid under the Commerce Clause. 215 Under Lopez, a regulated activity must constitute economic activity by involving economic enterprise and commercial transactions. 216 Land use regulation is more comparable to the non-economic activities at issue in Lopez and Morrison than it is to the Lopez Court’s list of activities exemplifying economic activity. 217 Although various economic considerations factor into land use regulation, this does not establish that land use regulation constitutes economic activity as defined by the Court. 218

Determining what activity a statute regulates is critical for evaluating the constitutionality of the statute under the Commerce Clause. 219 Congress enacted the RLUIPA to regulate local governments’ land use decisions due to concerns that local governments were overly burdening religion. 220 Therefore, to evaluate the constitutionality of the RLUIPA under the Commerce Clause, a court must determine whether land use regulation constitutes economic activity.

The Lopez Court listed several past Commerce Clause cases to demonstrate that its definition of economic activity was quite narrow. 221 The Court revealed that the activities of coal mining operations involved in the sale of coal, restaurants that purchase supplies and sell food, hotels that rent rooms, and loan sharks that engage in extortionate credit transactions all constitute economic activity. 222 These specific references suggest that, to meet the definition of economic activity, a regulated activity must involve economic enterprise and commercial transactions. 223

217. See id. at 559–61.
218. See supra Parts I.A, B.2.
221. See Lopez, 514 U.S. at 559–61.
222. See supra Part III.B.1.
In contrast, the regulated activities at issue in *Lopez* and *Morrison* did not constitute economic activity.\(^{224}\) Neither the act of carrying a gun in a school zone nor the act of committing a gender-motivated crime meets the narrow definition of economic activity set forth in *Lopez*.\(^{225}\) Although proponents of the GFSZA and the VAWA argued that the regulated activities impacted interstate commerce,\(^{226}\) the Court held that mere economic effects were insufficient to satisfy its definition of economic activity.\(^{227}\) Therefore, because the regulated activities were not economic in nature, neither congressional act was justified under the Commerce Clause.

As with the regulated activities in *Lopez* and *Morrison*, land use regulation fails to satisfy the Court’s narrow definition of economic activity. Land use regulation itself is not an economic endeavor and does not involve a transaction or the buying and selling of goods.\(^{228}\) Of course, land use regulation might encourage or prevent future economic transactions, such as the buying and selling of construction materials.\(^{229}\) Nonetheless, in *Lopez* and *Morrison*, the Court held that the economic nature of the activity is key, not the economic effects.\(^{230}\) Therefore, any future impact on a commercial transaction does not bring land use regulation within the Court’s narrow definition of economic activity.

Supporters of the RLUIPA may argue that land use determinations often involve economic considerations, thus making land use regulation an economic activity. For example, local governments often adopt land use regulations for the protection of aesthetics and public safety "with a view to conserving the value of buildings,"\(^{231}\) and one rationale for excluding religious uses from residential areas is the possible detrimental

\(\text{\(^{224}\) United States v. Morrison, 529 U.S. 598, 613 (2000); Lopez, 514 U.S. at 561.}\)
\(\text{\(^{225}\) Morrison, 529 U.S. at 613; Lopez, 514 U.S. at 561.}\)
\(\text{\(^{226}\) Morrison, 529 U.S. at 615; Lopez, 514 U.S. at 563–64.}\)
\(\text{\(^{227}\) Morrison, 529 U.S. at 610–11; Lopez, 514 U.S. at 559–61.}\)
\(\text{\(^{228}\) See supra Part I.}\)
\(\text{\(^{229}\) 146 CONG. REC. S7775 (daily ed. July 27, 2000) (exhibit 1).}\)
\(\text{\(^{230}\) See Morrison, 529 U.S. at 610–11; Lopez, 514 U.S. at 559–61. The Court has historically recognized that numerous trivial effects may substantially affect interstate commerce when considered in the aggregate. See, e.g., Wickard v. Filburn, 317 U.S. 111, 127–28 (1942). Nonetheless, the aggregate principle is unrelated to the determination of whether land use regulation constitutes economic activity; the *Morrison* Court suggested that the Court has invoked the aggregation principle only in those cases in which the regulated activity is economic in nature. *Morrison*, 529 U.S. at 613.}\)
\(\text{\(^{231}\) SZEA, supra note 15, at Appendix A, 215.}\)
effect on the property values in the community. The goal of preserving property values is an economic consideration. Further, in both City of Chicago Heights and Cornerstone Bible Church, courts upheld exclusion of religious uses from industrial zones to foster “economic stability and growth.” to stimulate commercial development, and to dedicate an area for industry.

These economic considerations, however, merely influence land use regulation. Non-economic factors play an equally important role. The SZEA suggests that land use zoning emerged out of a concern for public safety and aesthetics. Courts also have recognized that promoting public health and safety and aesthetics are primary purposes of land use regulation and local governments’ justifications for excluding religious uses frequently are based on aesthetic and nuisance-type concerns. Although economic considerations may play a role in land use determinations, they are merely part of a host of other factors that influence land use regulation. Furthermore, the mere consideration of economics, by itself, does not convert land use regulation into economic activity. Even if local governments consider economics in making land use determinations, this factor alone does not justify Congress’s regulation of such decisions based on the Commerce Clause.

Another connection between land use regulation and economics might occur if a land use regulation implicates the Takings Clause. A regulatory taking might implicate “economic activity” in two ways. First,

---

232. See, e.g., Lakewood, Ohio Congregation of Jehovah’s Witnesses, Inc. v. City of Lakewood, 699 F.2d 303, 305 (6th Cir. 1983); West Hartford Methodist Church v. Zoning Bd. of Appeals, 121 A.2d 640, 642–43 (Conn. 1956); Milwaukie Co. of Jehovah’s Witnesses v. Mullen, 330 P.2d 5, 17 (Or. 1958).


238. See supra Part III.B.1.


1281
the court must determine when the government has taken private property for public use by asking whether the land use regulation deprives the landowner of all economically beneficial or productive use of the land. Second, the government must pay landowners the value of their land when private property is taken for public use. These two economic implications of a taking, however, should be considered merely effects of land use regulation. The Court in *Lopez* and *Morrison* stated that the nature of the activity, not its effects, is key in determining whether land use regulation constitutes “economic activity.” Furthermore, the takings analysis does not suggest that land use regulation itself involves an economic transaction. Therefore, a court is not likely to find that a takings claim transforms land use regulation into “economic activity.”

Land use regulation does not fall within the *Lopez* Court’s narrow definition of economic activity. Land use regulation more closely resembles the regulated activities in *Lopez* and *Morrison*. A mere connection to economics does not suggest that the activity is economic in nature. Instead, an activity that is economic in nature must involve economic enterprise and commercial transactions. Because land use regulation is not an economic enterprise and does not involve commercial transactions, it cannot be characterized as economic activity for Commerce Clause purposes.

**B. The RLUIPA Does Not Regulate an Activity That Has Substantial Effects on Interstate Commerce**

Even if courts were to determine that the RLUIPA regulates economic activity, the RLUIPA does not satisfy the elements courts must evaluate to ensure that the regulated activity “substantially affects interstate commerce.” Thus, Congress lacked the Commerce Clause authority to enact the statute due to its tenuous effect on interstate commerce. Courts should hold that Congress has failed to ensure that the RLUIPA is limited to the regulation of interstate activity for the following reasons:

241. U.S. CONST. amend. V.
244. See infra Parts IV.B.1–B.3.
Religious Liberty and the Commerce Clause

(1) Congress failed to set forth findings indicating land use regulation has interstate effects;\(^{245}\) (2) the strength of the nexus between land use regulation and the economy is too tenuous;\(^{246}\) and, (3) the RLUIPA does not have an effective jurisdictional element.\(^{247}\)

1. **Congressional Findings Do Not Support the Interstate Commerce Effects of the RLUIPA**

Congress failed to set forth legislative findings regarding the effect land use regulations that burden religion have on interstate commerce. Although Congress passed the RLUIPA without committee action, it relied on the legislative history of the RLPA in enacting the RLUIPA.\(^{248}\) The RLPA’s legislative history contains a vast amount of evidence explaining the detrimental impact land use determinations can have on religion.\(^{249}\) Little, if any, of this material, however, focuses on the potential economic impact land use determinations might have on the interstate economy.\(^{250}\)

In support of the RLUIPA, Senators Hatch and Kennedy stated that the economic impact of land use decisions affecting religion was substantial.\(^{251}\) The Senators, however, incorrectly reported that Marc D. Stern’s testimony presented data to confirm such a statement.\(^{252}\) Stern emphasized the impact religious activity has on the economy.\(^{253}\) Stern’s statistical information regarding the value of property controlled by religious institutions was unrelated to any possible impact land use decisions affecting religion have on interstate commerce.\(^{254}\) Although he indicated that religious institutions spent $8 billion in 1992 on capital

\(^{245}\) See infra Part IV.B.1.

\(^{246}\) See infra Part IV.B.2.

\(^{247}\) See infra Part IV.B.3.


\(^{250}\) See H.R. REP. NO. 106-219, at 18–24 (1999); 146 CONG. REC. E1564–67 (statement of Rep. Hyde); 146 CONG. REC. S7774–75 (exhibit 1).

\(^{251}\) 146 CONG. REC. S7775 (exhibit 1).

\(^{252}\) See id.


improvements to facilities, his assertion that land use regulations may impact the interstate movement of supplies and services for the construction of facilities was unsupported. This lack of legislative findings regarding land use decisions affecting religious uses does not support Congress’s assertion that it has the authority to enact the RLUIPA under its Commerce Clause power.

2. The Nexus Between Land Use Decisions and Interstate Commerce Is Too Tenuous

The link between land use regulation and its effect on interstate commerce is too attenuated to give Congress authority to regulate local land use decisions under the Commerce Clause. Congress’s reasoning in support of the RLUIPA is similar to the “but-for causal chain” the Court rejected when evaluating the GFSZA and the VAWA. In Lopez, the Court also rejected the government’s logic of “piling inference upon inference” to conclude that the regulated activity impacted interstate economy. The same piling of inferences is needed to connect land use regulation to the interstate economy.

Land use regulation admittedly is more closely connected to the interstate economy than possessing a gun in a school zone or committing a gender-motivated crime. Nonetheless, the connection is still not sufficient. In Lopez, the government argued that violent crime caused by the possession of a gun in a school zone substantially affects interstate commerce. For example, gun regulation might decrease travel to the high crime area. To reach this conclusion, one would have to infer that possessing the gun in a school zone would actually lead to a violent crime. Furthermore, one would have to assume that the crime would actually cause people not to travel to that area, and that this decrease in travel would affect interstate commerce. The link between the regulated activity and interstate commerce arguably was more direct in Morrison

255. See 1998 RLPA, supra note 106 at 58.
258. Lopez, 514 U.S. at 567.
259. Compare supra, Part I (describing land use regulation) with Morrison, 529 U.S. at 615 and Lopez, 514 U.S. at 563–64 (detailing the government’s argument for how gender-motivated crimes and the possession of a gun in a school zone affect interstate commerce).
261. Id.
because the VAWA regulates the violent crime itself. \footnote{Kropf, supra note 152, at 402.} Thus, the significant inference one would have to make under \textit{Morrison} is that the gender-motivated crime affects interstate commerce, perhaps by deterring the victim from traveling interstate or from engaging in employment. \footnote{Morrison, 529 U.S. at 615.}

Similar to the regulated activities in \textit{Lopez} and \textit{Morrison}, a land use regulation restricting religious uses does not directly implicate interstate commerce. To show that religious land use regulation substantially impacts interstate commerce, one would have to infer that the regulation would cause a religious institution to refrain from undertaking an action. For example, in considering a land use ordinance restricting the ability of a church to expand, a court would first have to infer that a church intended to expand. The court would then have to make the additional inference that the church’s inability to expand would somehow affect interstate commerce, perhaps by decreasing the demand for interstate labor or interstate supplies. \footnote{See supra note 110 and accompanying text.} Thus, the number and type of inferences one has to make to conclude that a religious land use regulation affects interstate commerce is similar to the reasoning the U.S. Supreme Court rejected in \textit{Lopez} and \textit{Morrison}. \footnote{See supra Part III.B.2.b. The connection between a land use regulation and interstate commerce arguably may be more direct given alternative fact patterns. For example, a regulation preventing a church expansion might more directly affect interstate commerce if the church could prove with specific evidence that it intended to use a product that only was manufactured out of state. However, the U.S. Supreme Court rejected even these more closely-related connections in \textit{Morrison}, where the regulated activity was only two steps removed from interstate commerce. See \textit{Morrison}, 529 U.S. at 615–19.} Congress asserted that land use decisions prevented specific economic transactions such as the construction of a facility or the sale, purchase, or rental of a building. \footnote{146 CONG. REC. S7775 (daily ed. July 27, 2000) (exhibit 1).} The prevention of an interstate economic transaction is a possible effect of a land use regulation; however, a court would have to “pile inference upon inference” to conclusively determine that land use regulation substantially affects interstate commerce.

Accepting Congress’s reasoning for enacting the RLUIPA under its Commerce Clause authority would suggest that Congress could regulate all local land use regulations. The Court in \textit{Morrison} stated that allowing Congress to regulate gender-motivated crimes would allow Congress to
regulate any type of violent crime because gender-motivated violence, “as a subset of all violent crime, is certain to have lesser economic impacts than the larger class of which it is a part.”267 The Court feared that allowing Congress to regulate all violent crime would result in the obliteration of “the Constitution’s distinction between national and local authority.”268 Similarly, if the court affirmed Congress’s authority under the Commerce Clause, a land use determination, regardless of whether it burdened religious uses, would fall within Congress’s Commerce Clause authority because every land use decision would have some effect on interstate commerce.

Although every land use regulation may have a distant effect on interstate commerce, courts should be reluctant to approve legislation that would lead to unlimited congressional Commerce Clause power over a function that traditionally has been left to the states.269 The Court in Lopez stated that the Commerce Clause is subject to outer limits and must not be extended so far that it destroys the distinction between what is local and what is national, thus creating a centralized government.270 The Court rejected Congress’s reasoning for nationalizing regulation of crimes motivated by gender because the reasoning could be “applied equally as well to family law and other areas of traditional state regulation since the aggregate effect of marriage, divorce, and childrearing on the national economy is undoubtedly significant.”271 Similarly, Congress’s reasoning in support of the RLUIPA could be applied equally as well to all land use regulation, and the Court has held that land use regulation, like criminal law enforcement or education, is a function that historically has been performed by the states.272 This concern for preserving state sovereignty as well as the attenuated connection between land use regulation and interstate commerce suggests that Congress does not have authority under the Commerce Clause to enact the RLUIPA.

267. Morrison, 529 U.S. at 615.
268. Id.
271. Morrison, 529 U.S. at 615–16.
272. Solid Waste Agency of Northern Cook County, 531 U.S. at 174; Hess, 513 U.S. at 44.
3. The RLUIPA Does Not Contain an Effective Jurisdictional Element

The jurisdictional element contained in the RLUIPA fails to accomplish its required purpose. As expressed in *Lopez*, the purpose of a jurisdictional element is to ensure through a case-by-case inquiry that only activities with interstate effects are regulated. 273 The RLUIPA’s scope of application states that it only applies when a land use regulation affects interstate commerce. 274 In the joint statement of Senators Hatch and Kennedy, the Senators stated that the jurisdictional element requires that each regulated burden on religious exercise affect interstate commerce. 275

This limitation in the RLUIPA should be compared with the acceptable jurisdictional element in *Bass*. 276 In that case, the government had to prove that the firearm in question had previously traveled interstate, was moving interstate, was on an interstate facility at the time of the offense, or its possession affected commerce. 277 Perhaps with the exception of the last, it is relatively easy for the government to make these factual determinations.

In contrast, the limiting language in the RLUIPA does not lend itself to a case-by-case inquiry. In each case in which the government alleges a violation of the RLUIPA, it has the duty to demonstrate what effect the land use regulation at issue has on interstate commerce. 278 Unlike tracing the origin of a gun, it is much more difficult to demonstrate with specific, non-speculative data that each conflict between religion and a land use regulation has interstate commerce effects. 279 For example, in a case in which a church is unable to construct a facility due to a land use regulation, the government may argue that the regulation results in a decrease in the demand for labor or supplies. 280 To satisfy the jurisdictional element, the government would have to employ the reasoning the Court rejected in *Lopez* and *Morrison* of “piling inference upon inference” to reach the further conclusion that the church intended to...

---

273. See supra notes 198–206 and accompanying text.
277. Id.
278. See supra note 273 and accompanying text.
279. See supra Part IV.B.2.
280. See supra note 110 and accompanying text.
to use out of state labor or supplies originating from out of state.\textsuperscript{281} Thus, the purpose of the RLUIPA’s jurisdictional element is compromised by the difficulty a local government would face in making this determination. Although Congress included a jurisdictional element in the RLUIPA, courts would find it unhelpful in ensuring that only activities that have interstate effects are regulated by the statute. Without a valid jurisdictional element, there is no added assurance that a statute is limited in scope.

V. CONCLUSION

Congress’s third attempt at protecting religious liberty through the RLUIPA fails to adequately invoke the Commerce Clause power. This result might have been different had the U.S. Supreme Court adhered to its pre-\textit{Lopez} approach to the Commerce Clause. However, the Court’s current approach represents a return to the formulistic, restrictive interpretation of the Commerce Clause between 1888 and 1937. Thus, the RLUIPA is unconstitutional under the Court’s current interpretation of the Commerce Clause.

The U.S. Supreme Court’s relatively recent addition of the “economic activity” test substantially narrows Congress’s regulatory powers under the Commerce Clause. Only an activity involving an economic enterprise and commercial transactions satisfies the Court’s test. This reveals a critical defect of the RLUIPA: land use regulation does not constitute an economic enterprise or transaction as defined by the U.S. Supreme Court.

However, even if a court were to determine that land use regulation constitutes economic activity, Congress did not satisfy the elements the U.S. Supreme Court established to ensure that the statute regulates only those activities that affect interstate commerce. Most significantly, courts should be concerned that approving the RLUIPA would open the door for Congress to completely usurp local government authority over its land use. Therefore, Congress’s latest attempt at protecting religious liberty will be frustrated by limitations on the Commerce Clause power imposed by the U.S. Supreme Court.

\textsuperscript{281} See supra Part IV.B.2. It might be possible for a party to meet this burden under certain specific fact patterns. See supra note 266 and accompanying text.
WASHINGTON LAW REVIEW

INDEX

Volume 76

2001
## AUTHOR INDEX

<table>
<thead>
<tr>
<th>Author</th>
<th>Title</th>
<th>Volume: Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aronson, Robert H.</td>
<td>Attorney-Cleint Confidentiality and the Assessment of Claimants Who Allege Posttraumatic Stress Disorder</td>
<td>76:313</td>
</tr>
<tr>
<td>Art, Robert C.</td>
<td>Conversion and Merger of Disparate Business Entities</td>
<td>76:349</td>
</tr>
<tr>
<td>Azure, Anthony H.</td>
<td>Festo’s Effect on After-Arising Technology and the Doctrine of Equivalents</td>
<td>76:1153</td>
</tr>
<tr>
<td>Beane, Amanda J.</td>
<td>One Step Forward, Two Steps Back: Vasquez v. Hawthorne Wrongly Denied Washington’s Meretricious Relationship Doctrine to Same-Sex Couples</td>
<td>76:475</td>
</tr>
<tr>
<td>Bender, Deborah Leigh</td>
<td>Echazabal v. Chevron: A Direct Threat to Employers in the Ninth Circuit</td>
<td>76:859</td>
</tr>
<tr>
<td>Blomquist, Megan E.</td>
<td>A Shield, Not a Sword: Involuntary Leave Under the Family and Medical Leave Act</td>
<td>76:509</td>
</tr>
<tr>
<td>Cole, Douglas R.</td>
<td>E-Proxies for Sale? Corporate Vote-Buying in the Internet Age</td>
<td>76:793</td>
</tr>
<tr>
<td>Donohue, Alfred E.</td>
<td>Unopened Public Street Easements in Washington: Whose Right To Use That Land Is It, Anyway?</td>
<td>76:541</td>
</tr>
<tr>
<td>Farrow, Raymond J.</td>
<td>Qualifying Immunity: Protecting State Employees’ Right to Protect Their Employment Rights After Alden v. Maine</td>
<td>76:149</td>
</tr>
<tr>
<td>Glasgow, Rebecca R.</td>
<td>Can Students Sue When Schools Don’t Make the Grade? The Washington Assessment of Student Learning and Educational Malpractice</td>
<td>76:893</td>
</tr>
<tr>
<td>Harkins, Ryan P.</td>
<td>Holding Tortfeasors Accountable: Apportionment of Enhanced Injuries Under Washington’s Comparative Fault Scheme</td>
<td>76:1185</td>
</tr>
<tr>
<td>Hedberg, Matthew E.</td>
<td>Kim Ho Ma v. Reno: Cloaking Judicial Activism As Constitutional Avoidance</td>
<td>76:669</td>
</tr>
<tr>
<td>Hersey, Brian</td>
<td>Roberts v. Dudley: An Unnecessary Broadening of the Public Policy Exception to the Employment-at-Will Doctrine in Washington</td>
<td>76:179</td>
</tr>
<tr>
<td>Hoon, Carrie L.</td>
<td>The Reasonable Girl: A New Reasonableness Standard To Determine Sexual Harassment in Schools</td>
<td>76:213</td>
</tr>
<tr>
<td>Huebner, Kasey D.</td>
<td>In the Litigation Business: Insurance Company Liability for Acts Occurring in the Course of Litigation Under the Washington Consumer Protection Act</td>
<td>76:927</td>
</tr>
<tr>
<td>Hussong, Marni</td>
<td>Protecting the Tax-Exempt Status of Housing Developers Participating in Low-Income Housing Tax Credit Partnerships</td>
<td>76:243</td>
</tr>
<tr>
<td>Imwinkelried, Edward J.</td>
<td>DNA Typing: Emerging or Neglected Issues</td>
<td>76:413</td>
</tr>
<tr>
<td>Kaye, D.H.</td>
<td>DNA Typing: Emerging or Neglected Issues</td>
<td>76:413</td>
</tr>
<tr>
<td>Larson, Carlton F.W.</td>
<td>The Declaration of Independence: A 225th Anniversary Re-Interpretation</td>
<td>76:701</td>
</tr>
<tr>
<td>Martin, Eric B.</td>
<td>Grandma Got Run Over by the Supreme Court: Suggestions for a Constitutional Nonparental Visitation Statute After Troxel v. Granville</td>
<td>76:571</td>
</tr>
</tbody>
</table>
Miles, Shylah
Two Wrongs Do Not Make a Defense: Eliminating the Equal-Opportunity-Harasser Defense 76:603

Moomaw, Paul W.
Fire Sale! The Admissibility of Evidence of Environmental Contamination to Determine Just Compensation in Washington Eminent Domain Proceedings 76:1221

Mueller, Janice M.
No “Dilettante Affair”: Rethinking the Experimental Use Exception to Patent Infringement for Biomedical Research Tools 76:1

Murphy, Richard W.
Punitive Damages, Explanatory Verdicts, and the Hard Look 76:995

Perrine, Aaron
The First Amendment Versus the World Trade Organization: Emergency Powers and the Battle in Seattle 76:635

Reynolds, Sidney Charlotte
Closing a Discrimination Loophole: Using Title VII’s Anti-Retaliation Provision to Prevent Employers from Requiring Unlawful Arbitration Agreements As Conditions of Continued Employment 76:957

Rosen, Gerald M.
Attorney-Cleint Confidentiality and the Assessment of Claimants Who Alleg Posttraumatic Stress Disorder 76: 313

Rosenwald, Lonnie
Attorney-Cleint Confidentiality and the Assessment of Claimants Who Alleg Posttraumatic Stress Disorder 76: 313

Shapiro, Evan M.
The Religious Land Use and Institutionalized Persons Act: An Analysis Under the Commerce Clause 76:1255

Somerville, Thane
The Equal Pay Act As Appropriate Legislation Under Section 5 of the Fourteenth Amendment: Can State Employers Be Sued? 76:279

Zeigler, Donald H.
Rights, Rights of Action, and Remedies: An Integrated Approach 76:67
A Shield, Not a Sword: Involuntary Leave Under the Family and Medical Leave Act
Megan E. Blomquist 76:509

Attorney-Cleint Confidentiality and the Assessment of Claimants Who Allege Posttraumatic Stress Disorder
Robert H. Aronson, Lonnie Rosenwald, & Gerald M. Rosen 76:313

Can Students Sue When Schools Don’t Make the Grade? The Washington Assessment of Student Learning and Educational Malpractice
Rebecca R. Glasgow 76:893

Closing a Discrimination Loophole: Using Title VII’s Anti-Retaliation Provision to Prevent Employers from Requiring Unlawful Arbitration Agreements As Conditions of Continued Employment
Sidney Charlotte Reynolds 76:957

Conversion and Merger of Disparate Business Entities
Robert C. Art 76:349

DNA Typing: Emerging or Neglected Issues
Edward J. Imwinkelried & D.H. Kaye 76:413

Echazabal v. Chevron: A Direct Threat to Employers in the Ninth Circuit
Deborah Leigh Bender 76:859

E-Proxyx for Sale? Corporate Vote-Buying in the Internet Age
Douglas R. Cole 76:793

Festo’s Effect on After-Arising Technology and the Doctrine of Equivalents
Anthony H. Azure 76:1153

Fire Sale! The Admissibility of Evidence of Environmental Contamination To Determine Just Compensation in Washington Eminent Domain Proceedings
Paul W. Moomaw 76:1221

Grandma Got Run Over by the Supreme Court: Suggestions for a Constitutional Nonparental Visitation Statute After Troxel v. Granville
Eric B. Martin 76:571

Harry Maybury Cross: In Memoriam
William B. Stoebuck 76:991

Holding Tortfeasors Accountable: Apportionment of Enhanced Injuries Under Washington’s Comparative Fault Scheme
Ryan P. Harkins 76:1185

In the Litigation Business: Insurance Company Liability for Acts Occurring in the Course of Litigation Under the Washington Consumer Protection Act
Kasey D. Huebner 76:927

Kim Ho Ma v. Reno: Cloaking Judicial Activism As Constitutional Avoidance
Matthew E. Hedberg 76:669

No “Dilettante Affair”: Rethinking the Experimental Use Exception to Patent Infringement for Biomedical Research Tools
Janice M. Mueller 76:1

One Step Forward, Two Steps Back: Vasquez v. Hawthorne Wrongly Denied Washington’s Meretricious Relationship Doctrine to Same-Sex Couples
Amanda J. Beane 76:475
Protecting the Tax-Exempt Status of Housing Developers Participating in Low-Income Housing Tax Credit Partnerships
Marni Hussong 76:243

Punitivie Damages, Explanatory Verdicts, and the Hard Look
Richard W. Murphy 76:995

Qualifying Immunity: Protecting State Employees’ Right to Protect Their Employment Rights After Alden v. Maine
Raymond J. Farrow 76:149

Rights, Rights of Action, and Remedies: An Integrated Approach
Donald H. Zeigler 76:67

Roberts v. Dudley: An Unnecessary Broadening of the Public Policy Exception to the Employment-at-Will Doctrine in Washington
Brian Hersey 76:179

The Assymetry of State Sovereign Immunity
Richard H. Seamon 76:1067

The Declaration of Independence: A 225th Anniversary Re-Interpretation
Carlton F.W. Larson 76:701

The Equal Pay Act as Appropriate Legislation Under Section 5 of the Fourteenth Amendment: Can State Employers Be Sued?
Thane Somerville 76:279

The First Amendment Versus the World Trade Organization: Emergency Powers and the Battle in Seattle
Aaron Perrine 76:635

The Reasonable Girl: A New Reasonableness Standard To Determine Sexual Harassment in Schools
Carrie L. Hoon 76:213

The Religious Land Use and Institutionalized Persons Act: An Analysis Under the Commerce Clause
Evan M. Shapiro 76:1255

Two Wrongs Do Not Make a Defense: Eliminating the Equal-Opportunity-Harasser Defense
Shylah Miles 76:603

Unopened Public Street Easements in Washington: Whose Right To Use That Land Is It, Anyway?
Alfred E. Donohue 76:541
SUBJECT INDEX

Business
Conversion and Merger of Disparate Business Entities 76:349

Child Visitation
Grandma Got Run Over by the Supreme Court: Suggestions for a Constitutional Nonparental Visitation Statute After Troxel v. Granville 76:571

Commerce Clause
The Religious Land Use and Institutionalized Persons Act: An Analysis Under the Commerce Clause 76:
The Asymmetry of State Sovereign Immunity 76:1067

Community Property
One Step Forward, Two Steps Back: Vasquez v. Hawthorne Wrongly Denied Washington’s Meretricious Relationship Doctrine to Same-Sex Couples 76:476

Constitutional Law
The Asymmetry of State Sovereign Immunity 76:1067
DNA Typing: Emerging or Neglected Issues 76:413
Kim Ho Ma v. Reno: Cloaking Judicial Activism As Constitutional Avoidance 76:669
The Declaration of Independence: A 225th Anniversary Re-Interpretation 76:701
The Equal Pay Act as Appropriate Legislation Under Section 5 of the Fourteenth Amendment: Can State Employers Be Sued? 76:279

Consumer Protection
The First Amendment Versus the World Trade Organization: Emergency Powers and the Battle in Seattle 76:635

Corporations
Conversion and Merger of Disparate Business Entities 76:349
E-Proxies for Sale? Corporate Vote-Buying in the Internet Age 76:793

Damages
The Asymmetry of State Sovereign Immunity 76:1067
Holding Tortfeasors Accountable: Apportionment of Enhanced Injuries Under Washington’s Comparative Fault Scheme 76:1185
Punitive Damages, Explanatory Verdicts, and the Hard Look

Doctrine of Equivalents
Festo’s Effect on After-Arising Technology and the Doctrine of Equivalents 76:

Easements
Unopened Public Street Easements in Washington: Whose Right to Use That Land Is It, Anyway? 76:541

Educational Malpractice
Can Students Sue When Schools Don’t Make the Grade? The Washington Assessment of Student Learning and Educational Malpractice 76:893
Employment Discrimination

Roberts v. Dudley: An Unnecessary Broadening of the Public Policy Exception to the Employment-at-Will Doctrine in Washington 76:179

Two Wrongs Do Not Make a Defense: Eliminating the Equal-Opportunity-Harasser Defense 76:503

Employment Law

A Shield, Not a Sword: Involuntary Leave Under the Family and Medical Leave Act 76:509

Closing a Discrimination Loophole: Using Title VII's Anti-Retaliation Provision to Prevent Employers from Requiring Unlawful Arbitration Agreements As Conditions of Continued Employment 76:957

Echazaael v. Chevron: A Direct Threat to Employers in the Ninth Circuit 76:859

Two Wrongs Do Not Make a Defense: Eliminating the Equal-Opportunity-Harasser Defense 76:603

Equal Pay Act

The Equal Pay Act as Appropriate Legislation Under Section 5 of the Fourteenth Amendment: Can State Employers Be Sued? 76:279

Evidence

Attorney-Client Confidentiality and the Assessment of Claimants Who Allege Posttraumatic Stress Disorder 76:313

DNA Typing: Emerging or Neglected Issues 76:413

Fair Labor Standards Act

Qualifying Immunity: Protecting State Employees' Right to Protect Their Employment Rights After Alden v. Maine 76:149

Family Law

Grandma Got Run Over by the Supreme Court: Suggestions for a Constitutional Nonparental Visitation Statute After Troxel v. Granville 76:571

Family and Medical Leave

A Shield, Not a Sword: Involuntary Leave Under the Family and Medical Leave Act 76:509

Immigration Law

Kim Ho Ma v. Reno: Cloaking Judicial Activism As Constitutional Avoidance 76:669

Implied Cause of Action

Can Students Sue When Schools Don’t Make the Grade? The Washington Assessment of Student Learning and Educational Malpractice 76:893

Involuntary Leave

A Shield, Not a Sword: Involuntary Leave Under the Family and Medical Leave Act 76:509

Jurisprudence

Rights, Rights of Action, and Remedies: An Integrated Approach 76:67

Land Use Regulation

The Asymmetry of State Sovereign Immunity 76:1067

Fire Sale! The Admissibility of Evidence of Environmental Contamination to Determine Just Compensation in Washington Eminent Domain Proceedings 76:1221

The Religious Land Use and Institutionalized Persons Act: An Analysis Under the Commerce Clause 76:1255

Unopened Public Street Easements in Washington: Whose Right to Use That Land Is It, Anyway? 76:541
Mandatory Arbitration Agreements
Closing a Discrimination Loophole: Using Title VII’s Anti-Retaliation Provision to Prevent Employers from Requiring Unlawful Arbitration Agreements As Conditions of Continued Employment 76:957

Meretricious Relationships
One Step Forward, Two Steps Back: Vasquez v. Hawthorne Wrongly Denied Washington’s Meretricious Relationship Doctrine to Same-Sex Couples 76:476

Parental Rights
Grandma Got Run Over by the Supreme Court: Suggestions for a Constitutional Nonparental Visitation Statute After Troxel v. Granville 76:571

Patent Law
Festo’s Effect on After-Arising Technology and the Doctrine of Equivalents 76:

No “Dilettante Affair”: Rethinking the Experimental Use Exception to Patent Infringement for Biomedical Research Tools 76:1

Product Liability
Holding Tortfeasors Accountable: Apportionment of Enhanced Injuries Under Washington’s Comparative Fault Scheme 76:

Property
Unopened Public Street Easements in Washington: Whose Right to Use That Land Is It, Anyway? 76:541

Public School Law
Can Students Sue When Schools Don’t Make the Grade? The Washington Assessment of Student Learning and Educational Malpractice 76:893

Punitive Damages
Punitive Damages, Explanatory Verdicts, and the Hard Look 76:995

Qualified Immunity
Qualifying Immunity: Protecting State Employees’ Right to Protect Their Employment Rights After Alden v. Maine 76:149

Religious Liberty
The Religious Land Use and Institutionalized Persons Act: An Analysis Under the Commerce Clause 76:

Remedies
Rights, Rights of Action, and Remedies: An Integrated Approach 76:67

Same-Sex Marriage
One Step Forward, Two Steps Back: Vasquez v. Hawthorne Wrongly Denied Washington’s Meretricious Relationship Doctrine to Same-Sex Couples 76:476

Sexual Harassment
Two Wrongs Do Not Make a Defense: Eliminating the Equal-Opportunity-Harasser Defense 76:603

No “Dilettante Affair”: Rethinking the Experimental Use Exception to Patent Infringement for Biomedical Research Tools 76:1

The Reasonable Girl: A New Reasonableness Standard To Determine Sexual Harassment in Schools 76:213

Sovereign Immunity
The Asymmetry of State Sovereign Immunity 76:1067

The Equal Pay Act as Appropriate Legislation Under Section 5 of the Fourteenth Amendment: Can State Employers Be Sued? 76:279

Statutory Interpretation
Kim Ho Ma v. Reno: Cloaking Judicial Activism As Constitutional Avoidance 76:669

Roberts v. Dudley: An Unnecessary Broadening of the Public Policy Exception to the Employment-at-Will Doctrine in Washington 76:179
Tax

Protecting the Tax-Exempt Status of Housing Developers Participating in Low-Income Housing Tax Credit Partnerships 76:243

Torts

Holding Tortfeasors Accountable: Apportionment of Enhanced Injuries Under Washington’s Comparative Fault Scheme 76:1185

Wrongful Discharge in Violation of Public Policy

Roberts v. Dudley: An Unnecessary Broadening of the Public Policy Exception to the Employment-at-Will Doctrine in Washington 76:179
<table>
<thead>
<tr>
<th>CASE INDEX</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aladdin, Inc. v. Black Hawk County, 562 N.W.2d 608 (Iowa 1997) 76:1240</td>
</tr>
<tr>
<td>Am. Campaign Acad. v. Comm’r, 92 T.C. 1053 (1989) 76:258</td>
</tr>
<tr>
<td>Barnes v. Costle, 561 F.2d 983 (D.C. Cir. 1977) 76:615</td>
</tr>
<tr>
<td>Barrera-Echavarria v. Rison, 44 F.3d 1441 (9th Cir. 1995) 76:683</td>
</tr>
<tr>
<td>Baxter-Wyckoff Co. v. City of Seattle, 67 Wash. 2d 555, 408 P.2d 1012 (1965) 76:549</td>
</tr>
<tr>
<td>Bay Area Peace Navy v. United States, 914 F.2d 1224 (9th Cir. 1990) 76:649</td>
</tr>
<tr>
<td>Beagle v. Beagle, 678 So. 2d 1271 (Fla. 1996) 76:583</td>
</tr>
<tr>
<td>Belknap v. Schild, 161 U.S. 10 (1896) 76:1084</td>
</tr>
<tr>
<td>Bell v. Hood, 327 U.S. 678 (1946) 76:79</td>
</tr>
<tr>
<td>Bennett v. Hardy, 113 Wash. 2d 912, 784 P.2d 1258 (1990) 76:197, 900</td>
</tr>
<tr>
<td>Bentivegna v. United States Dep’t of Labor, 694 F.2d 619 (9th Cir. 1982) 76:873</td>
</tr>
<tr>
<td>Blessing v. Freestone, 520 U.S. 329 (1997) 76:139</td>
</tr>
<tr>
<td>Bonner v. Guccione, 178 F.3d 581 (2d Cir. 1999) 76:1036</td>
</tr>
<tr>
<td>Burton v. Neumann, 5 Wage &amp; Hour Cas. 2d (BNA) 1138 (S.D. Fla. Apr. 8, 1999) 76:528</td>
</tr>
<tr>
<td>Butcher’s Union Slaughter-House &amp; Live-Stock Landing Co. v. Crescent City Live-Stock Landing &amp; Slaughter-House Co., 111 U.S. 746 (1884) 76:708</td>
</tr>
<tr>
<td>B.S.W. Group, Inc. v. Comm’r, 70 T.C. 352 (1978) 76:252</td>
</tr>
<tr>
<td>Cannon v. Univ. of Chicago, 441 U.S. 677 (1979) 76:90, 221</td>
</tr>
<tr>
<td>Chisholm v. Georgia, 2 U.S. (2 Dall.) 419 (1793) 76:754</td>
</tr>
<tr>
<td>City of Boerne v. Flores, 521 U.S. 507 (1997) 76:281, 1263</td>
</tr>
<tr>
<td>In re City of Medina, 69 Wash. 2d 574, 418 P.2d 1020 (1966) 76:1227</td>
</tr>
<tr>
<td>Colegrove Water Co. v. City of Hollywood, 90 P. 1053 (Cal. 1907) 76:565</td>
</tr>
<tr>
<td>Collins v. Jordan, 110 F.3d 1363 (9th Cir. 1996) 76:640</td>
</tr>
<tr>
<td>Cornerstone Bible Church v. City of Hastings, 740 F. Supp. 654 (D. Minn. 1990) 76:1260</td>
</tr>
</tbody>
</table>

1299
<table>
<thead>
<tr>
<th>Case Name</th>
<th>Citation</th>
<th>Volume/Issue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corp. of Presiding Bishop of Church of Jesus Christ of Latter-Day Saints v. City of Porterville</td>
<td>203 P.2d 823 (6th Cir. 1983)</td>
<td>76:1259</td>
</tr>
<tr>
<td>Cort v. Ash</td>
<td>422 U.S. 66 (1975)</td>
<td>76:85,900</td>
</tr>
<tr>
<td>Creasman v. Boyle, 31 Wash. 2d 345, 196 P.2d 835 (1948)</td>
<td></td>
<td>76:1240</td>
</tr>
<tr>
<td>Daubert v. Merrell Dow Pharm., Inc.</td>
<td>509 U.S.</td>
<td>76:452</td>
</tr>
<tr>
<td>Daugherty v. City of El Paso, 56 F.3d 695 (5th Cir. 1995)</td>
<td></td>
<td>76:869</td>
</tr>
<tr>
<td>Davis v. Monroe County Bd. of Ed., 526 U.S. 629 (1999)</td>
<td></td>
<td>76:100,214</td>
</tr>
<tr>
<td>Davis v. Passman, 442 U.S. 228 (1979)</td>
<td></td>
<td>76:86</td>
</tr>
<tr>
<td>DeAngelis v. El Paso Mun. Police Officers Ass’n, 51 F.3d 591 (5th Cir. 1995)</td>
<td></td>
<td>76:227</td>
</tr>
<tr>
<td>Deuterium Corp. v. United States, 19 Cl. Ct. 624 (1990)</td>
<td></td>
<td>76:764</td>
</tr>
<tr>
<td>Donohue v. Copiague Union Free Sch. Dist., 391 N.E.2d 1352 (N.Y. 1979)</td>
<td></td>
<td>76:909</td>
</tr>
<tr>
<td>Dred Scott v. Sanford, 60 U.S. (19 How.) 393 (1857)</td>
<td></td>
<td>76:704</td>
</tr>
<tr>
<td>Duy Daec Ho v. Greene, 204 F.3d 1045 (10th Cir. 2000)</td>
<td></td>
<td>76:678</td>
</tr>
<tr>
<td>Echazabal v. Chevron USA, Inc., 226 F.3d 1063 (9th Cir. 2000)</td>
<td></td>
<td>76:859</td>
</tr>
<tr>
<td>EEOC v. Crown Zellerbach Corp., 720 F.2d 1008 (9th Cir. 1983)</td>
<td></td>
<td>76:967</td>
</tr>
<tr>
<td>EEOC v. Exxon Corp., 203 F.3d 871 (5th Cir. 2000)</td>
<td></td>
<td>76:869</td>
</tr>
<tr>
<td>Ellison v. Brady, 924 F.2d 872 (9th Cir. 1991)</td>
<td></td>
<td>76:225</td>
</tr>
<tr>
<td>Embrex, Inc. v. Serv. Eng’g Corp., 216 F.3d 1343 (Fed. Cir. 2000)</td>
<td></td>
<td>76:227</td>
</tr>
<tr>
<td>Erie R.R. Co. v. Tompkins, 304 U.S. 64 (1938)</td>
<td></td>
<td>76:118</td>
</tr>
<tr>
<td>Feiner v. New York, 340 U.S. 315 (1951)</td>
<td></td>
<td>76:641</td>
</tr>
<tr>
<td>Ferguson v. City of Charleston, 121 S. Ct. 1281 (2001)</td>
<td></td>
<td>76:434</td>
</tr>
<tr>
<td>Finkelstein v. Dep’t. of Transp., 656 So.2d 921 (Fla. 1995)</td>
<td></td>
<td>76:1243</td>
</tr>
<tr>
<td>First English Evangelical Lutheran Church v. County of Los Angeles, 482 U.S. 304 (1987)</td>
<td></td>
<td>76:1072</td>
</tr>
<tr>
<td>Ford Motor Co. v. Dep’t of Treasury, 323 U.S. 459 (1945)</td>
<td></td>
<td>76:165</td>
</tr>
<tr>
<td>Fry v. O’Leary, 141 Wash. 465, 252 P. 111 (1927)</td>
<td></td>
<td>76:549</td>
</tr>
<tr>
<td>Frye v. United States, 293 F. 1013 (D.C. Cir. 1923)</td>
<td></td>
<td>76:451</td>
</tr>
<tr>
<td>Fuller v. City of Oakland, 47 F.3d 1522 (9th Cir. 1995)</td>
<td></td>
<td>76:231</td>
</tr>
<tr>
<td>Gibbons v. Ogden, 22 U.S. (9 Wheat) 1 (1824)</td>
<td></td>
<td>76:1268</td>
</tr>
<tr>
<td>Gulf, Colorado &amp; Santa Fe Ry. Co. v. Ellis, 165 U.S. 150 (1897)</td>
<td></td>
<td>76:708</td>
</tr>
</tbody>
</table>
**Case Index**

<table>
<thead>
<tr>
<th>Case Title</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Petermann v. Int'l Bhd. of Teamsters</td>
<td>344 P.2d 25 (Cal. 1959) 76:182</td>
</tr>
<tr>
<td>Pierce v. Soc'y of Sisters</td>
<td>286 U.S. 510 (1925) 76:574</td>
</tr>
<tr>
<td>Plumstead Theatre Soc'y, Inc. v. Comm'r</td>
<td>675 F.2d 244 (9th Cir. 1982), aff'd 74 T.C. 1324 (1980) 76:253</td>
</tr>
<tr>
<td>Pierson v. Ray</td>
<td>386 U.S. 547 (1967) 76:161</td>
</tr>
<tr>
<td>Prince v. Massachusetts</td>
<td>321 U.S. 158 (1944) 76:574</td>
</tr>
<tr>
<td>Pierce v. Soc'y of Sisters</td>
<td>286 U.S. 510 (1925) 76:574</td>
</tr>
<tr>
<td>Rabidue v. Osceola Ref. Co.</td>
<td>805 F.2d 611 (6th Cir. 1986) 76:227</td>
</tr>
<tr>
<td>Reed v. Chrysler Corp.</td>
<td>494 N.W.2d 224 (Iowa 1992) 76:1204</td>
</tr>
<tr>
<td>Rhone-Poulenc Rorer, Inc. v. Home Indem. Co.</td>
<td>32 F.3d 851 (3d Cir. 1994) 76:1137</td>
</tr>
<tr>
<td>Russian Volunteer Fleet v. United States</td>
<td>282 U.S. 481 (1931) 76:1137</td>
</tr>
<tr>
<td>Schillinger v. United States</td>
<td>155 U.S. 163 (1894) 76:1091</td>
</tr>
<tr>
<td>Schneckloth v. Bustamonte</td>
<td>412 U.S. 218 (1973) 76:441</td>
</tr>
<tr>
<td>Schneider v. State</td>
<td>308 U.S. 147 (1939) 76:647</td>
</tr>
<tr>
<td>Schreiber v. Carney</td>
<td>447 A.2d 17 (Del. Ch. 1982) 76:798</td>
</tr>
</tbody>
</table>

| SEC v. Chenery Corp. | 318 U.S. 80 (1943) 76:1043 |
| Seward Chapel, Inc. v. City of Seward | 655 P.2d 1293 (Alaska 1982) 76:1259 |
| Sextant Avionique, S.A. v. Analog Devices, Inc. | 172 F.3d 817 (Fed. Cir. 1999) 76:1162 |
| State v. Koome | 84 Wash. 2d 901, 530 P.2d 260 (1975) 76:578 |
| State v. Mottman Mercantile Co. | 51 Wash. 2d 722, 321 P.2d 912 (1958) 76:1228 |
| Steiner v. Showboat Operating Co. | 25 F.3d 1459 (9th Cir. 1994) 76:227, 618 |
| In re Sumey | 94 Wash. 2d 757, 621 P.2d 108 (1980), 76:577 |
| Sutton v. Guild | 21 F. Cas. 554 (C.C.D. Mass. 1813) (No. 12, 391) 76:19 |
| Tenney v. Brandhove | 341 U.S. 367 (1951) 76:161 |