

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

A PROPERTY RIGHTS APPROACH TO
ANITRUST ANALYSIS

by
Timothy Dittmer

A dissertation submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

1998

Approved by 
Chairperson of Supervisory Committee

Program Authorized
to Offer Degree Department of Economics

Date 8/21/98

UMI Number: 9907891

**Copyright 1998 by
Dittmer, Timothy Paul**

All rights reserved.

**UMI Microform 9907891
Copyright 1998, by UMI Company. All rights reserved.**

**This microform edition is protected against unauthorized
copying under Title 17, United States Code.**


UMI
300 North Zeeb Road
Ann Arbor, MI 48103

© Copyright 1998

Timothy Dittmer

Doctoral Dissertation

In presenting this dissertation in partial fulfillment of the requirements for the Doctoral degree at the University of Washington, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for copying or reproduction of this dissertation may be referred to University Microfilms, 1490 Eisenhower Place, P.O. Box 975, Ann Arbor, MI 48106, to whom the author has granted "the right to reproduce and sell (a) copies of the manuscript in microform and/or (b) printed copies of the manuscript made from microform."

Signature 

Date 8/21/98

University of Washington

Abstract

A Property Rights Approach to Antitrust Analysis

by Timothy Dittmer

Chairperson of the Supervisory Committee

Associate Professor Keith Leffler

Department of Economics

In this dissertation I analyze a series of historically important antitrust cases, and demonstrate why the prevailing explanations are either incomplete or incorrect. By conducting fact intensive investigations of changes in demand, production costs, and ownership costs. I identify changes in property rights that explain the (alleged) illegal behaviors.

The first chapter presents many of the common property rights explanations of firm size, all of which relate ownership costs to production costs. My second chapter applies these ideas, very generally, to the changes in the US economy from the colonial period to the passage of the Sherman Antitrust Law. I argue that changes in shipping costs lead to the rise of both the modern corporation and antitrust laws.

The third chapter examines, in further depth, alternative ownership arrangements in the provision of canals. The fourth chapter dismisses the importance of predatory pricing in *Standard Oil of New Jersey*, and critiques the recent theory developed by Granitz and Klein. I propose an alternative explanation of Rockefeller's success, in which he exploits economies of scale in production by protecting quasirents from the railroads.

In the fifth chapter I argue that the formation of *US Steel* was not an attempt to create a dominant firm in the steel industry, but instead was protection for newly created downstream monopolies from future entry. The last chapter analyzes *United Shoe Machines* and its use of leasing, and presents existing property rights arguments that leasing was the optimal contract.

TABLE OF CONTENTS

LIST OF FIGURES	ii
LIST OF TABLES.....	iii
INTRODUCTION.....	1
CHAPTER 1: TRANSACTION COST EXPLANATIONS FOR THE FIRM	5
CHAPTER 2: THE DEVELOPMENT OF THE CORPORATION AND RESULTING ANTITRUST LAWS.....	31
PART I: THE DEVELOPMENT OF THE CORPORATION.....	31
PART II: THE DEVELOPMENT OF TRUSTS AND ANTITRUST LAWS	52
CHAPTER 3: THE CANALS	75
CHAPTER 4: STANDARD OIL OF NEW JERSEY	125
CHAPTER 5: THE UNITED STATES STEEL CORPORATION	173
CHAPTER 6: UNITED SHOE MACHINES	219
CONCLUSION.....	265
BIBLIOGRAPHY.....	268
APPENDIX A: EVIDENCE FROM AN ORIGINAL SOURCE.....	274
APPENDIX B: TECHNOLOGICAL INFORMATION RELATED TO CRUDE OIL AND REFINED PRODUCT.....	277
APPENDIX C: A CHRONOLOGY OF STANDARD OIL.....	280
APPENDIX D: MERGER LAW	284
APPENDIX E: THE DOMINANT FIRM MODEL.....	288

LIST OF FIGURES

<i>Number</i>	<i>Page</i>
Figure 1: Demand for Profitable and Unprofitable Canals.....	89
Figure 2: Demand with Average Cost	107
Figure 3: Demand and Fringe Supply.....	290
Figure 4: Residual Demand	291

LIST OF TABLES

<i>Number</i>	<i>Page</i>
Table 1: US Steel Market Share Over Time.....	194
Table 2: Alternative Explanations for Leasing.....	236

INTRODUCTION

This work is a series of interrelated essays in the application of economic theory to antitrust legal cases. My goal is explanation - making sense of behavior by demonstrating how it is consistent with the predictions of our theories. In this way I use both theory and economic history, but do not originate either.

I have selected antitrust cases for several reasons. First of all, the legal process results in the availability of large amounts of public information. "Data collection" has been a library, not field, enterprise.

Next, these cases already contain substantial analysis, for the opposing parties in the various suits provided alternative explanations for disputed behavior. This analysis was generally performed by lawyers, not economists, and is ripe for critique.

Finally, important legal cases tend to be dramatic. Large amounts of wealth were in dispute, as were the political careers of the prosecuting agents. The outcomes became legal precedents, and were therefore often more important than the existing statute law.

My work is not entirely based on cases. While the last three chapters are centered around cases the government brought against Standard Oil, United States Steel, and United Shoe Machines, the first three are oriented toward explaining the context for these cases.

The first chapter concerns various transaction cost explanations for why firms change the size or scope of their activities. In later chapters I will use various property rights arguments to explain the actions of firms. While I find these arguments interesting and compelling, they are also difficult to follow and often contradict each other. In this chapter I make sense out of these various explanations of the firm, and attempt to explain under what conditions they will be applicable.

The second chapter is a foray into economic history. Both antitrust law and the corporate form of business organization are relatively new. In chapter two I attempt to make sense of why each developed, and in doing so I introduce both the historical context for the later cases, and also the important antitrust statutes and early precedents.

The third chapter is the first to examine only one industry - the early US canals. These organizations were not involved in important antitrust cases, primarily because antitrust laws developed after their decline. But they are important in making sense out of antitrust, for they were some of the first major businesses organized as corporations, and they were almost

universally monopolies. The particular question I address in this chapter is this: under what conditions will monopoly assets such as canals be supplied by corporations, versus under what conditions will they be supplied by other forms of organizations such as governments?

The final three chapters center on important legal cases. I follow the same format for all three. First, I pose the important legal or economic question. Then I summarize the industry changes that were occurring at the time of the case. I believe that it is important to understand the larger technological changes occurring in an industry before examining the actions of a single firm. Finally I analyze the existing explanations for the alleged violations, and argue for what I believe is the correct analysis.

Standard Oil of New Jersey (chapter four) was not the first large corporation prosecuted under the Sherman Act. But its actions in the 1880's were instrumental in the Act's passage. The corporation was very important economically, and the case demonstrates many of the difficulties with antitrust enforcement. I will argue that the prevailing explanations for Standard's large market share are mistaken, and propose an alternative consistent with the material I present in chapter one.

Chapter five concerns the United States Steel antitrust case, and revolves around the creation of monopoly by merger. This is the case usually associated with the Dominant Firm Model of monopoly, and I will argue that this is the correct model, but only if applied in a manner different from the prevailing literature.

Chapter six, the final chapter, concerns the United Shoe Machines corporation. This case occurred much later than the previous two - in the 1950's and 60's instead of the turn of the century - and is not related to the formation of monopoly. Instead this case involves the maintenance of monopoly via contractual restrictions. I summarize many competing theories for these restrictions, and critique their application to the facts of this case. I argue for a transaction cost explanation, following that first present by Masten and Snyder.

CHAPTER 1: TRANSACTION COST EXPLANATIONS FOR THE FIRM

Firms, Production, Property Rights, and Transaction Costs.

Since the main agent in antitrust analysis is the firm, analysis presupposes some understanding of what a firm is and how it behaves. I define a firm as an institution (a series of enduring roles or relationships or contracts between agents) that organizes production. The firm acquires inputs from factor owners (labor, land, capital, ideas), and directs their transformation into some output. It has property rights over some assets, but is also the property of agents.

A firm is defined then, by its function; it organizes production. Production is the process of transforming some assets (inputs) into others (output). By this definition the process is both creative, since a new output is brought into being, and destructive, since at least some of the inputs are no longer available for other uses. The productive process may destroy an input in its entirety, as when a piece of sheet metal is molded into the side panel of a car door. Or only part of an input may be destroyed, as when a backhoe is rented for a day.

A firm must acquire the inputs into its production. Theft, violence and financial inducement are all common methods of changing the ownership of resources; while the first two are

fascinating topics, this analysis will concentrate on the voluntary exchange of property rights by either purchase or rental.

This exchange takes place in the market. Supermarkets are markets for retail food. Factories are markets for labor services. In each case assets (food, labor) are exchanged for money, and the prices of these assets perform the allocative role taught in introductory price theory classes.

There is then a distinction between production and exchange; production is the destruction of some assets to create others, and it always takes place within institutions called firms. Exchange is the transfer of assets from one owner to another, and if the exchange is voluntary we say it takes place across a market.¹

Exchange of assets does not mean moving them; it refers to changing who has control or property rights over them. To have property rights over an object means to expect to receive the benefits the object confers, either through direct consumption or in trade. The concept is forward looking; you have property rights over apples you plan to steal from your neighbor's

¹ Exchange is often surrounded by production. For example a stock broker would appear to only be involved in exchange - transferring stocks from one owner to another. But of course many assets are destroyed in this process - for example the stock broker's labor, resources used in communication, etc.

tree.² Property rights are, by this definition, generally incomplete. I only expect to receive part of the benefit of the apple tree in my front yard; my neighbors also have some property rights over the tree since they are likely to take some of the ripe apples.

Property rights are incomplete for reasons other than theft. Assets almost always have many valued “dimensions.” For example, I value both the weight and ripeness of bananas. If my grocer sells bananas by the pound, and does not sort them by ripeness, then I only pay him for the weight of any bananas I purchase. If I select those with only optimal ripeness - and pay the same per pound price as those overripe (I like them a little green) - then I have effected a transfer of wealth from the grocer to myself without me paying for it. By charging the same price for ripe and overripe bananas, and selling them by the pound, the grocer has given up property rights to the dimension of ripeness in his bananas. Selling by the pound means the grocer did not exercise “complete” property rights. Note that, in this example, it is the cost of measurement that limits property rights.

Markets generally lower the cost of buying and selling; they work so well that many economists assume they are free. But, as the previous example demonstrates, transacting in a market is generally not free. If it were, then the grocer would price every dimension of value for all of his products, and the associated measuring would be costless. One goal of this

² This definition and example are from Yoram Barzel.

work will demonstrate that the assumption of costless transactions is often inappropriate for antitrust analysis.

Transaction costs is another phrase, like property rights,³ that has been used with different meanings. Following Doug Allen,⁴ I will define transaction costs as any expenses incurred by an individual in order to expand or protect his property rights over an asset. Putting a fence around my apple tree expands my property rights to the future benefits of the tree. Measuring the important dimensions to an asset before its sale is another transaction cost.

Transaction costs, by this definition, are not synonymous with the costs of purchasing an asset through the market. Shipping costs, for example, are not transaction costs, but are included in the cost of market transactions.

I claim that a firm “owns” various factors of production - that it has purchased them from others. But how can something that is not a person, like a firm, have property rights? If to have property rights over an asset means to expect to receive the future benefits from it, how can anything that is not a human have expectations. This problem, that a non-person seems

³ The term property rights has two distinct meanings. The more normative idea refers to what people ought to be allowed to possess - what they have a “right” to. Legal rights are an example of this usage. You may have complete legal rights to a twenty dollar bill that falls out of your pocket - but if you fail to notice that it fell out, you may very well only have legal rights.

⁴ Allen, 1991.

to have property rights, disappears when we understand the firm as an *incomplete share contract*.

It is a share contract in this sense. When two business partners each contribute a million dollars to the purchase of a two million dollar building, each has exchanged his property rights over money for a share in a firm. Each does not own half the building (one the west side and the other the east side); the partnership owns the building, and each partner owns half of the partnership. A firm is a collection of jointly owned assets used for production. The expectations belong only to humans.

It is an *incomplete* contract, however, in this sense. Since property rights are incomplete, the “boundary” of the firm is never perfectly precise. For example, if you are hired to sweep floors for a firm, you are renting your labor to the firm. Your labor becomes the property of the firm for the contracted duration and subject to various contractual constraints. While the firm would like to buy your labor as an input into the production of clean floors, your labor can not be separated from the rest of you; it is not a commodity that may be transferred. While at work you still have considerable control over your labor (within the contract), and may use that control to produce new assets for yourself (like clean fingernails, or leisure).⁵

⁵ This ideas are influenced by Leffler 1986.

Who then really owns the labor you sold to the firm? Property rights over it are not perfectly defined. The formal owners of a firm (e.g. stockholders in a corporation) are not the only parties with property rights. Employees who have formally given up their property rights to their own labor still retain *de facto* control over the output of the firm, and expect to benefit from this control.

We will use these ideas, that measurement costs make property rights incomplete, and that transaction costs are incurred in order to make them more complete, to explain why differently sized and organized firms have different cost of production.

Firm Size and Costs

What determines the size and scope of a firm? Why is it that in the United States there are not two hundred and sixty million firms (the approximate population) or only one firm? Each is possible, since each factor owner could either form his own firm or instead perform production only for someone else's firm.

The answer, of course, is that for any given type of production, the firm size that maximizes profits is unlikely to be one employee or 260 million. Profits to factor owners will change with firm size for two related but distinct reasons; changes to cost, and changes to revenue. After all, profit is the difference between total revenue and total cost.

Antitrust economists often evaluate business practices using analogous, but not synonymous, categories. *Efficiency* explanations attempt to describe a business practice as a method of lowering costs. *Monopoly or Market Power* explanations appeal to raising revenue by somehow decreasing the elasticity of demand (e.g. eliminating rival suppliers). From the standpoint of business owners, a dollar earned via cost reduction is equivalent to a dollar earned by demand enhancement. Profit maximization implies that firms will take all actions that increase profits - that a firm is just as "greedy" when tries to lower cost as when it limits entry.

Alternative Costs of Alternative Ownership Arrangements

How then, does the cost of production vary with firm size? One tempting, but inadequate answer is to appeal to technological factors in production. For example, it may be that production costs are such that one large factory will produce the same quantity at a lower total cost than ten factories of one tenth the size.⁶ But this is not an adequate explanation of firm size, since the firms organizing production could own multiple factories, or perform their productive functions by renting only a small section of the largest factory. Firm size is an ownership question, not a technological scale question.

⁶ This is an example of economies to scale.

Transaction cost economics (often called property rights theory) attempts to explain and understand ownership costs. The next few sections describe various transaction costs and how they relate to firm size. All of them give, for some range of output, explanations of why the cost of production decreases if it occurs within one firm instead of multiple smaller firms (economies of scale in ownership). But we will first start with a strong argument for why there should be diseconomies of scale in ownership.

Moral hazard is a form of inefficiency that results from non-marginal payments. For example if I sew buttons on shirts for you, and I average one hundred per hour, you could equivalently pay me five cents per button or five dollars per hour. But if there is both random variation in my output (working equally hard I sometimes produce 90 and sometimes 110) and if monitoring my effort is excessively costly, I have an incentive to "shirk" - that is work less hard than we agreed - if you pay me by the hour. When ever it appears that I will produce 110, I could simply take a break and produce 90.

If I as a factor owner have my own firm, I have no incentive to shirk on the number of buttons I produce. I am paid for my output, so I receive less income if I produce less output. But if I work for you (you rent my time), you must monitor me to ensure that I am not shirking. This is an ownership cost of moral hazard - a transaction cost.

Moral Hazard is a force that creates diseconomies of scale. I, as a button maker, have a certain cost curve for making buttons. Presumably it is “U” shaped - average cost declines as I expand production because I am able to use more specialized tools. Eventually average (and of course marginal) cost rises as the factors of production have increasing opportunity cost (e.g. I work eighty hours a week instead of seventy).

If you pay me by the hour to make buttons, you may (approximately) take my average cost curve, and add to it the additional cost of shirking. Combining my cost curve with other similar workers will create a cost curve for you - but the minimum average cost will be larger than my minimum, since you have taken my costs and added shirking and monitoring costs.

The reader may object that economies of scale may outweigh shirking costs. Think of Adam Smith’s pin factory example. A single maker, working alone, could only make a few pins per day. When we combine workers, and have each specialize in one task, the total production is many times the sum of the single workers.

But this argument does not work, for each pin maker may specialize without being part of a firm. The worker that just makes pin heads may buy the upstream product, add the head, and sell his output to the downstream producer. Adding a firm does nothing. There may in fact be a reason why costs are lower if the workers are paid by the hour instead of the pin (I will provide these arguments latter), but it is not because of specialization in production.

What if we add a second layer of monitoring? If you, as a monitor, also have a “U” shaped average cost curve because of scale economies then diseconomies, you may wish to expand production by hiring others to do the monitoring, and spending your efforts monitoring the monitors. But then the minimum average cost will be the sum of the button maker’s cost, the lowest cost of monitoring, and the cost of monitoring the monitors. There may be economies of scale holding the number of layers of monitors constant, but adding layers will increase minimum average cost.

In this way, the average cost of production should increase with the number of layers in the hierarchy. This ignores all other transaction costs, and assumes that we evaluate average cost at its minimum point per level. For this reason, moral hazard is a force creating diseconomies of scale.

Yet what we observe is that most income comes from labor, and most laborers work for firms. So while moral hazard is a strong force for small firm size, there must be stronger economies of scale.⁷ But what are they?

⁷ Moral hazard in labor occurs because the valuable margin - effort - is not priced. Other inputs have analogous difficulties. For example a firm buys steel as an input, and pays per unit of mass. But variations in quality can also occur, so again the firm may not receive what it contracted for. This problem leads to different costs of different firm sizes.

Transaction Costs

In his 1937 paper *The Nature of the Firm* Ronald Coase was the first to ask these questions. He did not start with our definition of the firm - as the institution in which production occurs - but instead viewed the firm and market as alternative institutions for the organization of production. But his informal model of the firm still yielded an understanding of what determined firm size.

Coase spent considerable time and effort speaking to business managers, trying to discover why they decided to produce some products themselves, and purchase others through the market place. In order to produce some good (for example an input into one of the firm's products), managers would have to incur various organizational costs, including training, hiring and monitoring employees, purchasing equipment, etc. These organizational costs are presumably increasing for all of the same reasons marginal costs eventually increase. At some point the cost of producing another good within the firm exceed the costs of purchasing from another firm.

The cost of purchasing a good is greater than its price. There are various transaction costs associated with purchasing goods across markets, including such factors as ensuring you

receive the quantity and quality you contracted for.⁸ Presumably these costs increase as additional transactions occur across markets.

At some point the cost of another transaction through the market just equals the cost of organizing it within the firm. This point describes production efficiency, and this determines the size of the firm. Anything that increases the cost of using the market will increase firm size, and the converse. This was Coase's famous Transaction Cost analysis.

There are two difficulties with his analysis. First off all, Coase was answering the question, "why do we have firms at all when the market could organize production." Given my definition of both the firm and production, this question is not relevant. The more important difficulty is that Coase does not come up with specific transaction costs that vary with firm size. Since the nineteen thirties economists have begun to provide some of these costs.

Team Production

An important cost that results in economies of scale in ownership comes from Armen Alchian and Harold Demsetz in their 1972 paper *Production, Information Costs, and*

⁸see Doug Allen's paper for a more precise definition of transactions costs.

*Economic Organization.*⁹ They hypothesize that *team production* is the distinguishing attribute of production within firms.

If I sew on buttons for you (as described above) then moral hazard results diseconomies of scale. But if we change the example to where button sewing is a team production involving a number of us, then paying me based on my output (per button) often does not work. This is because with some productive processes, each worker does not produce any output by himself; the team works and together they produce output.

For example a fishing boat produces fish.¹⁰ But how many fish does the winch operator produce, versus one of the deck hands? Neither produces fish separately, and the actions of one change the output of the other in ways that are expensive to measure.¹¹ So it is not possible to pay the winch operator based on how many fish he produced, since he did not produce any by himself. It is possible to pay factor owners for a share of total (not

⁹AER, 62(5), 1872.

¹⁰ There are, evidently, several entirely different ways in which boats are used to catch fish. My example seems to confuse several of them.

¹¹The output of both the deck hand and the winch operator is the movement of a net full of fish into the boat. More fish may escape from the net if either worker shirks.

individual) output. Share contracts (splitting the total output) decreases the incentive to shirk, but as the number of team members increases, the incentive to shirk again increases.¹²

The normal result of team production is that the owners of the various factor inputs are paid based on their input (their time), not output. This, with variability in output, necessitates monitoring to control shirking. Thus team production leads us to the firm, with its per hour wages and monitors.

The classical firm is, for Alchian and Demsetz, a form of contract with the following characteristics:¹³ “(1) joint input production; (2) several input owners; (3) one party who is common to all the contracts of the joint owners; (4) who has rights to renegotiate any input’s contract independently of contracts with other input owners; (5) who holds the residual claim; and (6) who has the right to sell this central contractual residual status.”

The classical capitalist firm has an owner/manager who plays the role of monitoring inputs. In order to prevent the monitor from shirking, he is paid not on the basis of his input (time), but instead owns the variability in team output. If the team produces more than the payments to input owners, the owner profits. If the team produces less than the payments to input

¹² If my shirking reduces total output by one fish, then my cost in if I have one partner is half a fish. With nine partners, my cost is one tenth of a fish. Thus share contracts, while reducing the moral hazard contract, are not powerful for large teams.

¹³Demsetz p. 137.

owners, the owner must make up the difference. Since shirking comes out of the pocket of the owner, he is the ideal *residual claimant*; he has the proper incentives.

The size and scope of firms will be limited by team production. Firms will become larger as technology is developed that requires team production. This technology will have cost advantages that outweigh the cost of input monitoring. When technology is developed that reduces the requirements for teams, firms will decrease in size.¹⁴ An example might be transportation firms. Before the development of the railroad, transportation firms were relatively small due to the prevailing technology, horse driven wagons. Moral hazard considerations resulted in low cost production at the owner/operator size. With the advent of the railroad, drivers were not paid for their marginal output, since they produce no output independently of the rest of the transportation team. Thus firm size increased. Finally, the development of the interstate highway system and tractor trailer reduced firm size.

This is essentially a measurement cost issue; when measurement of output is "too" expensive, inputs will instead be measured and paid for. It is an important refinement of Coase, in that it explains with far greater precision what costs firms minimize. Finally, it explains the particular structure of firms (owner/managers), and makes predictions as to how firm structure changes as cost change.

¹⁴ Likewise, an increase in wages will result in more expenditures to avoid shirking, including direct measurement of marginal product.

Asset Specificity

Another organizational cost resulting in economies of scale is asset specificity.¹⁵ This hypothesis, created by Williamson, claims that assets that have low salvage value when separated will be owned by a single party. This is best understood by way of an example.

Suppose oil is discovered on property A. Furthermore, suppose the only feasible method of transporting the oil is by pipeline through the property of B to a harbor. The owner of A could pay the owner of B to build and operate a pipeline. But, after development, the costs of both the pipeline and oil field will be sunk; they will not have any salvage value. This presents the owner of each with an opportunity for holdup; B knows that A's costs are sunk, so B can hold up A for the difference between the full price of the oil and A's marginal costs. Of course in this example the opposite is also true; B's costs are sunk, so A need only pay the operating costs of the pipeline.

Since each party knows it is liable to lose the full cost of its investment, neither will incur the cost. Therefore in order for the oil field and pipeline to be constructed, some guarantee to each party must occur.

¹⁵See for example Williamson 1975.

Well delineated property rights and enforceable contracts help alleviate this problem. But ownership of both assets together is often the low cost method of solving the problem of asset specificity. This theory then claims that dedicated assets will tend to be owned by the same party. And since the wealth of labor and capital owned by most individuals is extremely limited, this leads to the use of firms. A business firm allows separate owners to pool their resources in such a manner that each asset is owned in common.

Firms will tend to expand to purchase assets that become dedicated or specific to their operations. Conversely, as other suppliers of inputs enter the market, it seems likely that specialization will decrease the scope of a firm's operations. This theory should be useful in explaining why in certain circumstances firms vertically integrate so as to "guarantee" their source of supply.

Measurement Costs

Another cost savings of larger firms has to do with the cost of measuring the attributes of a good. Yoram Barzel argues¹⁶ that when the attributes of a good are not measured, they will

¹⁶Barzel 1989, p. 59.

be produced and consumed as if they were free goods.¹⁷ When transactions occur between firms in sequential production, each good must be measured at each stage of production. But a vertically integrated firm will perform each step of the productive process by command, not price. So it is only necessary to measure the good at the first and last stage of the firm's production.

Suppose, for example, a cotton dress is made by first making fabric out of raw cotton, then dyeing the fabric, then cutting and sewing the fabric into its final form, and finally marketing the dress. And suppose each stage of production is performed by separate individuals transacting anonymously across markets.

The fabric maker must evaluate the quality of the cotton. If they do not do so, cotton producers will substitute lower cost substitutes, since they will be paid the same whether they produced expensive or inexpensive cotton. Once the cotton is processed into cloth, the dyer must evaluate both the quality of the weaving *and* the quality of the component cotton. If the dyer only evaluates the quality of the cloth, then the fabric maker has an incentive to buy low quality cotton and pass off the cloth as composed of high quality cotton. Note that the cotton has now been evaluated three times; once by the cotton farmer, once by the cloth maker, and once by the dyer.

¹⁷i.e. not produced at all or consumed until the marginal value is equal to zero.

Of course there are market institutions and practices that minimize this repetitive measuring. Statistical sampling techniques and the reputation of firms are two obvious responses. But when neither method is the cost minimizing option, vertical integration also decreases the incentive to overmeasure.¹⁸ It should be noted that this effect is opposite to the diseconomy of moral hazard, which arises from “too much” monitoring within a firm.¹⁹

Another cost of smaller firms, closely related to the theory provided by Demsetz and Alchian, also concerns measurement costs. One method for workers to use equipment valued beyond their wealth is for them to rent it. Shirking in road construction firms could be avoided if workers rent backhoes, dump trucks and graders from individuals wealthy enough to own such equipment.

But if it is difficult to measure wear and tear on equipment, and workers rent equipment by the hour, then they have an incentive to try to attain as much work as possible out of the equipment, and without regard to unmeasurable damage. Paying workers by the hour, instead of for how much they produce, decreases their incentive to over use equipment (or do any work). In the case of valuable equipment, for which it is difficult to measure wear and

¹⁸The production of Soviet factories, all of which were in once sense part of a large firm, were of notoriously poor quality. This explanation is obviously incomplete for this case.

¹⁹ As noted by Keith Leffler, vertical integration becomes more likely as the cost of measuring purchased inputs rises relative to the cost of monitoring labor inputs.

tear, the loss due to workers' shirking may be less than the loss from over using equipment. This gives us another cost advantage of paying individuals for their inputs instead of their output, and therefore why firms using expensive equipment may have cost advantages over firms with smaller output.

Wealth Constraints

Measurement cost was Barzel's first contribution to the theory of the firm. He developed another related explanation based on wealth constraints that contradicts two elements of the received wisdom; first, that businesses purchase insurance because of risk aversion, and second, that corporations exist in order to achieve limited liability.

For some goods, in some settings, the purchaser is uninterested in the identity of the producer. For example in the commodity exchanges, those parties buying a spot contract of wheat do not care which farmer is supplying the product. Identity is unimportant, since the commodity exchange itself (a very large monopolist) guarantees the product will meet certain specifications.

But in many transactions the identity of the purchaser is important, since the properties of the purchased good are not entirely identified at the time of sale. If the parties to the transaction remain anonymous, then we expect the purchaser to discount the quality of the good, since

moral hazard²⁰ implies that sellers will underprovide quality if it is not being measured. The buyer will then only be willing to pay for lower quality output, even if it is of high quality. Two factors that limit this are reputation and guarantees.

In order for buyers to be willing to pay for high quality output, they must be guaranteed not the difference in selling price between low and high quality output, but instead compensation for any damages from low quality when they thought they were purchasing high quality. Suppose, for example, that you are purchasing tires for your automobile. You may be willing to purchase low quality tires, but you will then drive only at slower speeds, etc. If you pay more for what you think are high quality tires, and then discover they are low quality (when you are traveling at 65 mph on a slick road), you will demand compensation of your damage (a wrecked car) not just the difference in price. If the guarantee that comes with the tires only pays up to the difference in price between what you thought you were getting and what you in fact received, you will not buy the high quality tires.

To Barzel, most goods have this problem. Almost all inputs into productive processes can cause damage if they are below specifications. Therefore they must be guaranteed. But my having a comparative advantage in the production of a good does not imply that I also have sufficient wealth to guarantee my output.

²⁰Or in some cases adverse selection.

Simple insurance does not solve this problem, since if my product is insured by someone else, my incentive to produce high quality is further reduced. Rather the only way that another party with sufficient wealth can guarantee my output is to alter the incentives I face as a producer.

The owners of a firm guarantee the output of all of the hired factors of production. They pay based on inputs (time, etc.) instead of outputs, thus eliminating the incentive to economize on quality. But in order to produce any output, the owners must monitor the hired workers. This both ensures the quality of output, and forces workers to do any work at all.

The limits of the firm are determined by how much output the wealth constraints of the owners can guarantee. Firms stop growing due to diseconomies of scale produced by paying factors based on their inputs (moral hazard), as contrasted to economies of scale due to the distribution of risk over larger amounts of guarantee capital.

For this reason it is not limited liability that drives the firm, but instead guaranteed liability, up to the wealth of the firm. From an individual shareholder's standpoint, limited liability is a benefit of investing in a corporation. By diversifying investments across a number of corporations, a catastrophe for one investment will not result in a loss of capital in other investments. But what increases demand for the corporation's product - and limits its size - is the guarantee capital embodied in the firm.

An example of this process is shipping insurance. Major US oil companies are both financially large and vertically integrated, performing tasks from exploring for oil to retailing gasoline. We would not expect these firms (or more precisely their owners) to be risk averse, since they are large enough to diversify away any single risk. Yet many of them buy shipping insurance for their oil tankers.²¹

Why would they buy insurance, if not because of risk aversion? Barzel's answer is that while they might have a comparative advantage in the production and distribution of petroleum products, it is unlikely that they are also specialist in maritime safety. Therefore they hire specialists.

These specialists may be capable of providing advice that reduces losses due to shipping accidents, but it is extremely difficult for the oil companies to know if they are receiving good advice (if they knew it was good advice, then they would be the specialists). If the shipping specialists are not being paid on the margin for the quality of their advice, they have an incentive to underprovide. The solution is then for the shipping specialists to guarantee the quality of their advice; they do so by insuring the oil companies' ships.

²¹ As reported by Yoram Barzel in personal conversation.

If a shipping specialist discovers procedures that reduce the rate of accidents (at a cost lower than their benefits), it can increase profits by offering lower rates if the oil companies to follow them. The importance of shipping insurance then is not merely the risk pooling; it is a way in which specialists can guarantee their product. Effectively the insurance company owns the oil companies' damages.

Political Explanations of the Firm.

Antitrust and other political considerations also change the cost of owning different size firms. For example firms wishing to price discriminate between two sets of customers may be constrained by the Clayton act. Under certain circumstances this may be avoided by purchasing one set of customers (integrating vertically).

In another example, the Federal Government claims, in its merger guidelines, that it will challenge mergers when the price of a good may increase by five percent or more.²² In the case of crude oil production, where the price of crude in the oil fields is only a few cents per gallon, a one cent increase would be far more than five percent. But in the case of gasoline

²² The merger guidelines actually state, "In attempting to determine objectively the effect of a 'small but significant and nontransitory' increase in price, the Agency, in most contexts, will use a price increase of five percent lasting for the foreseeable future. However, what constitutes a 'small but significant and nontransitory' increase in price will depend on the nature of the industry, and the Agency at times may use a price increase that is larger or smaller than five percent."

retailers, who sell gas at prices between a dollar and a dollar and a half per gallon, a one cent increase would be far less than five percent.

This is the case even though the actual markup above a dealer's marginal cost is far less than a dollar; it is closer to ten cents per gallon. But since the Justice Department only²³ looks at the price increase - not the percentage of margin increase - a merger of two crude oil producers might be challenged while a merger of similarly situated retailers would not be challenged. This and other quirks of government policy change a firm's cost, and therefore its size.

Summary of Cost Explanations of the Firm

When factor owners (i.e. workers) are paid for their difficult to measure inputs, like time and effort, they have an incentive to shirk. This implies that efficient production will occur when factor owners are paid for how much they produce. But much of production occurs within firms, where workers are paid by the hour. We need economies of scale in ownership to counter this diseconomy.

²³ The quote in the previous footnote indicates that there is ambiguity as to how the Department of Justice will act in these cases.

There are a number of theoretical reasons, most of them relating to measurement, why there are economies in ownership. Team production makes measuring individual output excessively expensive. Asset specificity requires one owner of multiple resources. Chains of vertically related producers face the cost of repeat measurement. Renters of equipment have an incentive to overuse it when measuring damage is expensive. Finally, suppliers have limited ability to guarantee their own output, leading to specialization in guarantee capital. All of these result in economies of scale in ownership which overcome the diseconomies due to moral hazard.

CHAPTER 2: THE DEVELOPMENT OF THE CORPORATION AND RESULTING ANTITRUST LAWS

Antitrust laws apply to all businesses - you will go to jail for price fixing whether you are the president of a corporation or a sole proprietor. But the phenomena of antitrust is really an outgrowth of the corporation. They developed and expanded at the same point in history, and the focus of antitrust enforcement has always been medium and large scale corporations.

For this reason, if we are to understand the development of antitrust laws and analysis, it is important to understand something of the history of the corporate form. This chapter has two parts. Part one traces and explains the development of the corporation in the US. Part two explains the rise of antitrust laws in reaction to the corporation, and presents the important early laws and court decisions.

PART I: THE DEVELOPMENT OF THE CORPORATION.

At the beginning of the eighteenth century, corporations were rarely used forms of business organization. By the beginning of the nineteenth century, they were the predominant form.

What brought about this change?

While there were many new inventions in the area of corporate management - for example modern accounting was developed by the railroads - appealing to new ideas is not a satisfactory explanation for the development of the corporation. For why were these ideas invented in the eighteen hundreds, and not the seventeen hundreds?

My explanation for the corporation's development is that it has cost advantages under certain conditions, and that these conditions first appeared in the nineteenth century. In particular, certain large scale transportation systems were most efficiently owned by corporations. Demographic and political changes made these transportation systems profitable, and the corporate form was developed in response. The resulting decrease in transportation costs created other changes, which in turn made the corporate form of organization efficient for many other industries.

The Economy of the United States circa 1800

While historians have traced the origin of the United States governmental regulation of business back at least as far as the early English Common Law, we will start our story somewhat midway, at approximately 1800.

The United States economy at that time was drastically different from today. It is sufficient to note several major differences:

1. Most of the population of European origin resided on the Eastern seaboard. Relatively inexpensive sea transportation existed both between port cities and with European shipping centers. Land based transportation was dramatically more expensive, with transportation costs ten times as much to ship from Pittsburgh to Philadelphia as from Philadelphia to London.²⁴

2. The United States consumer market was relatively small. In 1790 the population was approximately 3.9 million total, with 3.7 living in rural communities. There were no cities of 50,000, two between twenty five and fifty thousand, three between ten and twenty five thousand, seven over five thousand, and twelve over two thousand five hundred.²⁵ A significant portion of the population did not participate in the money economy.²⁶

3. Land was relatively cheap (as compared to European countries), and labor was very expensive. This resulted in land intensive economic activities, with little capital intensive manufacturing.

²⁴Mountfield p. 111

²⁵North 1961, p. 17.

²⁶North 1961, p. 18 writes: "There is no way to tell with any degree of precision what percentage of the rural population was producing for the market, and were themselves a regular part of the domestic market. Clearly a large percentage did not regularly engage in market production. For still a larger number, the contribution to the market production was peripheral and an irregular supplement to a way of life largely self-sufficient."

Larger manufacturing firms did not exist, both due to the scarcity of non-land capital, and the low cost of sea based importation from England. The United States economy produced goods in which it had a comparative advantage, and small firms most efficiently produced the resulting agricultural and raw material goods.²⁷

In addition to agriculture, the early United States economy had intermittent success with shipping. International politics drove the ebb and flow of shipping interest,²⁸ but as compared to the modern corporation, firm size was small. This will be explained by appealing to moral hazard (see below); with the extreme uncertainty of success and poor communications, residual claimancy was best left in the hands of those who could most control the outcome. Therefore ship captains and closely monitoring owners, not diversified shareholders, owned the resources they managed.

Regionally, the economics of transportation differed greatly. Once away from the seaports, the North East had relatively poor river transport. In contrast the South had navigable rivers

²⁷North 1966, p 38.

²⁸North 1966 chapter 5, especially p. 66. This was during the era of the Napoleonic wars in Europe, the location of the United States' most important trade partners. U.S. shipping trade boomed during periods in which the combatants recognized the rights of neutrals, since France attacked British shipping and Britain attacked French shipping, but neither preyed upon neutrals. At various other times Britain embargoed all U.S. shipping, leading to its virtual elimination (and the War of 1812).

far into the interior.²⁹ This resulted in more specialized market production in the South, and with it somewhat larger firm (farms and plantation) size. Low cost river trade routes from the Midwest to the South through New Orleans reinforced this trend. Effectively the Midwest fed the South with foodstuffs as the South specialized in production of commodities for international markets.³⁰

Small Firm Size as the Result of High Transportation Costs and Moral Hazard

The listed facts lead to small firm size for two related reasons. The first is expensive transportation. With *inexpensive* transportation, it is possible for large areas to be served by one large plant. With increasing returns to scale,³¹ firms with larger sized plants have cost advantages over smaller sized firms. When firms of differing sizes are engaged in price competition, the firms with lower average cost will presumably eliminate smaller rivals. But with *expensive* transportation, the cost savings resulting from economies of scale may be outweighed by shipping costs. Since land based transportation was very expensive (relative to by sea transport and subsequent railroad transport), large plants did not serve inland markets.³²

²⁹ North 1961, p. 125-126.

³⁰North 1961, p. 102.

³¹There are increasing returns to scale when the average cost of producing a unit decreases as a given plant produces larger quantities.

³²O'Brien, p. 20

This common argument for small firm size in the face of high transportation is incomplete without an additional element. The problem is that the argument confuses cost curves of plants with cost curves of firms.^{33,34} They are not equivalent, for cost curves of plants are not a subset of cost curves of firms, and cost curves of firms are not simply the summation of those of plants.

Suppose, for example, that small oil refineries operate at a higher average cost than large refineries (increasing returns to scale). If initially transportation costs are high, it may be the case that transportation costs are larger than the savings from having large refineries, and therefore smaller local refineries will produce at a lower net cost - they will win in price competition against their larger counter parts.

But this does not necessarily imply that oil producing *firms* will be small, for it may be the case that there are economies of scale in the ownership of refineries,³⁵ and therefore one firm

³³See the section on the nature of the firm in the first chapter.

³⁴For an example of this confusion, North 1966, p. 37 writes; "Economies of scale means that as the output of a firm increases, its cost per unit of output fall. The growing size of the market made possible larger firms. The optimum size of the firm changes with technology and has generally grown larger with modern technology."

³⁵ For example, since entrepreneurial talent is a particularly scarce resource, one superior manager may cost less than many local inferior managers. Economies of scale must be discovered by fact and experience intensive research in particular industries, and do not seem to be the type of knowledge suitable for generalizations.

will own many refineries, each in its own local market.³⁶ Economies of scale related to transportation only are important to the size of plants, not the number of plants owned by a firm. We need another element if we are going to argue that high transportation costs lead to small firm size. That element is moral hazard.

Moral hazard is, in this case, the incentive a worker has to underproduce (shirk) when he is not paid for how much he produces. When a worker (or other factor owner) sells output directly to a market, that worker will produce every unit in which the marginal cost is less than the price (assuming he is a price taker). No monitoring of the worker is necessary. The worker does not have an incentive to slack.

But if instead we increase firm size from one worker to one hundred, and make one of them the “owner”³⁷ of the firm and the rest employees, then the other 99 now have an incentive to reduce their efforts. If the workers’ pay now is hourly, and does not change with the amount they produce, then they will presumably produce less than the number where $P=MC$. This assumes there is both natural variation in output and a cost to monitoring.³⁸

³⁶ Chapter one provides reasons why large plant size will be accompanied by large firm size, but in this case high transportation costs limit plant size, making these factors unimportant.

³⁷ The new owner “rents” the other labor inputs on an input measured basis and owns the intermediate output.

³⁸ This is phrased in quantity terms. If quantity is easy to measure, and quality is expensive to measure, then the “shirking” will occur in the quality dimension.

As firms grow larger, hiring more workers and producing more output, a larger number of managers must be hired to monitor the additional workers. But if these managers are also paid employees, the same moral hazard condition exists, and the total cost of management should increase. Moral hazard is one reason why there should be diseconomies of scale in firm size. High transportation costs lead to small firms because plant size will be small and moral hazard gives higher costs to multiplant firms than to locally owned plants

The second related fact supporting small firm size is communication costs. Early nineteenth century America did not have telephones or telegraphs, paved roads or automobiles. Owners had to monitor their employees by being on location, which limited the number of plants a single owner could monitor. High communication costs exacerbated moral hazard costs, and resulted in smaller firms.

Moral Hazard and Firm Size in Shipping Firms.

Each of these explanations (small market size and expensive monitoring) lead to small sized shipping firms. If transactions were homogeneous, as occurs with a specialized firm in a large market, there exist economies of scale to monitoring. The residual claimant or owner may create one set of rules for his fixed wage employees to follow. The owner's monitoring expense is then simply after the fact auditing; no further direction is needed. This monitoring

has decreasing average cost, since the one fixed cost (creating the rule) is divided by increasing numbers of transactions.

But with small markets, as prevailed in the early nineteenth century, transactions were not homogenous. Managers of ships had to make decisions as to which goods to buy and which markets to sail to. These decisions were constantly changing with relative prices. One set of rules would not have been sufficient, and a fixed wage employee did not have the same incentives as the owner of the ship. For these reasons we did not see large, multiship transportation firms, but instead ship managers were either more closely monitored or were themselves owners.³⁹

The Rise of Low Cost Transportation.

In 1825 the Erie canal opened, linking New York City and Buffalo. This reduced freight Charges between Buffalo and NYC from \$100 per ton to \$5, and transit time changed from twenty days to six.⁴⁰ The canal was build by the State of New York, and financed by New York State Bonds. A number of feeder lines were built by private corporations created under special charters.⁴¹

³⁹ See the appendix to this chapter for evidence both for and against this claim.

⁴⁰Mountifled, p 112.

⁴¹Seavoy p. 43.

The initial change brought about by the opening of the canal was the obvious decrease in transportation costs in upper New York State. Farmers before this time were generalists, since the poor quality mud roads precluded shipping any goods other than those of the highest value per weight. So instead of producing specialized goods for the market, early farmers spent more of their time producing goods they used themselves.

The Erie canal allowed these same farmers to exploit comparative advantage by producing a smaller number of specialized goods in greater quantity, and sell them to the New York City markets, where they could be exported by sea to anywhere in the world. Of course this was accompanied by movement in the opposite direction; goods that were less expensive to produce elsewhere now became viable alternatives for locally manufactured goods. Economies of scale in manufacturing and farming became accessible, which again changed the prevailing firm size.⁴²

The opening of Western New York was an economic boom to the shipping facilities of New York City. For the west end of the canal at Buffalo allowed access to Lake Erie, which led to direct shipping to the other great lakes. With relatively minor improvements, the system of rivers spanning the Mississippi became linked to this transportation net. The low cost transportation allowed the rapid opening of the productive agricultural land in the West

⁴²See O'Brien p. 20

(now Midwest) and changed the demographics of the early colonies. In 1810 the Western states had 15% of United States population; by 1830 this skyrocketed to 28%.⁴³

The Limits of the Existing Business Form.

For the reasons I have discussed, early American business firms were small. They were also, almost without exception, either sole proprietorships or partnerships. The corporate institutional structure did exist, but exclusively for quasi-governmental, religious, educational or charitable organizations.⁴⁴ The early American states did not even possess general incorporation laws for businesses.

The existing legal institutions were not “primitive” or merely awaiting the innovation of the modern corporation. Rather the legal structure was endogenous; the law reflected appropriate institutions for the existing economic conditions. The size of the local market determined the efficient size of business firms. For example in the case of shipping companies, the extreme dangers of sea travel and long sailing times gave owner run firms advantages over diversified ownership. Free rider and moral hazard costs outweighed the gains from inexpensive access to capital.

⁴³Frey, pp. xiv.

⁴⁴Seavoy, p. 5.

There were, of course, transactions that required more capital than could readily be provided by one owner or group of partners. Markets for loans did exist. Either the government, or religious or charitable corporate institutions completed the rest of these transactions.

The canal, followed quickly by the railroad, changed all this. Canals required integrated systems to run efficiently -- both in the development and operation. Local ownership was not efficient, since local decisions as to location and operation had large spillover effects to other locations. The difficulty with independent development is the same with all systems; the parts have to fit together. This problem plagued the early canals; when independent canals linked into integrated systems, the carrying capacity of the narrowest canal limited traffic on the entire system.

The difficulty with divided control over operations may be illustrated with an example from the railroads. According to Miller,⁴⁵ early railroad corporations were founded by individuals with an interest in local commercial centers. Since through traffic (trains using the tracks without interacting with local commercial enterprises) could potentially result in a loss in local comparative advantage, early tracks were intentionally designed to be of differing gauge (width). But of course this prevented the efficient movement of freight across long distances; freight would have to be off loaded then reloaded at every difference in gauge. Under divided ownership and management, local municipalities, which faced the loss of

⁴⁵Miller, p. 55

rents under “through trains”, were able to prevent the efficient integration of the system. Only when local control over railroads was eliminated and ownership became centralized, did the lines become standardized in gage.⁴⁶

State Involvement in Infrastructure Development

Since state laws did not permit the joint stock corporation, political entrepreneurs built the earliest canals with financial assistance from state governments. The initial success of the New York legislature’s backing of the Erie Canal stimulated similar attempts by other state legislatures.

There are, however, severe institutional inefficiencies in having state legislatures finance or otherwise control the construction of canals. Slight differences in the route of a canal can change its cost substantially. Since a canal must be level, crossing hills or drainage is extremely expensive. Changes in elevation must be compensated for by raising the canal above or below the level of the ground, or by locks, which raise and lower canal traffic to a different level. When a profit maximizing firm builds a canal, it will select a route that

⁴⁶ The Coase theorem claims that if bargaining is inexpensive and property rights are well defined, bargaining will lead to the efficient outcome. This implies that the local interests would be paid off to allow transshipment. Why didn’t this occur? Presumably both conditions were not met; bargaining was expensive due to the divided local ownership, and property rights to assets that were about to lose their value were intrinsically ill defined.

maximizes the difference between the total revenues from traffic and the total expense of building to that traffic.

The legislators who vote for the passage of a canal bill do not own the profits of the canal. Instead they have some economic rights over their elected office, and land in particular locations. Therefore in making tradeoffs between routes that increase profits and those that increase an incumbent's chance of reelection, we expect politicians to favor the latter. One way to benefit your constituency is by having a canal built through your district. For this reason state legislatures would choose routes that were less efficient than those selected by profit seeking firms.

A number of states, following New York's example of the Erie Canal, attempted to develop transportation infrastructure. They were all, to varying degrees, failures. For example the State of Illinois⁴⁷ committed itself to a plan of building 1300 miles of railroad in Southern Illinois, as well as additional turnpikes and river improvements. The State, which had a population of 300,000 at the end of the eighteen thirties, committed itself to twenty million dollars of liability. These state bonds were to be paid out of general tax revenue, and were not dependent on income from the transportation projects themselves.

⁴⁷The Illinois example is drawn from Miller pp. 50-51.

Shortly after the plan was enacted, the depression of the late thirties struck, and with only fifty five miles of completed track, the state stopped construction. It later sold these few miles after concluding that they did not earn sufficient revenue to cover their maintenance expenses. The State of Illinois was left with a debt of fourteen million dollars, and was forced to default on its interest payments.

Illinois' new constitution of 1848 forbid direct involvement in infrastructure, but required the passage of general incorporation statutes, which allowed corporations to form without individual approval of the state legislature. Because of similar experiences in other elsewhere, State Governments were not a source for the large capital requirements of integrated transportation systems.⁴⁸

The Federal Government was the next obvious institution capable of funding the construction of canals. But in fact it played almost no role in their early development. Explanations for this are many. One factor was that as the result of expensive transportation and communications, the power or scope of the Federal Government was very limited.. Society itself was more local and regional, and so problems for which the government was the low cost solution also tended to be local or regional.

⁴⁸This is not to imply that had always been the case. The truth seems to be the oppose; public works such as roads and bridges were up to this point usually constructed by or for State governments. See Miller, pp. 43.

The political geography of transportation also made Federal Government involvement rare. When the benefit of a particular improvement was almost entirely local, or perhaps even detrimental to other regions, national political support is difficult to sustain. Take for example, the transportation network leading to the Midwest.

When finally settled (middle of the nineteenth century), the plains of Illinois, Indiana, Wisconsin and Minnesota proved to be very productive in grain crops. But grain is a low value per weight commodity, which required relatively inexpensive transportation if it was to compete with local produce. There were two feasible routes for Midwest output; either South along the Mississippi to the port at New Orleans or East through the Great Lakes to the Erie Canal and New York City. Each route had difficulties.⁴⁹ Mississippi traffic was by barge and steam boat, and suffered the inconvenience of a short transportation season due to ice on the rivers. The Mississippi also was not entirely navigable, and sections of rapids required offloading and reloading.

The Eastern route had a longer transportation season, but was initially unconnected to much of the Midwest. For while the Midwest contains large expanses of territory with access to navigable rivers, none of them connect to the Great Lakes. Eventually canals and railroads

⁴⁹Miller, Chapter 1.

were built linking the Mississippi drainage system with the Great Lakes, but until that time the Southern route was less expensive.

Any infrastructure improvement would result in large changes in relative shipping cost, and changes in fortune of either New York (and the other Northern Cities) or New Orleans. Canals or railroads linking the river system with the Great Lakes tended to divert more of the traffic away from the Southern Port, while improvements in the river system had the opposite effect. Southern/Northern political and economic rivalry then ensured that the Federal Government, which required the cooperation of both regions, would not engage in transportation improvements.

The Corporation

The lack of a suitable institutional structure prevented the initial gains from building canals. Yet institutional structures are endogenous. As argued above, general incorporation laws for business did not exist because there was not a demand for them. Only as the demand for new legislation appeared did the institutional structure change. With the successful development of the canal in England,⁵⁰ its successful adoption in the United States only awaited political

⁵⁰ The canal era began in England in the 1760's. By 1830 4,000 miles of waterways (canal and river) were in use in England, which combined with coastal sailing and horse drawn carriages resulted in the most advanced transportation system in the world (for the era). The Railroad was developed in England after the turn of the century. Each system was studied and copied by American Engineers. See Mountfield chapter 1 and Frey, page xiv.

entrepreneur who would capture some of the benefits of canals by changing the rules of the game.

This is what happened. For example in New York, joint stock limited liability⁵¹ companies were created by state charter. Much of the stock of these corporations ended up in the possession of state legislators and their associates. Thus the law adapted to take advantage of the new potential gains from trade.

Initially there was a limited number of feasible opportunities for canals (later Railroads), leading to the success of individual charters.⁵² The resulting lowering of transportation costs lead to increasing economies of scale in many forms of commerce, for example manufacturing. Eventually then, state laws were further modified to allow general incorporation. Any firm could incorporate, as long as it satisfied certain conditions. The political details are unimportant; the driving force of change was clearly the decrease in transportation costs and the resulting efficiency of firms with corporate organization.

⁵¹Limited liability was not a uniform feature of early corporations.

⁵²The canals, including the Erie canal, and all early railroads were incorporated under special state charters. Almost every railroad in Illinois was the result of special charters, even after the passage of general railroad incorporation statutes.

The corporate form was first used in business for the building of canals. This was followed shortly after by the introduction of the railroad⁵³ and telegraph.⁵⁴ Miles of tract increased from 23 in 1830, to 2,818 in 1840, to 9,021 in 1850 to 30,626 by 1860.⁵⁵ The resulting dramatic decrease⁵⁶ in transportation and communication costs enlarged markets, and resulted in economic advantage to larger, more specialized plants and firms. This in turn created a demand for general incorporation laws for all types of business.

The Advantages of the Corporation

As transportation and communication decrease in cost, the advantages of the Corporation over sole proprietorships increase. The corporation allows the pooling of capital for transactions most efficiently managed by a single party, but beyond the wealth constraints of any single individual or family. While many individuals could own one mile sections of a

⁵³There is some debate as to how important the railroad really was. While some authors have attributed most of the economic development of the United States to the railroad, others have argued that it was the canal system that resulted in the more dramatic decrease in transportation costs, and that the railroad was only a slight improvement over the canal. See O'Brien.

⁵⁴ The invention of the telegraph is credited to Samuel F.B. Morse (1791-1872), who expanded his own and others' ideas in the 1830's. The first telegraph line ran along the route of the B&O railroad, and sent its first message in 1844. It was the railroads (who had long right of ways) and eventually the Western Union telegraph monopoly that implemented the telegraph system, which in turn was replaced by the telephone system in the 1940's and 50's. See Frey, pp. 280-282.

⁵⁵ Frey, p. xviii.

canal, the ability of one party to make decisions incompatible with the interests of everyone else leads to the potential for holdup. And no one will make the investment in a dedicated asset when the quasirents can be captured by someone else.

Therefore if a canal is going to be build, it seems likely that one party will control the entire canal. Yet no one owned enough wealth to build a canal from New York City to Erie, so some pooling arrangement was necessary. The corporation, with central authority but diversified ownership, allowed this pooling.

Corporations also had scale advantages over partnerships. Most particularly, as the number of partners grows, the cost of agreement increases. With a corporation, those suppliers of capital who disagree with the management's decisions have another option besides the breakup of the organization; they can sell their shares. Considerations including the replacement of deceased or retired partners, liability for errors, and ease of court representation all favor the corporation over the equivalently sized partnership as the capital requirements increase.

Summary

⁵⁶ Ibid. Frey claims that in 1853 the cost of moving one ton of freight one mile by turnpike was \$15, by railroad \$1 to \$2, and by river \$.37.

Features of early American economic life, particularly expensive transportation and communication, made sole proprietorships and partnerships the efficient form of business organization. Changes in technology⁵⁷ and political boundaries⁵⁸ allowed projects requiring the pooling of much larger amounts of capital. These transportation systems were at first developed by state governments and corporations incorporated under special charters, but after various governmental failures, the predominant organizational form became the corporation (often with governmental subsidy).

This introduced the corporation as a business organization - but only in the realm of transportation. However the corresponding reduction in transportation and communication costs changed the characteristics of other markets, making the corporation advantageous in other fields as well. Therefore we see that the revolution in transportation brought the corporation to the US by two means; it was the efficient organizational form for the transportation systems, and the existence of these systems made the corporation efficient in many other fields.

⁵⁷ For example the invention of canal and railroad technology.

⁵⁸ Specifically the ending of warfare in upper New York State, as well as the expansion of the US via the Louisiana purchase.

PART II: THE DEVELOPMENT OF TRUSTS AND ANTITRUST LAWS

Price Discrimination and Quasirents

The influence of railroads on corporate form and behavior did not stop with the enactment of general incorporation laws. The particular cost circumstances that made the corporation such an efficient institution for the creation of the railroads also led to widespread price discrimination, which in turn elicited attempts to control their behavior through antitrust laws. Thus the cost conditions of railroads led both to the rise of the corporation and to its regulation. This attempt at government control first took the form of state level laws designed to regulate prices, and only later (as we shall see) resulted in Federal action.

The early railroads engaged in widespread price discrimination. For example, the Wabash, St. Louis & Pacific Railroad, which ran a line between Peoria, Illinois, and New York City, charged less to ship between these two points than from Gilman, Illinois and New York. Peoria was a shipping center, and had several competing railroads (a “competitive point”). Gilman was one of many small stops between Peoria and New York for the Wabash (but not its competitors), so the railroad was charging more to travel a shorter distance than it did for the longer distance.

The importance of the Railroad’s cost structure is that distance is not the most important cost. Some marginal costs, such as fuel for a heavier train, certainly did vary with distance.

But many, including loading and unloading, marketing, and administrative costs, all did not vary significantly with distance. It is difficult to say whether the marginal cost of a shipment from Gilman or Peoria was larger.

As well as the difficulty in applying overhead costs to any particular transaction,⁵⁹ there is the further difficulty that marginal costs only make up a small share of the total cost of operating any particular line. For a disproportionate share of the costs of a line are either sunk costs, such as the installation of track and other equipment, or costs that do not vary with the volume of shipping (e.g. maintenance costs).

With rising marginal costs, setting the price for *all* units equal to the marginal cost of the *last* unit leaves rents from low marginal cost units to pay for fixed costs. But in the case of the railroads with fairly constant and relatively low marginal costs, a price set to this level did not cover overhead costs. Therefore for average cost to be exceeded, the price for some shipping transactions would have to exceed marginal costs, or the railroad would shut down.

⁵⁹It is interesting to note that modern accounting, which makes some attempt to measure and control costs, was invented by the railroads. See Chandler, p. 54.

Price Discrimination

The other attribute of railroads that made price discrimination so likely was the ease of both sorting customers and preventing arbitrage. In the case of Peoria, alternative railroad lines made demand very elastic; competition drove the price down toward marginal costs.

But these lines took different routes between shipping centers, and for smaller localities such as Gilman, demand was insufficient to warrant more than one railroad. This implied that the demand faced by the one line that did enter Gilman was relatively inelastic. Railroads were well acquainted with alternative transportation, and therefore sorting customers by elasticity was not difficult. Preventing arbitrage was equally simple, since customers from Peoria could not resell their low priced shipping to customers from Gilman.

Because marginal costs were such a small portion of total cost, and because sorting of customers and preventing arbitrage was relatively easy, price discrimination was almost universally practiced. Shipping between points with substantial alternative transportation sources was relatively inexpensive, while shipping from locations with few alternatives was often very expensive. It was demand, not cost, that determined rates. This obviously caused considerable consternation on the part of those wishing to ship from single rail towns.

Price discrimination is universally listed as the primary evil perpetrated by the railroads. Differences in shipping costs were instrumental in the rise or fall of particular commercial

centers. Communities with low rates prospered, while those without alternative transportation suffered. The railroads determined these rates, not on the basis of differing costs, but instead on the basis of their customers' alternatives.

Railroads not only discriminated between locations, but also between individual customers. The most famous example of this is the Standard Oil Company, which officially paid the same rates as its rivals, but then received secret rebates.⁶⁰ This sort of price discrimination occurred because some customers, often heavy users of transportation services, had more elastic demands. The cost of finding a low price does not vary with the total volume of shipping. As compared to small customers, customers with large shipments are more likely to incur these search and bargaining costs, since the total savings are greater. Railroads are also more likely to provide a price discount when threatened with the loss of a large shipment. These discounts gave a cost advantage to large volume customers, and this advantage was unrelated to other resource costs (except those that made the large customers succeed in the first place).

A number of economists have criticized price discrimination for this reason, since it changes for reasons unrelated to "resource cost", business size and therefore the types of services provided. Of course since we do not know the cost of the next best arrangement, we can not say if price discrimination is inefficient or not.

⁶⁰See the fourth chapter.

Rent Seeking and the Protection of Property Rights

Propaganda on the part of those wishing for government regulation of railroad pricing asserted that railroads were abstracting all of the hard earned rents (quasirents) from farmers who had settled the frontier. In outrage these farmers rose up and passed the Illinois Granger laws of the 1870's, restricting the railroads' power to whimsically alter the fate of entire communities.

More recent research⁶¹ casts doubt on the accuracy and usefulness of this account. Before the coming of the railroad, only Midwest farms with access to river transport engaged in production for distant markets. While land away from navigable rivers might be well suited to grain production, overland shipping costs prohibited profitable production. Therefore farmland away from accessible transport was either not settled, or crops were produced that were not shipped to distant markets.

As railroads became financially and technically feasible, new lines were created that opened production to the cash economy. These lines obviously could only increase the value of the surrounding farms. But any farmers considering switching resources to market production knew that they would be supplied transportation by a monopolist. These farmers would

⁶¹See Chalmers, p. 4

presumably only incur sunk costs if they anticipated either the monopoly price or a satisfactory control (perhaps political) over the rate the railroad would charge.

Therefore the railroads did not abstract or “steal” the quasirents from the farmers; they instead created new rents by providing lower cost transportation. Of course once the railroad incurred the substantial expense of laying track, it had quasirents that were potentially extractable. Investors expected to receive a certain rate of return (their opportunity cost) on the large capital expenditures involved in setting up particular lines. But if returns were only enough to cover operating expenditures, the sunk nature of the investment implies that railroad services would still be provided. Thus the railroads were liable for “holdup” of their quasirents.

Given this holdup potential, we expect that the railroads would only be willing to incur sunk cost if they had what they deemed to be sufficient guarantees. In general, guarantees or protection from holdup may take many forms, including either common ownership or explicit contracts. Each form was used in railroad development.

Many of the first railroads were founded by members of the communities they served;⁶² since the members owned the local railroads through ownership in corporate stock, their incentive to abstract the railroad’s quasirents were reduced. But economic conditions changed, this

⁶²Miller, p. 55

became a less secure mechanism of protection, for increasing specialization in both agriculture and commerce made larger scale transportation networks economically feasible.

Initially railroad lines were primarily local, with differing gages requiring unloading and loading of cargo. The civil war forced railroad managers to coordinate transshipment of goods, and in the eighteen sixties, seventies, and eighties local railroads were increasingly consolidated into larger railroad lines. This consolidation, and expansion of new lines, created a tremendous demand for capital, far beyond the wealth constraints of the local users of these lines. The result was the much more extensive development of financial markets, allowing Eastern and European investors access to a liquid investment. In fact this expansion of railroad firms created most of the trappings of modern corporate capitalism, including stock markets, modern accounting, and large vertically integrated firms run by hired managers.⁶³ But it was through this absentee ownership that the interests of the railroads and its local users began to differ.

Railroads also attempted to prevent holdup through contractual arrangements. In particular, once the railroad were in place, state regulation could restrict railroad rates, leaving just enough return to pay operating expenses. Owners of these quasirents were usually not

⁶³Chandler, p.54

residence of the state⁶⁴ in which the railroad was built. Its customers, who would benefit from rate regulation, were both residents, and more numerous, giving them a substantial advantage in state legislatures.

A second form of protection against this form of holdup was contractual. Railroad entrepreneurs in Illinois formed their corporations under special state charters, and reserved rate setting powers for the corporations. These special charters were contracts between the railroads and state governments protecting the railroads from governmental abstraction of quasirents. Illinois had general incorporation laws for railroads by 1849, but by 1870 not a single railroad had incorporated under them. Corporations preferred special incorporations, for the general incorporation laws reserved some rate making ability for the state legislature.

The railroads had, for a substantial part of their development, one final form of protection. As soon as a railroad line was complete, its customers would prefer rate regulation. But if the state legislature abstracted quasirents from one line, railroad investors were unlikely to construct lines elsewhere in the state. Since initially most of the state did not have railroad lines, the “have not’s” outnumbered the “have’s”, and the railroads had some protection via the support of communities desiring a railroad. Of course as construction continued, this support undermined itself.

⁶⁴Railroad capital was, to a large extent, provided by English investors. See Miller p. 51 for details of English concern with expropriation of quasirents.

The common story then, that the railroads “exploited” rural farmers, seems doubtful if we understand this as the abstraction of quasirents from settling and farming land. Railroads did not limit transportation options, but instead made production for the cash economy feasible.

Rather it was the railroads that had extractable quasirents, and investors were only willing to incur sunk costs with some expectations of protection. Several forms of protection were available, including legal guarantees and threat of canceling future construction.

State Regulation of Corporations - The Granger Laws

While firms will only incur sunk cost if they expect some protection of their investment, in a world of imperfect information, complete protection is prohibitively costly. The Railroads’ protection through the eighteen forties, fifties and sixties was fairly successful throughout the nation, primarily because the number of communities with railroads was exceeded by those wishing to acquire them.

This is not to imply that the railroads did not have to incur expenditures to protect their assets. Long before the Granger Laws of the 1870’s, bills were submitted to state legislatures compelling the political regulation of railroad rates. For example in 1854 Rowland Hazard, a

legislator in Rhode Island, submitted a bill requiring railroads to charge rates at every station proportional to their lowest rates at any station (prorata).⁶⁵

This bill passed the Rhode Island Senate but was defeated in the House. The bill was supported by commercial interests in Rhode Island, for competition for the Boston to New York through route was driving shipping prices below rates between Rhode Island and these two cities. The price discrimination resulted in losses to commercial firms in Rhode Island. It is important to note that it was the loss of rents to commercial parties, not agriculture, that drove this and other early bills.

Similar bills were supported by the Clinton League in New York. This League was composed of businesses with resources specialized for use with the Erie Canal, and its goal was to increase railroad rates so as to make the Canal a more viable alternative.⁶⁶

By the 1870's, Illinois passed a series of bills, misnamed the Granger Laws,⁶⁷ which created state control over some railroad fares. The first Granger Railroad Law, passed in 1869, outlawed rebates. But its enforcement depended on individual shippers bringing suit in a

⁶⁵Miller, p. 33

⁶⁶Miller, p.34

⁶⁷Similar laws were also passed in other Midwestern states, and were also called Granger Laws. These bills were misnamed because they were neither created, nor often supported, by the Granger Movement.

court of law, which effectively made it toothless since the firms receiving a rebate would be unlikely to bring suit, and firms excluded from rebates were unlikely to have proof. And of course it was rarely worth the legal expense, since triple damages were not available.⁶⁸ A much more serious law, the law of 1871, required each railroad to charge less for its short hauls than long ones. Furthermore, in order to prevent raising the rates for the more competitive long hauls routes, the rate charged at any point could not exceed the rate charged in 1870. This forced railroads to set all of their rates equal to their previous lowest rates. Finally, a commission was appointed to enforce the act by bringing suite.

The greatest difficulty with the law of 1871 was that it would undoubtedly drive many marginal railroads out of business. There were simply too many differing lines, each with different cost and demand conditions, for the Illinois State Legislature to examine each individually. The legislature had to make some rule, and the one it selected would clearly deprive a large number of communities of rail service, since railroads would be forced to charge the rate at their most competitive location. This rate, probably close to marginal cost, would not pay for the substantial fixed costs required to maintain a line. Once this became known, it effectively divided the legislature into “weak” and “strong” road factions, with the weak road faction fearing the loss of service.

⁶⁸Miller, p. 75

The Granger law of 1873, while still odious to the railroads, revised these failings with the following provisions:

1. Any railroad corporation charging more than a fair and reasonable rate of toll would be deemed guilty of extortion.
2. If any railroad charged higher rates for a long haul than for a short haul, it would be “deemed and taken against such railroad corporation” as prima facie evidence of unjust discrimination.
3. It was not deemed a sufficient excuse or justification for any of the above discriminations that the station at which the lower charge was imposed was a point at which there existed competition with another railroad or some other means of transportation.
4. The Railroad and Warehouse Commission (the Strong Commission) was to investigate the enforcement of this provision and prosecute for violations.
5. The Commission was directed to prepare for each railroad operating within the state a schedule of prima facie reasonable maximum rates. The schedules were to be revised as often as required and published whenever changes were made.⁶⁹

This law had the effect of making price discrimination illegal, but allowed railroads to raise rates at competitive points so as to ensure sufficient returns to cover fixed cost. It also allowed railroads to determine their own rates, subject to court (not legislative) review.

⁶⁹Miller, p. 93

The Order of Patrons of Husbandry, or Grange, was founded in 1867 by Oliver H. Kelley, an official of the United States Department of Agriculture. It assumed a position of leadership among rural fraternities in the early 1870's, and gave its name to the broader movement. It sought to restore and enhance the quality of rural life in America through programs of education and self-help.⁷⁰

The leadership of the Granger movement did not take an active interest in railroad legislation until 1873, for the simple reason that it was not yet sufficiently organized to do so. Thus the most radical act (1871), is quite mistakenly given the name Granger. The 1873 act was in fact opposed by the Granger leadership. The reason for this misnaming is presented by Miller:

“Merchants at way points and terminal markets, in fact, were more apt to be adversely affected by this abuse (price discrimination) than were farmers who often had a choice of collecting points and a choice of buyers”. “In most cases, initiative and leadership were provided by shippers and businessmen working through their boards of trade, mercantile associations, commercial conventions, and the regular political machinery of the state.”

⁷⁰This paragraph is directly from Miller.

“To disregard their active participation in the movement would be quite wrong. Granger literature was filled with commentary on the railroad problem and with claims that the Grange led the drive for controls. But eastern journalists, led by E. L. Godkin of the *Nation*, were too quick to identify the movement for rural organization and the movement for rate regulation as one and the same and to characterize the legislation of 1874 as Granger legislation. It seems clear that they did so with some malice aforethought in order to affix the stigma of agrarian radicalism on all subsequent efforts to regulate rates”.⁷¹

The early attempts at state government regulation of corporations (the Granger Laws) arose out of the commercial conflict over transportation quasirents. Railroads were then both the source of the corporation, and the source of its initial regulation. The main parties to this conflict were Midwestern commercial interests⁷² facing discriminating rates and Eastern and European capitalist owners of railroads. The revolt of the farmers, or Granger movement, while supportive of attempts to limit the power of these new corporations, did not play the role initially claimed by both itself and its enemies.

⁷¹Miller 164-5.

⁷²See Chandler p 72.

Corporate Mergers and Trusts.

The steady expansion of the United States market, both because many regional markets became unified by the expanding railroad, and also because of the increase in population, resulted in both new specialized industries and many new industrial firms. Unlike the local monopolies that existed before inexpensive shipping, many of these new firms were forced into price competition.

Attempts were made to minimize price competition through industry trade associations. But in the face of rapid change, both in supply and demand conditions, cartel arrangements were exceedingly difficult.⁷³ In a few industries,⁷⁴ trusts were formed that held the stock of the major firms, and these trusts were better able to enforce common interest. The most famous example, the Standard Oil Trust formed in 1882, is the topic of the fourth chapter. Holding corporations, which entirely replaced trusts in industries merging to prevent price competition, did not exist, for it was not for another seven years until the state of New Jersey passed incorporation laws permitting corporations to hold shares in other corporations.⁷⁵

⁷³See the fourth chapter for a more explicit treatment of cartel arrangements.

⁷⁴All in the refining and distilling industries. Chandler, 74.

⁷⁵Chandler, p. 73

The trusts, and much more common trade groups, were opposed by less successful organizations. For any collusion to work (raise prices), the involved firms must somehow prevent new entrants from taking advantage of the higher prices. There will naturally be conflict then between those firms seeking high prices and those taking advantage of them.

For example, as the fourth chapter will explain in greater detail, Standard Oil elicited the formation of several "independent" oil producer organizations, whose main purpose seemed organizing for conflict with Standard. Once again the railroads were an essential part of the conflict, for price discrimination in favor of Standard gave it a decisive edge. These opposing trade groups battled in the market through price competition, and in the political arena through state politics and court battles.

All of the early political battles took place in state courts and legislatures. Because businesses were almost entirely local before the rise of inexpensive transportation, there was almost no demand for national antitrust laws. But after the expansion of the railroads and resulting national scope of industry, and the increased power of the national government brought about by the civil war, the focus eventually shifted to the Federal Government. Independent oil producers and analogous associations in other industries lobbied Congress to pass Federal antitrust legislation. As part of this populist movement against large corporations, the Sherman Antitrust Act was passed by congress in 1890.

The Sherman Antitrust Act

The *Antitrust* is almost certainly the wrong word to associate with the purpose of the act, which regulates both contractual restraints of trade and monopolies. Trusts, which placed assets of possibly different owners under the control of one party,⁷⁶ were not used for the purpose of restraining competition after 1890. But it was not the Sherman Act which eliminated this use, but rather the prior New Jersey incorporation laws that allowed more efficient holding companies. The name Antitrust was selected because of its political connotations. Some might argue that this confusion at the beginning of the law characterizes the rest of its history.

The important sections of the act are:

Section 1.

"Every contract, combination in the form of trust or otherwise, or conspiracy, in restraint of trade or commerce among the several States or with foreign nations, is hereby declared to be illegal..."

⁷⁶Trusts developed their name from their purpose; a trust agent manages assets for someone else's benefit. The beneficiary has to "trust" the management of the agent. This form of agency is particularly common when the beneficiary is unable to responsibly manage their own assets, as in the case of children. Imagine that there exists a substantial body of law, professional cadre of defense attorneys, a division of the Department of Justice, and college courses all devoted to controlling trusts.

Section 2.

“Every person who shall monopolize, or attempt to monopolize, or combine or conspire with any other person or persons to monopolize any part of the trade or commerce among the several States, or with foreign nations, shall be deemed guilty of a misdemeanor (later changed to felony) ...”

As the law now stands, section one deals with “restraint of trade” issues, while section two violations involve “monopolization”.

The Relation of the Act to the Common Law

John Sherman, a Senator from Ohio who sponsored the act, did not believe that this act substantially changed the law. He argued that this bill simply codified the common law against monopoly,⁷⁷ and in early cases the courts interpreted the law in this manner. But as conflicting Supreme Court rulings indicated (each with precedent), the English and American Common law was far from unified in its treatment of monopoly.

⁷⁷The Act was not viewed as a substantial change in the law, as evidenced by its early infrequent use: the Harrison administration brought seven antitrust cases before the 1892 election, but won only two. Letwin, p. 116.

Letwin identifies four branches of common law related to monopoly and restraint of trade, and he notes that two were terminated by statute, and two were constrained by English courts after the American Revolution. United States common law, which differed from its parent after the Revolution, did provide some support for Sherman's interpretation. But as Letwin writes, "Nevertheless, the common law was very unwieldy material from which to have constructed a law to control modern corporations and big businesses; the most that could be said for it was that nothing better was available."⁷⁸

The common law had restricted (or refused to enforce) unreasonable contracts "in restraint of trade." By reasonable, the common law meant the restriction was not "more injurious to the public than is required to afford a fair protection to the party in whose favor it is secured."⁷⁹ A first question for the courts to consider under the Sherman Act was whether "in restraint of trade" followed its common law definition, and referred only to unreasonable restraints, or whether it applied to all restraints on interstate commerce.

A second question for the courts to decide referred to section two; monopolizing or attempting to monopolize any part of trade. Was sheer market share sufficient for violation of the act, or were some explicit and illegal acts required as proof of violation? The English common law had varied its opinion concerning monopolies, but tended to view them as

⁷⁸Letwin, p. 15. See Letwin, chapter two for a complete analysis of previous common law.

⁷⁹52 Fed. 118. Quoted in Letwin, p 147.

special or exclusive privileges provided by the government. During some sections of its history, English law had attempted to actively discourage the provision of governmental monopolies, while at others the courts had defended the government's legal right to restrict trade.

Early Decisions

The earliest ruling on these questions (subsequent to the passage of the Sherman Act), was one of the whiskey trust cases, *In re Greene*.⁸⁰ In this appellate case Judge Jackson ruled that in order to violate the common law and legislative limits on monopoly, an monopolist would both (1) have to have exclusive right or privilege and (2) impose some restriction on others who might infringe on that privilege.⁸¹ Existence was insufficient as proof of a violation.

Two subsequent cases illustrated the outer limits of this ruling. In one, *Trans-Missouri Freight Association*,⁸² the lower courts ruled that an agreement to fix railroad rates by competing lines did not violate the statute since the defendants did not have an exclusive right to carry freight.⁸³ On the other hand, in *United States v. Patterson*,⁸⁴ Judge Putnam

⁸⁰52 Fed. 104

⁸¹Letwin, p. 147.

⁸²53 Fed. 440. Note that the railroads are involved in almost all of early antitrust history.

⁸³Letwin, p. 152

⁸⁴55 Fed. 605.

ruled a cash register trust violated the act because it used various coercive practices (including violence and bribery) to maintain its monopoly.

While Jackson's ruling held sway in the courts for several years, by 1897 the Supreme Court reversed the lower court's decision on *Trans-Missouri Freight Association*. Justice Peckham ruled that the Sherman Act explicitly prohibited all restraints, reasonable or unreasonable. In the same ruling, Justice White wrote a dissent that introduced the idea that eventually became entitled the *Rule of Reason*. The rule of reason evaluates contractual restrictions on the basis of their reasonableness, generally as defined by precedent. These two conflicting ideas, that acts should either be strictly illegal or illegal independent of their consequences (later called *per se illegality*), as opposed to the *Rule of Reason*, continue to represent polar extremes of how the court judges conduct.

This question was not purely academic. For example, if a baker only took on apprentices if they agreed not to set up competing bakeries nearby,⁸⁵ the baker would be in violation of a *per se* rule against any restraint of trade. But the baker may not be willing to create his own competition, so the creation of any alternatives may require contracts in restraint of trade, and so the restraint would be judged acceptable by the *Rule of Reason*.

⁸⁵ Assume just across a nearby state boarder, so we have the element of interstate commerce.

A literal application of a rule against any restraint of trade would rule out all partnerships (which limit competition between partners), and any other element that restricts commerce. Of course this was unacceptable to the Court, so Justice Peckham modified this understanding so that it did not apply to ancillary purposes; contracts whose primary purpose was a restraint of trade were judged as *per se* illegal, but any ancillary elements of these contracts would be judged under the *Rule of Reason*.⁸⁶

Summary

The United States economy changed dramatically during the last century. Initially high transportation costs resulted in small markets and firms. The combination of technological advancement (supply) and expanding population and area (demand) lead to a transportation revolution, first with the canal then the railroad. This inexpensive transportation increased firm size, and eventually created the current limited liability corporation.

The railroads were a central component of this change. Their large scale required the creation of the corporate business form. Combined with the canal, they dramatically reduced transportation costs. Railroad corporations invented modern accounting and the salaried managerial class. And the wide spread use of price discrimination, brought about by the

⁸⁶Letwin, p. 170.

railroads' cost and demand conditions, created the demand for government regulation of large business.

Since most firms operated within states, early regulatory efforts occurred in state legislatures and courts. The Granger Laws of the 1870s, which attempted to control railroad prices and policies, are typical of state regulation of business. As firms became larger, and as the Federal government became relatively more powerful, the demand for Federal regulation of corporations arose. In 1890 the Sherman Act was passed, allowing the Federal government to prosecute firms that conspired to restrain trade or monopolize commerce.

The Federal courts were forced to rule on what business practices were prohibited under the Sherman Act. Judge Jackson, in an appellate ruling, developed the doctrine that to violate the Sherman Act, the defendant would both (1) have to have exclusive rights or privileges, and (2) impose some restriction on others who might infringe on that privilege. This doctrine was later undermined by the Supreme Court in the decision *Trans-Missouri Freight Association*. The evolution of the law to its current form will be a topic in the remaining chapters.

CHAPTER 3: THE CANALS

The central focus of property rights theory has been on how the characteristics of a good result in the institutional arrangements under which it is produced and traded. Questions of firm size and scope,⁸⁷ which attributes will be priced and measured, and residual claimancy all form the core of property rights theory.

Since these essays are related to property rights and antitrust, this chapter concerns alternative contractual forms in the development of monopolies. Much of property rights analysis has focused on efficiency explanations for monopoly - why one (usually large) firm is the only seller of a good or service. I wish to abstract from questions as to whether the good could be supplied by either a competitive market or monopoly, and instead investigate alternative arrangements in the provision of monopoly assets.

The example I have selected is the development of the early United States canals. Because of the large expense of their construction, the low operating costs, and the extremely restricted geography through which they could operate, canals were usually monopolies. The price they charged was never driven to *marginal* cost through price competition with rival transportation systems, since they generally had marginal costs far below their average

costs, due to large fixed expenditures.⁸⁸ So, while prices were constrained by competition with other forms of transportation, they were not restricted to anything near marginal cost.

Given that canals would be monopolies, under what form of contract would they be built? Would they be developed by private individuals, corporations, or governmental bodies? I will argue that there are ownership considerations, varying with circumstances, that explain the differing ownership patterns we see in the historical record.

Background Information

The first burst of canal construction occurred in England in the half century before 1825. Most were built by joint stock companies, and most paid substantial dividends to their shareholders. The profitability of these canals resulted from their locations - they linked urban or industrial areas, and had relatively few natural obstacles to overcome.⁸⁹ These two elements - the guarantee of revenue, and the minimization of costs - will be important in the rest of the analysis.

⁸⁷ Both what assets will be owned, and what assets will be produced.

⁸⁸ Canals had very large initial construction costs, and moderate maintenance costs that did not change substantially with the amount of traffic, and low marginal costs. This form of costs leads to a (roughly) horizontal marginal cost curve and average variable cost curve. The average total cost curve, which includes the sunk cost of construction, is declining up until the point of canal capacity.

The first wave of US canal construction began as the English construction tapered off. Why was the introduction of canals delayed in the US relative to Britain? First of all, the US had a considerably lower population density than England. This meant that number of customers per canal was lower in the US than in Britain. A canal needs to join large markets in order to generate traffic and revenue, and the US did not have these markets when England first started building canals. Large demographic changes in the first decades of the nineteenth century changed this.

Just as importantly, the geography of the two countries was significantly different. The US population in 1800 resided on the Eastern coastline, and used Atlantic shipping for transportation between urban centers. A number of Eastern US canals were (eventually) built to improve transportation between these urban centers, but the potential gain was much less than in inland England.

The highest transportation costs for the US - and the greatest potential for improvement - was for goods shipped between the Midwest and East Coast. While sparsely settled by whites, the Midwest was the source of the fur trade as well as an agricultural area of great potential. The difficulty with exploiting this opportunity was the Appalachian Mountain chain, which proceeds Northeast to Southwest with only one gap. The relevance of the mountains is that

⁸⁹ Goodridge 1961, p. 2-3.

canals change elevation by means of locks. These are expensive to build and maintain, and they slow transit times. Even with locks, canals can not cross peaked elevations unless there is a source of water at the top, since water will only run downhill without pumping.

Four Canals

The Erie Canal

The one break in the Appalachian chain occurs in a series of connected valleys in New York.⁹⁰ This route was previously unavailable to the English settlers, for Northern New York had been in an almost constant state of warfare between the English, Dutch, French and Mohawk. With the end of the Revolutionary war this changed, and the increase in Western settlement created an even larger demand for transportation services. Trade routes, at the end of the war, were three. First, the French brought trade goods down the St. Lawrence sea way, around various falls, and into the great lakes. Next, the Mississippi River was a natural route to the Gulf. And the major all-land route was via the Pennsylvania Turnpike (and others), a toll road for teamsters.

The route that became available through New York was primarily rivers, with a few short portages between them. In the early 1790's the State legislature incorporated the Western

⁹⁰ This section draws its facts from Rubin 1961.

Company, which combined canals and river dredging in an attempt to create an all water route from the Hudson to the great lakes. The company's shareholders were primarily owners of land along the route. As well as private capital, the corporation also received state aid in the form of both loans and subscriptions of stock.

The company was not successful, both because of difficulty in acquiring capital, as well as technical difficulties in using rivers for transport. The company never did complete an all water route to the Great Lakes, and only rarely paid dividends. While it was used for intrastate transportation, its relatively high tolls usually made teamsters a better choice for shipping.

Perhaps in reaction to the relative failure of the Western Company, the New York State legislature itself undertook ownership of an all canal route to Lake Erie. While the project was both extremely large and risky, it also proved to be quite profitable - even before its completion. As sections of the canal were completed, upper New York State experienced rapid settlement.⁹¹ Once the link was made to the Great Lakes (1825) much of the developing trade with the Midwest was diverted to the canal.

The Erie immediately began to change trade patterns, with most low margin goods now being shipped over the canal instead of carried overland. Cities such as Philadelphia, which

⁹¹ New York State was in the midst of rapid development before the beginning of the canal.

depended on the Pennsylvania turnpike, and Baltimore, which depended on shipping from New Orleans, began to loose trade.⁹²

The success of the Erie Canal prompted imitation in a number of other locations in the United States, particularly those that would lose trade to New York. The exchange economy of the time was dominated by small merchant houses, which used specialized knowledge - especially interpersonal connections and reputation - to serve local markets. A change in shipping patterns would result in a loss in the value of these quasi-rents, and it was pressure from these merchants that generated demand for local alternatives to the Erie.

The Pennsylvania Mainline

In the case of Philadelphia, a group of canal proponents, made up primarily of merchants, began a lobbying campaign to convince the Pennsylvania legislature to develop a canal connecting Philadelphia with the West. Initially the lobbyist sponsored research into the merits of railroads versus canals, and their investigative agent unexpectedly reported that a railroad was a far superior alternative (and history proved that he was correct). After bitter infighting, the leadership of the society overruled their researcher, and lobbied exclusively

⁹² Philadelphia and Baltimore were each involved in both sea and land shipping.

for a canal. Eventually they were able to convince the state government to build a canal at tax payer expense.

So began the development of the Pennsylvania mainline. The project had a number of faults, at least in hindsight, that gave it the appearance of a boondoggle. First of all, it was in competition with the Erie canal, the construction costs of which were already sunk, and therefore the mainline was unlikely to generate much revenue. Next, technological developments in railroads were such that the railroad was about to exceed the canal as a cost effective method of transportation - particularly in mountainous terrain. The society promoting a canal probably knew this (or knew it with some probability), but were unwilling to wait the additional time to develop the technology in the United States. Railroad proponents, in opposing the canal, pointed out that in a few years the canal would be made obsolete by railroads, and they were correct.

But the greatest fault lay in geography. As previously mentioned, the Erie canal was a success because it took advantage of the one break in the Appalachian Mountain chain. Pennsylvania did not have such a gap, and various alternatives were proposed for overcoming the mountain barriers. These included extensive locks, a tunnel under the mountains, and stationary engines to pull rail cars over the mountains. The last method was eventually adopted, but construction was started at each end of the canal without any clear idea of how to pass over the mountains.

The final product never did link Philadelphia and Ohio by an all water route - at either end. Barges on the Ohio river were forced to unload their haulage and reload it on railroads (!), which transported it to the Western end of the canal. There the goods were again unloaded, and reloaded on barges. After many locks, the barges eventually reached the mountains, at which point the goods were again unloaded, and loaded in cars pulled over the mountains by stationary engines. After unloading, the goods were loaded on barges, carried East, unloaded then loaded on trains, before finally entering Philadelphia.

Once the Pennsylvania railroad was completed, the same category of haulage could be loaded in cars in Ohio, and shipped directly over the mountains. And since much of the expense of shipping was in the labor intensive loading and unloading, both the railroad and all water Erie canal proved to be much less expensive for most of the goods shipped from the Midwest.

The Mainline had other disadvantages as compared to its rivals. The Erie canal's completion opened up much of the northern and western portion of New York State for agricultural development. This was of advantage not only to the property owners in upper New York, but also the merchants of New York City, since the canal outlet (the Hudson) used the city as a port. This meant that a single canal could generate revenue by both local and interstate traffic. Because of this, it did not prove necessary to include additional projects in order to form a majority coalition.

The Pennsylvania mainline did not lead to extensive agricultural development in Pennsylvania; it traversed the mountains and was intended to provide transportation services from the Midwest to Philadelphia. Since the population of Pennsylvania was also somewhat more dispersed than in New York, a canal of primary benefit to Philadelphia merchants could not pass the state legislature without extensive additional local projects.

Because of these various problems, the Mainline system never developed the volume of traffic its promoters predicted. Virtually all of the local canals were financial disasters, and most were not completed or were given away. The Mainline was eventually sold to the Pennsylvania railroad, which dismantled it after a few years of operations. In the end the State incurred something like \$58 million dollars in expenditures on various canals.⁹³

In contrast to this example of government owned and operated monopoly, there were two canals developed in New Jersey by private (for profit) corporations. These two canals had a number of features that distinguish them from the Pennsylvania example, and the difference between the two private canals will also illustrate circumstances under which private corporations are more likely than governmental bodies to develop assets.

⁹³ Goodrich 1960, p. 69

The Delaware and Raritan

The Delaware and Raritan canal connected the Delaware and Raritan rivers. Its completion created an inland waterway connection between New York City and Philadelphia. The canal was relatively short, and, more importantly, crossed a low and flat space of land.

The canal had been proposed before the turn of the century, and two previous corporations were formed to develop it. The first failed due to technical reasons,⁹⁴ and the second was unable to negotiate a contract successfully with Pennsylvania to use water from the Delaware river.⁹⁵

Several bills were proposed in the mid 1820's for the state to build or finance the Delaware and Raritan, but opposition from a competing railroad line,⁹⁶ as well as opposition from counties that would not directly benefit from the canal, prevented passage. Eventually canal and railroad supporters joined together, and parallel incorporation bills were passed. A short time later the newly created corporations merged.

⁹⁴ It tried to make a canal by expanding existing streams, which in the end did not provide sufficient water.

⁹⁵ The Coase theorem should be applicable to this case.

⁹⁶ The Camden and Amboy, which had not yet been built.

The terms of the incorporations⁹⁷ are quite interesting, and I will use these facts later in analyzing the question of residual clemency. Each company was given a monopoly, in the sense that no competing canal could be built within five miles of the Delaware and Raritan, nor any railroad within five miles of the Camden and Amboy. In return for these monopoly grants, the state reserved the right to buy at par value a quarter of the original shares of the companies, and could buy each project 30 years after its completion. Each had maximum tolls constrained by law, and each had to pay “transit duties” to the state. The canal corporation also made a “gift” to the state of one hundred thousand dollars.

After opening in 1834, the canal earned a low rate of return on the initial investment (less than 1%) during the first decade of its operation, but it proved profitable in later decades, with very large returns during the Civil War. In some years the revenue the state government received from the joint canal/railroad company was half of the government’s total income.

The Morris Canal

The Morris canal went across the more elevated Northern section of New Jersey, from Newark west to the Delaware River, and had the effect of connecting the coal rich Leigh

⁹⁷ These corporations were formed before the era of general incorporation laws.

Valley to New York.⁹⁸ Because of the relatively great change in elevation, and low level of urbanization of the areas it traversed, the canal had both greater expenses and lower revenue than the Delaware and Raritan.

While its promoters insisted that under state ownership it would cost the state almost nothing, and be an extremely large source of future revenue, they were unable to develop a majority in the state legislature. In the end, a bill was passed creating a dual banking/canal corporation with the banking subsidizing the transportation.⁹⁹ Various restrictions were placed on the firm to ensure the canal would in fact be built. These restrictions were necessary to the canal's completion, as it was apparent to the corporate officers that the canal would not generate enough revenue to cover the construction costs.

After various cost overruns (the final costs were more than two hundred percent of the estimates), the canal was eventually completed. After one business failure in 1841, the canal became profitable in the 1860's, but by the end of the decade the completion of a rival rail road line eliminated profits once again. The canal did continue operations for another fifty years.

⁹⁸ Goodrich, p. 116.

⁹⁹ The bill's passage was in the midst of a banking charter frenzy. Banks were perceived as very profitable, and state legislatures were unwilling to provide charters without some compensation.

Conventional Analysis

H. Jerome Cranmer, in his essay on the New Jersey canals,¹⁰⁰ repeats a distinction he attributes to the early writers on canals. He considers a canal a “developmental” project if its purpose was “the economic development of the region traversed,” and he considered the construction of such a canal justifiable even if it never proved profitable, as the result of the economic development it precipitated. Such canals, he reports, must be owned or subsidized by a government. In opposition to this are “exploitative” canals, which are likely to prove profitable almost immediately. These, he claims, may be provided by either for profit corporations, or revenue seeking governments.

Developmental canals are supposed to provide transportation to a region without it. In doing so, they convert other assets, such as land, from valueless to valuable. These canals must be subsidized or owned by the government, because the value they create is in higher land/industry value, which can not be captured in the rates they charge.

An exploitative canal supplies lower cost transportation to a region that is already “developed.” Because markets already exist for its services, it can be profitable

¹⁰⁰ This a chapter in Goodridge 1961. See page 157 for the developmental/exploitative distinction. The facts relating to canals are drawn from this essay and others in Goodrich 1961. The particular arguments in which these facts are used are my own.

immediately.¹⁰¹ It does not cause a widespread increase in the value of land and industry - much of the new wealth created by an exploitative canal is captured in the toll. Because of this, private corporations may be used as methods for constructing these canals.

It is argued that the Erie, Pennsylvania Mainline, and Morris canals were all developmental, and that the Delaware and Raritan and the English canals were all exploitative. As I will argue later, I do not believe this distinction, and the corresponding prediction for governmental versus private action, is well supported by the facts.

More Contemporary Analysis

Using somewhat more contemporary language, many economists argue for governmental subsidy when demand is such that the revenue generated from an asset is insufficient to cover its cost - yet the total value is greater than the total cost.

¹⁰¹ One element of the distinction is to argue that private corporations will not build developmental canals because, since the area is not yet developed, there is not any current traffic to carry. This is a very incomplete, and probably mistaken idea. It does not matter, to a private corporation, if the profits resulting from current expenditures are in the future; the appropriate comparison is between the present value of the costs and the present value of the revenue. To argue that only the government can build developmental canals for this reason is to argue that the government has a substantially lower discount rate.

For example Figure 1a could be construed as a exploitative canal, relative to figure 1b.¹⁰² In each diagram, marginal cost is constant, which is probably a satisfactory approximation for canals. The fixed cost of constructing and maintaining the canal is not shown, nor is average cost, which must be declining.

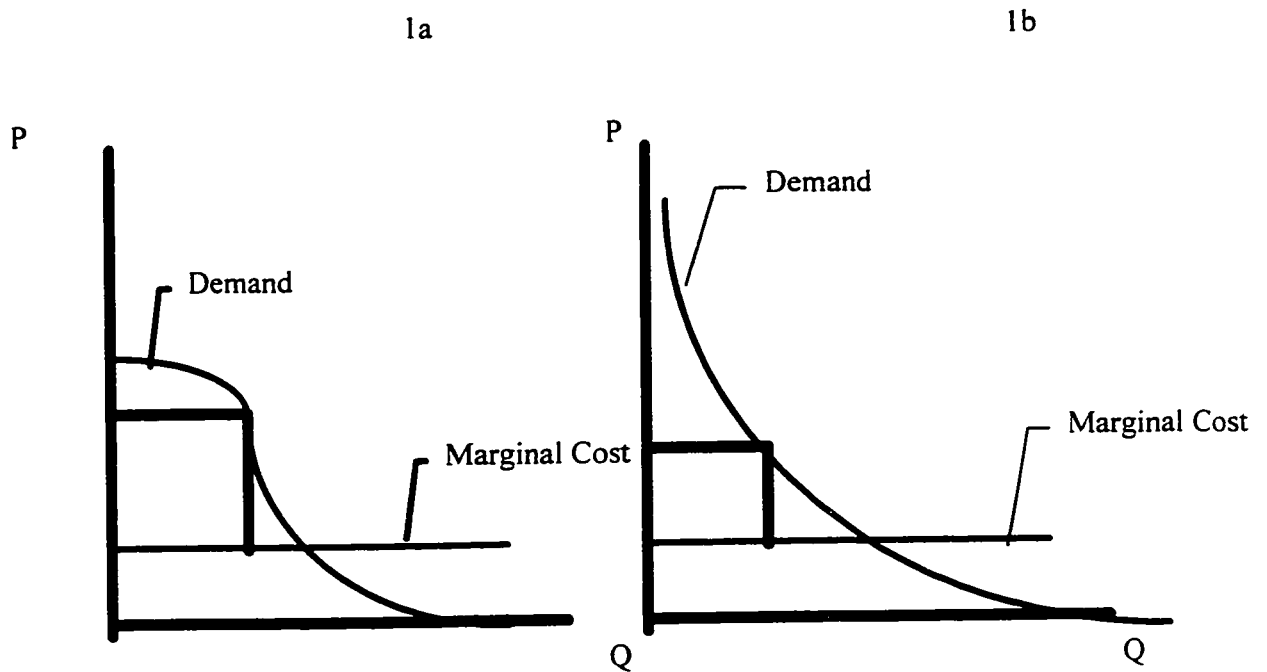


Figure 1: Demand for Profitable and Unprofitable Canals

¹⁰² These diagrams, and the demand analysis, are taken directly from Spence 1976. Their application to the developmental/exploitative distinction is original to this work.

The only difference between the diagrams is demand. Figure 1a, the exploitative canal, has relatively elastic demand. This might occur due to the availability of higher cost substitutes, such as coastal shipping or teamsters. Figure 1b, the developmental canal, has relatively inelastic demand. Upper New York State farmers, due to the low purchase price of their land, might have received considerable consumer surplus from the Erie Canal. But due to the continuum of distance from New York City, and perhaps the extent of through traffic from the Midwest, reservation prices are rapidly declining.

What is important to note is that the revenue above varying costs, or equivalently the profit less fixed costs, is larger for the *elastic* demand curve. This is the shaded box in each diagram. Given two canal projects with the same construction costs, the canal facing the more elastic demand (the exploitative canal) would be more likely to be profitable than the one facing inelastic demand (the developmental canal).

This reasoning - that the wrong product is being selected by the criteria of profitability - has been used extensively by economists in analyzing the efficiency of the market system. The common form of argument - as seen for example in Tyrol¹⁰³ - is to compare the (theoretical) behavior of for-profit firms to some "social optimum."

¹⁰³ For example Tyrol, p. 284.

I would like to take a slightly different approach. If a particular form of business activity - say private development of canals - is inefficient, this implies that there is some other feasible alternative that results in greater total wealth. Why did the organizers of the activity not select the contract that yielded greater wealth? Was there really a *feasible* alternative given the existing constraints?

If there exists inefficiency, there also exists a profit opportunity. What contract will entrepreneurs take in order to exploit these opportunities?

In this way I view institutions (governments, corporations) as contracts - as alternative arrangements that individuals select in order to maximize their own benefits. By this approach corporations can not be "greedy," since they do not have any will of their own. Likewise governments are not "beneficent," since they are only the tools of individuals maximizing their own utility.¹⁰⁴

This approach undermines, I hope, the view that governments exist to solve market failure. and that the appropriate role of the economist is to direct the government in its intervention.

¹⁰⁴ This approach is an organizing principle. I hope it is useful in making sense of economic history, and in predicting future action. It will be useful because it restricts attention to certain types of behavior (the economic, the rational, etc.). But for this same reason it will be wrong - if it is viewed as a complete description. Governments, in my experience, may be just and moral, or immoral. But viewing thought the ethical lens has not gotten us very far in the past.

But it also should undermine the idea that private action is superior to public action - for it questions the very nature of the distinction.

For while the public/private distinction has been much trumpeted, what could it really mean? Proponents of government action generally claim to be acting for "the public good" - even though certain projects benefit some individuals and impose costs on others - who counts as "the public"? Likewise, many of these same proponents claim their project is designed to benefit "people not corporations", knowing the entire time that corporations are composed of people (workers, shareholders, managers, etc.).

On the other side of the public/private divide, there were few large scale developmental projects that did not involve public actions. The Delaware and Raritan canal was privately owned. But the corporation was created by an act of the New Jersey legislature. Various fees were paid to the state government. The corporation received various legal privileges, including a monopoly exclusion, rights of eminent domain, and assistance in acquiring water rights from Pennsylvania. The state, because of the large revenue it derived from the project, was a residual claimant to the project. Was this really a "private" canal?

My point is that there are numerous contractual arrangements that may be used in the development of an asset, including different margins that may be priced, and different methods of paying for non-marginal costs. Each contract leaves some non-priced margins and resulting inefficiencies.

I will attempt to demonstrate some of the various margins left in the public domain by alternate contractual forms, and show how alternative contracts were selected as the value of these inefficiencies changed from situation to situation. In doing so, let me divide different canal projects by their profitability under different institutional arrangements:

1. Canals that were profitable and owned by private corporations.
2. Canals that were profitable, and owned by the government.
3. Canals that were not profitable under either arrangement, and the total value was less than the total cost.
4. Canals that were not be profitable under either arrangement, and their total value was greater than the total cost.

Canals that maximized profit under private ownership

Under most conditions, an asset owned by a private corporation will be more valuable than if it is owned by a government. Why is this?

Each form of ownership may be thought of as a share contract. In each case capital is pooled. When a corporation is formed to develop an asset, individuals buy stock, and have a

claim on a share of future profits. In the case where a state government builds a canal, citizens are taxed for the construction costs, and will pay lower taxes if the project turns a profit. So why do I expect that most assets will generate greater profits if they are owned by corporations instead of governments?

There are two categories of reasons for this. First, private corporations are more likely to select profitable projects - to develop assets that are profitable. Next, private corporations are more likely to operate assets in a way that maximizes profits.

Why are private corporations more likely to select profitable projects? Consider the organizers of a canal planning its route. A private corporation will concern itself primarily with profitability - selecting a route that maximizes the difference between revenue and cost. Each independent feeder will be evaluated purely on the amount of traffic it will add as compared to the cost of its construction. There will not exist any incentive to include locations that will not be profitable to serve.¹⁰⁵

¹⁰⁵ This analysis applies only when the shareholders are investing because they expect to derive revenue from their holding of canal shares - not by the existence of the canal itself. In the case of the river improvement companies that preceded the Erie canal, landowners that would benefit from the improvements provided much of the capital. But as is well understood in the Public Finance literature, public goods are less likely to be provided by private investors as the number of investors increases - the free rider problem dominates. Based on these considerations, it seems that as the number of investors per asset increases, the direct profitability of the asset must also increase if it is to be provided without government subsidy.

In contrast to a private corporation selecting the route of a canal (or deciding what attributes of an asset to develop), a state developed canal will face different constraints. With government development and ownership of assets, it is normally the case that some individuals receive disproportionate benefit from the development of an asset. In the case of developmental canals, for example, property owners along the canal route receive direct benefit, while property owners in other sections of the state pay a portion of the cost of the canal without receiving any benefit.

Given majority voting and large information costs as to the relative merits of various projects, it is possible that a majority will support a project that is not profitable, because it is a vehicle to transfer assets from the tax paying minority to the benefit receiving majority. In lobbying for a particular project, the beneficiaries are expected to compare the benefits to themselves versus the costs to themselves, ignoring the tax burden on others.

Because of the constraints of coalition building, organizers will at times select routes not because they are profitable, but instead because of the value of the extra votes in building the coalition.

Not only will private corporations select assets more likely to be profitable, but they are also likely to *operate* them to maximize profits. Before an investor purchases shares in a corporation, he must be convinced that the management will not use the corporation for its

benefit - or at least that the constraints on management will be significant. Contracts such as stock options and other executive compensation based on profitability align (imperfectly) management interest with profitability.

Furthermore corporate ownership is much more concentrated than state ownership (in the sense of number of shareholders versus number of tax payers), providing stronger incentives for shareholders to monitor efficient operation. Extremely inefficient management may be expelled by means of one party purchasing enough stock to take corporate control. An analogous process exists in the case of government owned assets - a political entrepreneur may seek control from voters in exchange for decreasing inefficiency - but since the payoff (getting elected) is not proportional to the profit saved, this process will be underused.

For various reasons, the managers of government owned assets rarely have their compensation tied to profitability. In many cases the asset is not designed to produce revenue - for example freeways or infantry battalions - so profitability is not an issue. I will discuss these cases below. Even in institutions designed to raise revenue, such as the Internal Revenue Service or Customs Service, maximizing profits is rarely an institutional goal.

If these conclusions are correct, then I expect that canals built in order to raise revenue - profits - will be owned by private corporations. If there exists two competing sets of canal promoters, one for a private development of a canal, and one for state development of the

same canal - then the private canal group will be able to make larger offers for the assets needed in construction, including state approval.

Even in cases where governments have cost advantages in ownership, there are reasons to expect private ownership of profitable assets. As I have previously discussed, organizers of corporate production become, at least to a large degree, residual claimants to the profitability of the asset. Even if the same asset may be produced at a lower cost by a government agency, the organizers of government production will be residual claimants in a much less direct way, resulting in lower power incentives to develop assets.

This point is well illustrated in by the example of monopoly pricing. We all understand that the gains from trade are greater from marginal cost instead of monopoly pricing. Government agencies are more likely to select a price closer to marginal cost, for the direct compensation of the decision makers will not increase by monopoly pricing. If costs functions were the same for private corporations and governmental agencies (generally an unreasonable assumption), then it would seem that there are larger gains from trade from government ownership.

But these gains are not concentrated. The gain of efficient pricing is spread out among purchasers. If purchasers are price takers - an assumption I have been making throughout this analysis - then each purchaser has a trivial incentive to organize production. So even in cases where governments have a cost advantage, since organizers of government

production have low powered incentives, such assets are less likely to be owned by governments.

This is also the logic behind the Clayton act, which allows triple damages for parties hurt by antitrust violations. It is fairly obvious under these circumstance that private parties have stronger incentives to act. Yet governments do own assets and bring law suits. A correct comparison is between the share contract incentives offered by corporations, and the various incentives offered government agents. I will explore this latter topic in the section on unprofitable assets.

Canals that maximized profits under state ownership

While the profitability of assets will, for the reasons I have just described, generally be maximized when owned by corporations, there are obvious exceptions to this. Governments have several economic advantages, and sometimes these advantages outweigh the increase in other ownership costs.

One advantage is that the interest rates most (state) governments must pay are substantially less than those paid by private corporations. When the major continuing costs of a project, such as canals, are interest on initial sunk costs, governments may have a cost advantage in owning an asset. Of course this is limited by the fact that many internal improvements in the United States were built by private corporations with bonds guaranteed by state governments.

In this manner some of the cost minimizing behavior of private corporations was matched with low government interest rates.

Another government cost advantage relates to bargaining power. State governments have the power of eminent domain, which may be used to avoid holdout problems in negotiating the purchase of rights of way. Similarly, an early private attempt to build the Delaware and Raritan was foiled by the corporation's inability to gain water rights to the Delaware River. Once again, however, these powers may be "sold" by the state to private companies.

One further cost advantage of state ownership comes out of the contracting literature. Absent risk aversion, the party with the most control over the variability in the value of an asset should be the residual claimant. When is the state the party with the most control over the value of an asset?

First of all, states that are *not* bound by the rule of law usually have the most control over the value of assets (or some categories of assets) within their domain. Since these states have the power to seize assets of value, other parties are unlikely to develop assets just to have them confiscated. Projects such as canals, which are composed of substantial sunk costs, are ideal candidates for confiscation. Since private individuals or corporations will under-invest in these circumstances,¹⁰⁶ states may own these assets at a lower cost.

¹⁰⁶ See Grossman and Hart.

Even in rule of law countries, confiscation via taxes is always a possibility.¹⁰⁷ Indeed, in industries where costs and technology are well understood - where first best results may be imposed on an agent by a principal - the cost of confiscation and government ownership are small, leading to under investment on the part of private individuals. This may be the reason that so many government projects seem stagnant (without major innovation). It is assumed that the government is not innovative, but it may instead be the case that only the government will invest in projects in which innovation is not an important attribute.

Low cost metering is another cost advantage of governments, but one that is not particularly relevant in the case of canals. It was important in the case of other transportation infrastructure, as illustrated by the example of toll roads versus freeways.

For a toll road, a significant marginal cost of its operation was in metering its use. If clients were to be charged efficiently, they should have been charged for every mile they used the road. But since roads cover long distances, this perfect measuring of use required a large number of toll booths and other controls to prevent travelers from exiting before the booth and rejoining the road after the booth. Because of the high costs of measuring use, private toll roads were generally not profitable. This is an example of a case where even when the

¹⁰⁷ Note for example the recent tobacco legislation.

total value exceeded the total cost, total revenue from even the most efficient pricing schemes was insufficient.

The common “solution” to the road metering case is to have governments provide roads, and to have users pay for them via gasoline taxes, which are proportional to total usage. The disadvantage to this system is that prices will not be used fully in determining either the location or metering of roads (hence crowding), since the gasoline pump does not measure where the gas is used. And as discussed in the case of profitable canals, governments are less likely to select road locations or usage that maximize profits.

Why did private corporations not use this metering technique? In other examples, corporations have used another complementary good for metering (IBM and punch cards). The difficulty would be controlling the supply of the related good. A gasoline company could build roads to increase the demand for its product, but then customers would use the gasoline company’s roads and a rival’s gasoline. Unless the alternative product is also supplied monopolistically, and is less expensive to meter, the tie in will not work.

Neither the railroads nor canals faced this metering problem. In each case exit and entry from the road was costly, and therefore owners spent few resources ensuring clients paid for their use. Railroads in particular avoided these metering expenses by two methods; first, they provided the transportation itself (vertically integrated into railroad engines), and when they did allow non-company engines on their track, the number was sufficiently small that

metering was not expensive. This is probably a major reason why railroads were privately built and operated, and also why railroad companies were vertically integrated into both the road and the train.

Canals avoided metering problems because of the large expense of entering and exiting the canal. Few users would find it worthwhile to avoid the toll by such methods, and therefore few resources (true transaction costs) were spent ensuring payment. *Expensive metering* was not a reason governments had a cost advantage in owning canals.

Profitable Canals Reexamined

Of the four canals I have previously described, two were profitable. One was owned by a private corporation, and the other by a state government, and the characteristics of each canal conform to the institutional considerations that determine ownership.

The Delaware and Raritan was an attractive investment for a private corporation. The canal had guaranteed traffic, since it connected Philadelphia and New York City. It had few obstacles to overcome. The Erie canal was in the process of being completed, and therefore a domestic work force had been trained in canal construction.

These same characteristics were common to the profitable English canals. By analogy alone, it was a business venture that should have been profitable (unlike other analogies drawn

between the English canals and the Morris canal, or between the Erie and Pennsylvania Mainline).

Yet at the same time, there were few reasons for state ownership. It was not a satisfactory mechanism for redistribution of wealth within the state, for it provided transportation services to merchants in New York and Philadelphia. Only the owners of the canal, not its within state users, would receive benefits. So why did not the state government own it, and in doing so receive the rents that went to the owners?

First off, there remained substantial risk. Two previous corporation had been formed to build the canal, and both had failed. Canals were not yet a proven method for states to raise revenue, although the Erie was just beginning to demonstrate some promise. There is evidence that the representatives of other sections of New Jersey believed that the canal posed risk of higher property tax rates statewide.

Also, the state government was able to receive rents from the canal for attributes it already owned. The canal corporation needed a charter from the state, as well as aid in acquiring access to water from the Delaware river. In return for the purchase of these attributes from the state, the corporation paid both a lump sum fee, and a share of the revenue in the form of state taxes on its shipping.

The state did not own this canal, for its promoters were able to profit by its direct ownership. It was a money making venture - and if the state owned the canal, the profits would have been far more diluted among taxpayers than among shareholders of a corporation.

One way to view the contract under which the canal was created is as a share contract. The capitalists provided the initial investment in the canal, and retained control over its operation and pricing (although there were state imposed maximum tolls). By retaining this control, the canal was more likely to be profitable than under state ownership.

The state provided a charter, powers of eminent domain, legal monopoly power, and assistance in bargaining with the State of Pennsylvania (which owned the water rights to the Delaware). But the state did not have control over construction or continuing operations. It received a share in the future revenue as an incentive to avoid appropriating the canal after the initial investment was sunk.

The Erie canal is an example of a canal that was profitable under government ownership. The tolls from the canal - which were actually reduced a number of times - raised enough revenue to pay for the operating expenses, the interest on the construction bonds, the initial construction costs, improvements, and a substantial contribution to the state treasury.

Could the canal also have been profitable for a private corporation? Clearly it could have been run as a profitable venture. Toll collection and upkeep did not require any attributes not

held by private corporations. And certainly some elements of the initial construction could have been completed more profitably by a for-profit corporation - for example the elimination of certain feeder canals that were included only as an inducement to counties that would not have otherwise benefited from the canal.

But two factors make it unlikely that the Erie canal could have been profitably developed by a private corporation. The first is a credit constraint. The second relates to preventing competition.

The Erie canal was, for its time, a tremendously large project. The capital required for its construction was provided by private English investors - it is doubtful that any other group had sufficient capital. Could a private US corporation have raised this much capital? The State of New York could pledge not only the value of (unknown) future tolls against its canal bonds, but also its ability to raise taxes. This greatly decreased the risk of default, and the associated risk premium. New York State could borrow at a lower cost than any private canal company, and the cost of borrowing was a significant expense of canal construction.¹⁰⁸

¹⁰⁸ The greater wealth - guarantee capital - of the state was one reason for the lower risk premium. Perhaps more importantly, English investors knew that the New York State government would be a frequent borrower, resulting in repeat interaction (or reputation) and lower risk of default.

The other potential cost advantage of state ownership relates to preventing competition. After the completion of the canal, the state government was very careful to ensure that New York railroads did not compete directly with the canal. Railroads that did run along the canal route were forced to pay canal tolls on the goods they hauled. These restrictions were eliminated only after other railroads, in other states, overcame the technical difficulties of crossing the Appalachian chain. The state government, with its ability to eliminate some forms of competition, was able to increase the expected profitability of the canal. In doing so it reduced the risk premium it was forced to pay on canal bonds.

Canals that were not profitable under either public or private ownership, yet the total value was greater than the total cost.

Canal promoters often claimed that certain canals would be economically worthwhile even if they were unprofitable. Figure 2, which is figure 1b with the addition of average cost, illustrates why this might occur. In this case there does not exist a price on the demand curve which is above average cost. Yet at many prices the loss from producing the good (the difference between the average cost and the price, times the quantity) is less than the consumer surplus - meaning the total value is greater than the total cost.

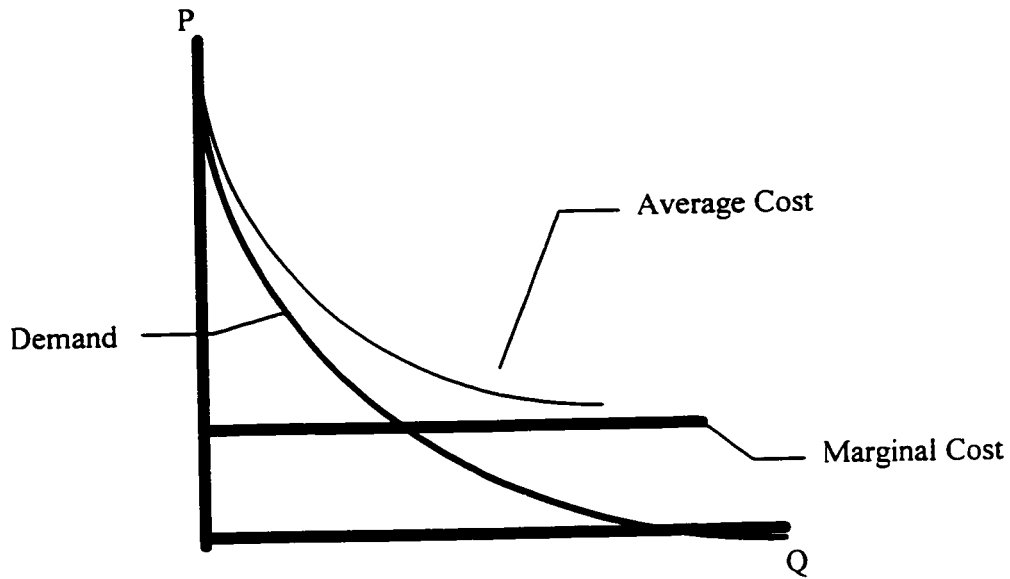


Figure 2: Demand with Average Cost

In fact the quantity that maximizes the net value of this good is where the marginal cost curve crosses the demand curve, since once the initial fixed cost is incurred, it is efficient to produce any units for which the marginal value is greater than the marginal cost. This

illustrates, using demand and costs but not any explanation of why they are as portrayed, an example of a good which is not profitable, yet is efficient to produce.

Graphically, it is not difficult to calculate when the total value of a canal is less than its total cost (at the optimum point). If we integrate the demand curve from zero to the quantity where the marginal cost curve crosses the demand curve, we have the optimum total value. The average cost, times this quantity, gives us the total cost. If the total value is less than the total cost, the product is inefficient to produce (inefficient for who, I am not sure). There are various complications - does the demand curve capture all of the value of the good; does the cost curve capture all of the opportunity costs of production - and some of these may not be swept under the rug.

A single price per unit is only one form of contract. It has the advantage of inexpensive use - the seller does not have to acquire any information concerning the buyer. But it has the disadvantage that it only allows the seller to acquire a constant amount of revenue per unit, when the marginal value per unit varies. It also has the disadvantage that it creates the monopoly distortion - leaving some gains from trade unexploited.

A pricing scheme that does not have these disadvantages is one where consumers are charged the marginal cost of production, plus a lump sum fee to cover fixed costs.

This solution is often approximated in situations in which the consumers provide the capital investment. A club of which I am a member, for example, charges marginal cost pricing for

the use of its facilities and the goods it sells, and then has an annual membership fee. Heavy users of the facilities are encouraged to “volunteer” for work parties.

These arrangements also work in more commercial settings. For example early attempts to improve water transportation in upper New York State were attempted by private companies without the prospect of substantial (or positive) profits. These companies’ shares were purchased by landowners along the routes of improvement. Volunteer fire companies, private roads in housing developments, and all sorts of services provided by clubs are other examples.

But clubs are not a universal solution. Given low bargaining costs brought about by a small number of beneficiaries, these non-governmental programs work well, both because sufficient investment is assembled, and because the residual claimants have sufficient incentive to monitor their agents. But if the beneficiaries of a project are large in number, then voluntary contributions rarely are sufficient. For this reason the private corporations developing water transport in New York were never very successful, while the state sponsored Erie was able to raise sufficient capital to complete a project of tremendous size.

This analysis has some predictive power. As a general rule, assets of this type will only be built privately if the number beneficiaries is reasonably small. Small size is important, because it reduces the cost of measuring valuations. With a “club good” and marginal cost pricing, high valuation consumers must pay more of the non-variable costs than low marginal

valuation consumers - this is what makes it economical to produce the good at all. With a small number of users, sorting users by valuation is less expensive. Thus we see enthusiasts paying more, in the form of donating time, being elected to office, etc. As the number of consumers increases, free riding dominates, and this form of organization becomes unlikely.

I have noted a few examples where transportation improvements took on a form similar to club goods. But unless land holdings are very concentrated, this form is unlikely to be selected, and history demonstrates that it generally was not.

Even when there are too many consumers for a "club" type of contract, uniform pricing is still not the only alternative. If consumers may be inexpensively sorted by their valuation, and arbitrage prevented, price discrimination will generate greater revenue than a single price. Was there something about the demand for transportation services that made price discrimination costly?

The Railroads, in providing transportation services to the West, widely practiced price discrimination. Since the increase in the value of land was proportional to the distance to markets, fares in excess of marginal cost could be proportional to distance and capture much of this value.

Consider, for example, land on the frontier. Property near existing transportation services would obviously be worth more than land farther away; farmers would have to pay teamsters

to ship their output, and this cost would increase with distance, decreasing the value of land. After a certain point, land would be worthless - even in present value terms. Now consider the addition of a canal to the frontier. The land that was the farthest from the markets would increase in value the most, while farmers closer to the markets would still have alternative (more expensive) transportation services that would limit the increase in the value of their land brought about by the addition of a canal. A tariff that was linear in distance would capture much of the increase in land values.

The two standard problems with price discrimination - the difficulty in sorting customers by marginal value and the difficulty in preventing arbitrage - should not have been a problem in the case of canals. A linear tariff, combined with discounts for different classes of goods (i.e. coal versus fresh produce), should have allowed sorting by marginal value, and periodic toll booths should have prevented arbitrage (as opposed to having one ticket for both short and long trips on the canal). Since price discrimination should have worked,¹⁰⁹ why would a canal not be profitable, yet create net surplus?

¹⁰⁹ Limited price discrimination was in fact practiced by the canals. The example of Railroad Price discrimination was also somewhat more complicated than I have described here. See my chapter on the economic development of the US.

Canals with total benefits less than the total cost.

According to the Coase theorem, canals that cost more to construct and operate than produce in total benefits, in expected terms, should not exist, unless there are substantial bargaining costs or poorly defined property rights. After all, bargaining should lead to the efficient solution.

When a government owned canal is built, there is a transfer of wealth from taxpayers in general to those who benefit from the canal's existence. Since the canals I am describing produce total benefits less than the total costs, the value the beneficiaries receive is less than the amount the taxpayers lose. For this reason government inefficiency in canal development is something of a mystery that needs to be resolved.

Given that a canal is, in expected terms, not going to be profitable, what form of organization or ownership do we expect? Assuming we do not have the sort of asset that may be provided by clubs (see the previous section), organizers are going to have to approach governments for subsidies. Government ownership or subsidy will have two effects. First, it will either create or destroy wealth, depending on the efficiency of the project. Next, it will result in wealth transfer from tax payers to recipients of consumer surplus.

When the beneficiaries are easily identified and taxed in proportion to their marginal valuation, these projects do not incur opposition. The difficulty is that the characteristics that

allow this - easy recognition of marginal benefit - also make price discrimination inexpensive, which means these projects are likely to be profitable and undertaken by a private organization. As a general rule, then, I expect governments to only subsidize those projects where price discrimination is expensive, which in turn requires transfers from non-beneficiaries.

Since votes may be traded, groups of individuals will be willing to subsidize each other (i.e. I will support your project that requires general taxation if you support mine). To the extent that projects create larger amounts of wealth, trading allows entrepreneurs to increase the scale of their activities. Of course government trading is not entirely analogous to the market; voting allows the majority to coerce resources from minorities - exchange may be one sided. In this way majority coalitions may be formed involving projects that in each case transfer wealth from the majority to a minority - as long as a sufficient number of minority recipients exchange votes. This is probably what drives the creation of political parties.

Those minorities that are net losers from this process will incur expenses - true transaction costs - to minimize their losses. But in the case of projects that are efficient, the beneficiaries will gain more from the transfer than the others lose. This means those parties that are net losers on a project will not make better offers to coalition partners - because the beneficiaries are able to outbid them.

To the extent that projects destroy wealth and only result in transfer, I expect that those not benefiting will be able to outbid their rivals for the coalition support. This is a powerful force that results in the government supporting efficient but unprofitable projects.

So, once again, why are inefficient projects undertaken? The obvious reason that assets in this category (net wealth destroyers) are developed is that it is expensive to distinguish between these projects and those in the previous category - projects that are not profitable yet the total benefits are greater than the total cost. If benefits were inexpensive to measure, then price discrimination would also be relatively inexpensive - and profit maximizing corporations would be selected. Why is it so difficult to measure the net benefits of a canal?

Since much of the expense of a canal is the initial cost of construction, and since these construction expenses varied tremendously across time and space, there is a substantial difference between expected and actual. Even after a canal was built and operated, it is difficult to determine whether it was efficient to produce. There exists a large and indecisive literature, with contributions by distinguished economists including Douglas North, questioning the efficiency of various railroads and canals. If current economists, with our more advanced knowledge of theoretical and practical problems in measuring net gain, are unable to say precisely if a canal was *ex post* efficient, then what sort of accuracy do we expect the individuals making the decision had?

Furthermore, the analysis we are doing now questions whether a canal was efficient after its completion. The relevant considerations for the decision makers is whether a proposed canal was efficient in expected terms. Varying estimates of net gain obviously could shift almost any project from net gain to net loss - or the opposite. Are we surprised that some canals were built that created more in costs than in value?

Both promoters and detractors will make attempts at measuring the cost and benefits of a project, both for the purpose of deciding what it is worth spending to pass or defeat the project, and also for propaganda purposes. But case by case measurement is expensive and inaccurate, and I expect rule of thumb measures, based on collective experience, to prevail in deciding which projects will be undertaken by different branches of governments.

Classes of projects that have been successful in the past will be included in the future. Types of projects that have proven to be wealth destroying will not be attempted. Changes in demand and cost will switch projects from one category to another, and eventually legislatures will follow suit. In the case of new types of assets, inexpensive test projects will be supported, and future funding will depend on initial success.¹¹⁰

¹¹⁰ As noted by Yoram Barzel, these implications are not very operational (subject to falsification).

At any given time then, I would expect subsidized projects to be a mixture of the efficient and inefficient. To the extent that it becomes easier to determine that a type of project is inefficient, then that category is less likely to be constructed. To the extent that the cost or valuation of projects change, the sorts of projects receiving subsidies will change. This explanation, rather than changing ideology, is probably at the core of changes in government involvement in infrastructure.

The Morris Canal

The Morris canal illustrates many of these points. First, promoters will attempt to use the government when their project appears to be unprofitable in expected terms. Next, promoters and detractors will each develop cost estimates biased in their favor. Those losing wealth from such projects will incur transaction costs in order to minimize these losses. Finally, coalitions will be assembled in order to turn projects benefiting minorities into larger projects benefiting majorities.

The foremost entrepreneur behind the creation of the Morris canal was the president of the Morris County Agricultural Society, George P. M'Culloch. In his propaganda, he estimated the construction costs would be \$300,000 or less, and gross revenue would be \$81,000 per year. This is a substantial return on investment (27%), so why did not M'Culloch set up a corporation to exploit these profits?

M'Culloch, in arguing for state construction of the canal, claimed that the benefits of the canals should go to "the people" instead of speculators. But, for several reasons, this is almost certainly *not* the reason why he approached the state legislature. First, even if the canal did have this profit potential, M'Culloch would have had to cut the state government into the profits. There were not any general incorporation laws at the time, which means a bill would have to be passed creating a corporation to construct the canal - state approval was needed. The promoters of the Delaware and Raritan found that this cost its owners a \$100,000 contribution to the state.

But M'Culloch did not just want to buy various property rights of the state government for his own use - he wanted the state to pay for and own the canal. Why? Even to George M'Culloch, and contrary to his writings, the canal would not be profitable. *Ex Post*, the canal cost \$4 million to construct, and after being completed in 1837, failed in 1841.¹¹¹ *Ex Ante*, the canal was also not a profitable investment. M'Culloch's correspondence indicates he was aware that investors perceived such canals (developmental) as unlikely to succeed, which means he was unlikely to find sufficient capital.

¹¹¹ The canal was purchased by a transportation company in 1844, was enlarged, and made profits of around 7% in the 1860's. But by the late 60's railroad competition eliminated its profitability.

Furthermore canal opponents easily savaged his revenue estimates. Most of his claims of revenue depended on coal shipments from the Lehigh valley, but in order to generate the forecasted revenue, coal would have to replace other fuels entirely in New Jersey and New York industry (which it turned out, did happen), and no other sources of supply or transport would become available.

Given that the project would not be profitable, could M'Culloch have organized its construction in some other method than a government subsidy? The beneficiaries of the canal would have been farmers, landowners, and industrialists that received lower cost transportation services. Since New Jersey had been settled for several hundred years already, land was already owned by a large number of different farmers, making it unlikely that M'Culloch could have received a subsidy from a concentrated number of land speculators. The relatively large number of beneficiaries, and resulting free rider problem, probably made an entirely cooperative organization impossible.

Given that the canal needed a subsidy - a transfer from the majority - its promoters needed to build a coalition. An attempt was made to politically link the project to the Delaware and Raritan, but since that canal was expected to be profitable, its promoters had nothing to gain by the linkage.

Eventually a coalition was achieved, but instead of a government owned canal, a corporation was chartered that had banking privileges dependent on its expenditures for canal construction.

This arrangement had the advantage that it did not directly require tax increases for the subsidy, and the state government was not residual claimant to a canal failure. But why were the canal promoters successful in linking their project to a profitable bank, when they had been unsuccessful in linking it to the profitable Delaware and Raritan?

While I do not understand the complexities of banking incorporations in early nineteenth century America, I hypothesize that there were significant externalities to banking speculation. Allowing the incorporation of a bank (which did most of its business in New York) was much like a tax that transferred wealth from the citizens of the state to the bank's shareholders. The bank would be profitable, yet it needed coalition support for incorporation, since its existence imposed costs on others.

So there was a trade between supporters of the Morris canal and organizers of the bank. While other sections of the state were able to prevent direct subsidy (transfer) to the canal project, the costs of the bank were sufficiently obscure that with only the direct support of the canal beneficiaries, a majority coalition was built. Of course the bank supporters could have instead contributed cash to the state - reducing the general level of taxation. This is what occurred with the Delaware and Raritan. It may have been the case that each project would not have passed on its own - but together they formed a majority coalition.

The Pennsylvania Mainline

The Pennsylvania Mainline also had the characteristics that would only allow it to be constructed by government subsidy. The process of its development illustrates various inefficient transfers that may result from government ownership.

The canal's promoters made a number of familiar claims. Three quoted by Rubin include;

"We have unquestionably the best route for a canal between the eastern and western waters," wrote the majority of the Canal Commissioners to the Governor in November 1824.

The Governor in turn told the legislature, it was the "cheapest and best route."

And the foremost promoter claimed during the public debate "The Pennsylvania canal will have immense advantages over the Erie."

Since by the time of these debates, local traffic alone was paying the interest on the Erie's canal debt, why were these supporters besieging the legislature for a government owned canal instead of a corporate charter? Was the Pennsylvania Mainline the sort of project the would be most profitable under government ownership?

The canal's proponents hoped to create value for the merchants of Philadelphia by retaining a substantial share of the Western trade that would soon be diverted to the Erie. Due to the

much larger expense of developing a canal over mountainous Pennsylvania, the canal was certain to not be profitable. Its various disadvantages were widely known - everyone knew that the canal would have to cross a mountain range avoided by the Erie.

Why was the canal not built by the city of Philadelphia, where the proposed benefits would be concentrated? After all, even if the free rider problem prevented a voluntary project, local ownership would have ensured that the total benefits were greater than the total cost.

Of course that is exactly the point. The City of Philadelphia did not build the canal for its merchants, for the simple reason that the canal was not efficient - the total cost of the canal was less than the benefits to Philadelphia merchants. In later years, the City did in fact invest in the Pennsylvania Railroad - which proved profitable.

The only way to gain a transfer from the rest of the state was to link the project to others also needing subsidies. Thus, when one looks at the map of mountainous Pennsylvania, one is struck by the sheer number of canals. When the state began construction of the 395 mile Mainline, it also was forced to construct 314 miles of lateral lines.¹¹² Some of these eventually proved profitable in moving iron ore and coal. But in 1824 this was probably unforeseen.

¹¹² Segal, p. 178.

Even with the linkage to other questionable projects, its proponents had a difficult time gaining a majority. This is perhaps why absurd comparisons were made with the Erie, absurd in that the cost estimates were far too low, and the revenue possibilities far too great (the Mainline's proponents claimed it would generate sufficient revenue to pay for an education for every child in New Jersey - while in the case of the Morris canal, revenue would endow Princeton to such a degree that it would be the world's leading university).¹¹³

The Pennsylvania mainline was a transfer from the general taxpayers of Pennsylvania to merchants of Philadelphia. Since the Mainline itself could not support a majority coalition, the various ancillary canals were included in the construction project, and these canal provided benefits to additional citizens. The backers of the mainline promised, however, that the canal would generate revenue beyond its costs. This did not prove to be the case, and the resulting transfer of wealth obviously resulted in constraints on future projects.

Those who were net losers from the canals were unable to prevent the construction, but they were able to avoid payment by direct taxation. The canal system was paid for via bonds, many of which were sold to English investors. When the promised canal revenue did not

¹¹³ The proponents of the Erie canal were much more modest in their claims, as reported in Rubin.

appear, the state was forced to default on these bonds.¹¹⁴ At this point it was apparent that the canal was a mechanism for wealth transfer, not creation, and those losing wealth took measures to prevent continuing losses.

Pennsylvania had, as the result of the canals, the largest construction debt in the US.¹¹⁵ Constraints on continuing losses included authorization for the sale of all state owned canal properties. The Mainline experience demonstrated that canals were net wealth destroyers - so those losing wealth prevented state involvement in these types of projects. Taxes were also imposed on corporate salaries and "trades, occupations and professions," which presumably had the effect of returning some of the net losses from the users of the canals to the farmers in the rest of the state. This demonstrates that types of projects proven to be net wealth destroyers will eventually be eliminated from the realm of state action.

Conclusion

Over time, we observe improvements such as canals were constructed by both private and governmental organizations. One explanation for the variance in ownership patterns is ideological - government came in and out of fashion. The problem with this approach is that it is analogous to change in taste arguments for changes in demand - they may occur but are

¹¹⁴ This same story was repeated in a number of other states at the same time.

¹¹⁵ Segal, p. 203.

not predictable. Explanations that instead depend on changes in measurable phenomena - particularly cost - leave us with operational theories.

In this chapter I have argued that the pattern of ownership of canals conforms to explanations based on ownership costs. For reasons that center on tying the payment of organizers to profitability, corporations were more likely to select canal routes that maximized profits, and to operate canals at minimum cost. Unlike the example of England, there were few possibilities for the profitable operation of a canal in early 19th century America, but the example of the Delaware and Raritan, with high demand and few natural obstacles, is one example that accords with my expectations.

An example of an exception to corporate ownership advantage is illustrated by the history of the Erie canal. The State of New York was able to operate the canal at a profit. But imitation by other states demonstrated that it was factors particular to the geography and timing of the Erie that made it profitable.

New York and other governments built canals under conditions that were not favorable to profitability, and created canals that were either efficient or inefficient. As the result of experience with inefficient canals, various constraints were imposed on future projects that would result in the net loss of wealth. While we do observe that governments first entered then exited the canal business, this change may be explained by changing opportunities and increased understanding of the costs of alternative contractual forms.

CHAPTER 4: STANDARD OIL OF NEW JERSEY

The most fundamental question of antitrust analysis, and topic of this chapter, is this: how do firms become monopolists? How do firms initially acquire monopolies? How do they keep rivals from entering their market? Is monopoly “natural” in many industry, and therefore inevitable and inefficient to prosecute? Or is competition between firms the natural state of industry, with monopoly the result of inefficient and sinister behavior?

Obviously these topics are too broad to answer in one chapter (or one book), so this chapter will analyze the narrower topic of how the Standard Oil Company acquired and maintained an almost complete monopoly in the refining of crude oil during the last quarter of the nineteenth century. It will begin with select economic facts describing the oil industry at the time of Standard’s formation. Competing explanations of Standard’s monopoly position will be offered and criticized. It will end with a portrayal of Standard Oil’s political and legal battles.

Oil Industry Structure in the 1860’s.¹¹⁶

¹¹⁶ See the appendix for a more complete explanation of the production of kerosine.

With the expansion of population and industrial technology, the demand for liquid fuel used in lighting increased throughout the 19th century. The most common¹¹⁷ initial product, whale oil, eventually suffered severe supply limitations, resulting in a switch to various substitutes, especially kerosene. Since the same economic phenomena also occurred in Europe, which did not have an independent source of petroleum until the late 1870's, the demand for United States kerosene exceeded the growth of the United States market.

Today petroleum is primarily produced by large, vertically integrated firms, which perform every task from searching for crude oil to selling it at corner gas stations. But this was initially not the case; separate firms discovered and extracted the crude oil, transported it to the refineries, refined it, and transported it from refineries to markets. The industry then included three (not double counting transportation) vertically related types of firms.

Crude Oil

Entry into the business of crude oil discovery and production was relatively inexpensive, and had few economies of scale. Oil was discovered by drilling test holes in likely places, and the cost of extracting the discovered crude was relatively small. Since pooling of capital was

¹¹⁷ This is not quite correct. Camphene, a redistilled spirit spirit of turpentine, was developed in the 1830's and became the most popular lamp illuminant in the 1840's. It was substantially cheaper than whale oil, but suffered the disadvantage of a low ignition point, making it quite dangerous. See Williamson and Daum, pp. 33-34.

unimportant, the product was fairly homogenous, and the local drilling areas were heterogeneous, moral hazard considerations imply that firm size would be small (see chapter one). If local managers had not been owners, but instead paid employees of one large firm, they would have an incentive to shirk.¹¹⁸ Predictably, small firms dominated.

Most crude was found in the oil fields of Western Pennsylvania (and later Ohio and Indiana); these fields were the only significant source of U.S. crude during the nineteenth century. Europe was eventually (1870's) supplied by some crude from the Baku region of Russia, and the discovery of oil in Texas (the gulf area), California and Oklahoma occurred at the same time as Standard lost its monopoly position in refining (the second decade of this century).

Refining

Entry into the crude refining business¹¹⁹ was also economical on a small scale. The simple technology¹²⁰ again allowed efficient operation at a level which required little pooling of

¹¹⁸For example, shirking could take the form of digging test wells in areas that are more easily accessible. If local managers are paid the same amount whether they dig test wells in easy or difficult areas, they will presumably pick the easier areas.

¹¹⁹Until 1900, most crude oil was refined into kerosene used in lighting. Other products included lubricants, solvents, heating oil and medicinal oils. Most of the kerosene was exported (69% in the 1880's). It was not until 1900 that the demand for gasoline appeared. Chandler, p. 92.

capital. For example in 1870, when Standard incorporated, its works were the largest in the world. But the largest in the world still only meant four percent¹²¹ of United States capacity, and equal to approximately the combined size of the next three largest refineries in Cleveland.

Plant capacity did steadily increase throughout the century, with Standard's Cleveland plant doubling its capacity between 1865 and 1869.¹²² By the end of the century, Standard's largest plant (located at Whiting, Indiana) was capable of refining 36,000 barrels of crude per day; the largest refineries in the 1860's processed 2,000 barrels *per week*.¹²³

Transportation

The transportation section of the industry involved two stages; transporting crude oil from the wells to the refineries, and transporting the refined kerosene to market. Early in the industry's history, crude was placed in wooden barrels, which were then transported by teamsters across mud roads to either river boats (on the Allegheny River) for shipment to

¹²⁰ Refining crude into kerosene involved separating out the kerosene from the other elements by means of distillation, and then chemical treatment to remove sulfuric compounds. Since the kerosene made up only 50-70% of the crude oil, refining substantially reduced the volume of crude. See the appendix for additional details.

¹²¹ Bringhurst p. 10 claims Standard had 10 percent of U.S. capacity, but cites Ida Tarbell's critical *History of the Standard Oil Company*.

¹²² Granitz p. 1 and Chandler p. 93.

Pittsburgh, or to railheads for transportation to refineries in Cleveland, Philadelphia or New York.¹²⁴

The refined product was also placed in wooden barrels, which were then shipped to markets in the East (for either direct consumption or further shipment to Europe). Most refined product was shipped by Railroad, with the exception that Cleveland refiners had the option of using the Erie Canal (when it was not frozen over), and New York and Philadelphia were already located in the principal markets.

Within a decade of the initial discovery of oil, bulk shipments began to reduce transportation costs. First, pipelines (gathering lines) replaced teamsters in transporting crude from the wells to the railroads.¹²⁵ Railroads responded by transporting bulk crude in tank cars.

By the mid 1880's, railroads themselves began to be replaced, in the transportation of crude, by long distance pipelines. They continued to transport coal and other inputs to refineries, and ship refined products to the market. Their market share in this last category decreased with the eventual creation of pipelines from refineries to major markets.

¹²³ Williamson & Daum, p. 228, 628.

¹²⁴ These, as well as the oil region itself, were the principal locations of refineries. A few existed in locations such as Boston and Maine.

¹²⁵ This is a fairly unambiguous case of a monopoly (a pipeline) replacing a competitive industry (teamsters). It is also an example of both Bertrand and "limit" pricing; the price fell just enough to prevent reentry. Williamson & Daum, p. 188.

The oil industry was served by three “trunk” lines,¹²⁶ which connected the grain producing Midwest to the markets of the East, and a number of smaller “feeder” lines, which connected the more remote local areas to the feeder lines. The three trunk lines were competitors for traffic to and from the Midwest, which resulted in lower (per mile) rates for any traffic classified as “through”, as opposed to “local”.

When two or more railroads served the same route, pricing became somewhat more complicated than either competitive or monopoly industries. In the conventional industry analysis popularized by Jacob Viner, the industry price will equal long run average cost, which will equal marginal cost. Firms do not have an incentive to undercut each other’s prices, since each is already operating at marginal cost.

But in the case of the railroad, costs were composed of large fixed costs (payments for the tract and engines, etc., most of which was sunk) and very small marginal costs (extra fuel and wages). In competition each railroad firm had an incentive to undercut rivals price to some level above marginal cost, but below average cost.

¹²⁶ They were: the Pennsylvania Railroad, which ran through Southern Pennsylvania; the Erie Railroad, which ran through Southern New York; and the New York Central Railroad, which ran North of the Erie Canal. See Williamson & Daun, p. 171.

This follows from the fact that the railroads had to pay the fixed costs regardless of how many miles of service they sold. Any price above marginal cost increased profits (or decreased losses). Thus price competition often lead to below average cost pricing, and losses for the competing railroads.¹²⁷ The obvious solution to the problem was collusion.¹²⁸

¹²⁷This simple analysis is, at the very least, incomplete. Why would the railroads incur sunk costs if they did not have some reassurance that they could cover their average costs? Railroads, and their investors, were very concerned with the danger of appropriating quasi-rents. For an example from another context, consider railroad charters. Railroad general incorporation (charter) laws of the 1850's allowed state control over maximum rates, as well as other legislative control over corporations formed under these general laws. Effectively no railroads were built under these laws; all projects brought to completion were formed under special charters guaranteeing the railroads freedom to select their own prices. See Miller, p. 47.

Why then did the railroads ever put themselves in such a position? First, entire transcontinental lines were not build from the ground up. Instead, long distance lines were formed from the merger of many local lines. The actual extra expenditures for entering the oil fields were relatively small. Next, the capitalists who created these railroad companies both anticipated and put into practice agreements to limit price competition. While these agreements were imperfect, resulting in occasional losses, in general railroads were successful businesses.

¹²⁸ Is there a non-collusive equilibrium? During this period in history the courts would not enforce contracts "in restraint of trade" (see chapter 2), but a group of firms would generally not be prosecuted by a government organization if they agreed on a price. While there may also be an equilibrium with only tacit collusion (where the firms never meet to discuss price, but still follow a strategy of not undercutting each other), the explicit collusion is easier to achieve and yields results that are at least as profitable. Is there an equilibrium that does not involve any type of collusion, explicit or tacit? Is there a core? One requirement would seem to be rising marginal costs, since otherwise there is always some incentive to undercut the market price.

Due to the limited number of railroads serving any given area,¹²⁹ the costs of negotiating a collusive price was relatively low. But even with a price agreement in place, each firm had an incentive to “cheat” by secretly undercutting its rivals’ price.¹³⁰ The decision to cheat on a collusive price was then a comparison of the long term profits from continuing in the collusion, versus the short term profits of capturing a large percentage of the market at a price just below the collusive one, followed by potentially below cost pricing as the agreement broke down.¹³¹

The cycle of railroad pricing followed this pattern. Railroads would undercut each other’s price in order to expand sales. Eventually the price was driven down below average but above marginal cost. Firms would lose money, until they once more agreed to charge an

¹²⁹There were three major railroads serving the western Pennsylvania oil fields, The New York Central, the Erie, and the Pennsylvania. All three provided access to Eastern Ports, allowing kerosene to be shipped to Europe.

¹³⁰ Why did the railroads depend on collusion when they could have signed a contract with a large penalty for cheating? The use of contracts depends critically on the availability of a third party with the power to enforce the terms of the contract. Contracts without this third party must be self enforcing, meaning that no party to the contract has an incentive to cheat. Governments, particularly court systems, often will enforce contracts. In the nineteenth century, the U.S. common law deemed price fixing illegal, and would not enforce contracts including it. This did not make it a crime to engage in price fixing; this followed the passage of the Sherman Act.

¹³¹ This is a situation in which dynamic analysis is required. Cooperative behavior in pricing is the most basic dynamic analysis taught in game theory.

equal but high price. Then a new event would trigger cheating, which would return the market to price competition.¹³²

One example of this is the *Empire Rate War* of 1876-77. The Empire Transportation Company, a firm associated with the Pennsylvania Railroad, began purchasing and constructing refineries. This would allow the Pennsylvania to increase its market share, since it would be transporting its own oil and kerosene, which would be difficult to monitor by rival shippers. After talks between the Pennsylvania and Standard (which was receiving large discounts from all shippers) broke down, Standard stopped using the Pennsylvania Railroad.

The Pennsylvania responded by cutting shipping rates to crude producers and independent refineries, leading to a price war between all three railroads. The Erie Railroad lowered its rate for crude to 35 cents per barrel, less than one third of the pre-price war rate. After substantial losses were experienced by all three railroads¹³³ a new agreement was negotiated.

¹³²This explanation is consistent with both the folk theorem and the results of Green and Porter. Many events could lead to the breakdown in collusive pricing, but the common element is change, be it in demand or supply conditions. The second half of the nineteenth century was a time of radical economic change in the U.S., which implies that collusive arrangements were comparatively more difficult to enforce.

¹³³Granitz p. 33-36.

Most of the oil region rate wars occurred before Standard's rise, but it must be noted that the Empire Rate War was only one episode in the history of railroad price competition. The entire system of secret rebates, short and long haul rates, and general price discrimination (see chapter 2) effected every industry dependent on transportation, and was generally referred to as the "railroad problem". It was not "resolved" until the creation of the Interstate Commerce and the regulation of railroad rates prevented extensive price competition.¹³⁴

Quasirents

While railroads had the problem of rate wars, it was only at this final stage of production, that is transportation, that a single party had the capability of capturing quasirents. At any other stage, any attempt to raise price substantially above cost would lead to new entry and the undercutting of the price. With inexpensive entry into the crude production and refining stages,¹³⁵ producers could only expect to make normal returns on average.

What made collusion and above normal returns possible in transportation was the great expense, legal difficulties, and time required for acquiring rights of way and construction of additional track. Any firm that did enter would have to do so on such a large scale that the

¹³⁴ The Hepburn Act of 1906 gave the Interstate Commerce Commission power to set railroad rates. This limited the use of price discrimination, and was supported by some railroads that were being forced by large shippers into giving discounts. See Frey p. 39.

resulting market price would likely decrease. If the three railroads acted in unison, they could be monopoly suppliers for extended periods.

*A Brief History of Standard Oil*¹³⁶

John D. Rockefeller, a successful Cleveland commodities merchant, began investing in oil refineries in the early 1860's. By 1870, when he and his business partners formed the Standard Oil Company of Ohio, his two Cleveland refineries had an estimated capacity of 1500 barrels of crude daily.¹³⁷ Following a period of refining over capacity¹³⁸ and shutdown, in 1871-72 Rockefeller purchased or merged with twenty one Cleveland rivals, effectively the entire refining capacity of Cleveland, giving him twenty five percent of the national industry capacity.¹³⁹ During the same period Standard invested in extensive marketing, storage and transportation infrastructure.

Standard's ownership of Cleveland refiners closely linked it with the two Northern Railroads, which ran their trunk lines through Cleveland. This effectively created two "alliances", with

¹³⁵ Crude production also had a large number of producers, making collusion difficult. Refining also had a large number of firms until the consolidation of the 1870's.

¹³⁶ See the appendix for a more complete chronology.

¹³⁷ Williamson & Daum, p. 305.

¹³⁸ In 1871-72 total refining capacity (annual) was 12 million barrels; refinery receipts were between five and six million barrels. Williamson & Daum, p. 344.

¹³⁹ Williamson & Daum, pp. 352-3.

Standard providing the bulk of refining for the Northern railroads, while the remaining (and dominant) Pennsylvania Railroad used independent refineries in Pittsburgh.

While avoiding the direct ownership or development of oil wells, Standard extended its operations into gathering lines, connecting oil wells to the railroads. In 1873 Standard bought a one third interest in the United Pipe Line Company and formed its own American Transfer Company.¹⁴⁰ These were not the largest pipeline companies; the Empire Transportation Company, associated with the Pennsylvania Railroad, owned the largest networks.

Standard continued to expand its refinery capacity through both new construction and the purchase of, or merger with, rival refineries. There remains some controversy at what price Standard acquired its rivals' assets; often it seemed to be below construction costs, yet many of these refineries were unprofitable because of industry over capacity. Many of the leading refinery entrepreneurs sold their businesses to Standard and became Standard executives.¹⁴¹

Following the Empire Rate War (1876-77), and the resulting consolidation of gathering lines under Standard's control, Standard purchased most of the remaining independent refineries. By 1878 Standard controlled through ownership or lease over 90 percent of United States

¹⁴⁰ Williamson & Daum, p. 412.

¹⁴¹ Williamson & Daum, p. 417.

refining capacity.¹⁴² Standard also had extensive control of transportation, both through its virtual monopoly over gathering lines, and through its ownership or lease of railroad cars used in bulk transportation.

Standard's control over transportation was threatened in 1879 by the completion of the Tidewater long distance pipeline, which directly connected the oil fields to the coast. The railroads immediately began a rate war with Tidewater¹⁴³ (which, due to the lower cost of transporting crude by pipeline, they were bound to lose). Upon observing the successful operation of the Tidewater pipeline (of which there was substantial prior doubt), Standard built its own pipeline system. By 1882, the competing transportation systems came to terms, with Standard purchasing a minority interest in Tidewater.

The rest of the century saw continuing battles of this pattern, with ever increasing crude output and refinery capacity, and a shift in new crude discoveries from Pennsylvania to Indiana and Ohio. Standard continued its majority position in gathering lines, refining capacity, and marketing; but it did so in the face of ever changing market conditions.

¹⁴² Williamson & Daum, p. 429.

¹⁴³ Standard executives claim that they opposed a rate war, and instead proposed cooperating with Tidewater. That is, in fact, what occurred, but only after Standard completed its own pipeline and after all parties lost money in price competition. Williamson & Daum, p. 447.

Standard never experienced “the quite life” that is the often claimed goal of monopolists. The extensive network of gathering lines in the Pennsylvania oil fields became less valuable as the fields became exhausted and new oil deposits were discovered in Ohio and Indiana. Standard’s market share in Europe and Asia decreased as Russian, then Dutch Sumatra crude oil penetrated both markets. Finally, Standard was involved in almost continuous legal conflict (see below).¹⁴⁴

Competing Explanations of Standard’s Monopoly Position

Standard’s success has been explained with four, not necessarily exclusive, explanations. They include economies of scale, predatory pricing, merger to monopoly, and enforcement agent for a railroad cartel.

1. Economies of Scale.

When the average cost of production changes with the total amount produced, we call this economies of scale. There are positive economies when, as production increases, average cost decrease, and diseconomies of scale when increasing production increases average cost. Economists normally explain the size of firms within an industry by appealing to economies

¹⁴⁴ The first appendix to this chapter contains additional events in Standard’s business history.

of scale; firms with the lowest cost win the price competition and drive inefficient size firms out of business. Thus the firms we observe are either the low cost producers or are losing money.¹⁴⁵

Our initial presumption then, is that Standard Oil became as large as it did because that was the efficient scale of production. According to this explanation, it is historical coincidence that it was Standard, and not a rival, that grew to this size, since any firm that purchased the capacity to produce at a large scale would have had a cost advantage over smaller rivals.

What sort of cost savings did Standard experience that gave it an advantage over its rivals?¹⁴⁶

First off, Standard exploited plant level economies - cost savings that resulted from using

¹⁴⁵ Stigler (pp. 71-94) proposes the "survivor test" as a measurement of efficient size. He divides existing firms into various categories by percent of industry output, and then tracks over successive years whether the firms within a given category produce more or less of industry output. A given size is efficient if it tends to produce more of total output over time. For example, if small firms (.5% of total output each) initially make 30% of industry output (there are 60 of them), and later make 90% of total industry output, this is an efficient size. His main conclusion, after applying this test to several industries, is that there is a wide range of optimum sizes: "the long-run marginal and average cost curves of the firms are customarily horizontal over a long range of sizes." This of course undermines the usefulness of economies of scale arguments in explaining the size of any particular firm.

¹⁴⁶ And what type of training is appropriate to discover and report economies of scale? Engineers certainly are better able to report on the economies of mechanical processes than economists. But as I have argued in chapter one, costs are not simply the result of mechanical processes (an apple cost x amount for fertilizer, y for the harvester, etc). Costs vary with contractual arrangement, which is the topic of economics. In the end, I must conclude that it is the combination of engineering and economics that is best able to study economies of scale. It is businessmen that are likely to discover economies of scale.

plants of the greatest capacities. Standard Oil started with the largest scale plants and constantly increase its plant size.¹⁴⁷ Standard's initial plant in Cleveland decreased refining costs by 40%, from five to three cents per gallon.¹⁴⁸ With the 1882 reorganization of the Trust, Standard produced almost twenty five percent of the world's kerosene in three plants.¹⁴⁹

Standard finally lost its monopoly position in refining when two events occurred: new oil fields were discovered on the Texas coastline, and other firms built refineries there that were as large as Standard's plants. The second factor provides some evidence for economies of scale arguments, for it might be argued that demand increased to such a point that the efficient sized firm could not satisfy industry demand, and therefore other firms were able to enter.

In chapter one, I present the several arguments why productive processes involving many inputs (like oil refineries) may most economically be owned by one firm. If an oil refinery contains substantial elements of team production, then one party will most likely own the

¹⁴⁷ Refineries, or at least individual distillation tanks, definitely experienced first increasing, then decreasing returns to scale. Since one or two laborers could run a one hundred or one thousand gallon still with equal ease, there existed economies of scale in labor as stills became larger. But diseconomies quickly appeared; for example the ratio of surface area to volume decreases as a tank increases in size, and this ratio was of some importance to the techniques of refining.

¹⁴⁸ Chandler, p. 93.

¹⁴⁹ Chandler, p. 25.

entire plant. While sweeping the floors may not include team production (and therefore may be contracted out), the interdependence of the various refining tasks seems¹⁵⁰ to require team production.

Other explanations, including the measurement of the product at different stages of production and the danger of working with equipment that essentially boils a flammable liquid, suggest that workers will be paid for their inputs and that one party will be residual claimant. This suggests that the firm will be large enough to own an entire plant. Standard was so big because its refinery was large.

But these economies of scale that arise from economies of plant size do not quite get us to an explanation of Standard Oil's size, since Standard was always much larger than any of its plants. We must also appeal to economies of scale that are unrelated to plant size. What other economies are there? This is equivalent to asking what other costs do oil refiners incur. One is management costs. Is it possible that the cost of making decisions is subject to increasing returns? Is the cost of managing two plants less than two times the cost of managing one? If the production process is fairly homogenous, and communication costs are low, then decisions made at one plant are likely also to apply to another. The same applies to the cost of research into productive processes, consumer preferences, etc.

¹⁵⁰ I really ought to say 'requires team production' instead of seems. When one does not have evidence for a statement, and it is very expensive to obtain that evidence, it is best to be bold.

Chandler attributes almost all of Standard's success to being the first firm to achieve economies in plant and organization size. He points out that Standard was the first to "rationalize" its corporate structure, develop a centralized marketing department, and develop over seas markets.

His explanation is insufficient for two related reasons: first, he does not explain what the technological economies of scale are, and second, he does not explain why having a larger firm gives access to these scale economies. The next section explains Standard's success without depending on economies of scale.

2. Predatory pricing.

When the Supreme court found Standard Oil of New Jersey guilty of monopolizing the oil industry, it did so primarily on the grounds of illegal practices. The Court did not believe that Standard grew to its large size because of economies of scale, but instead because it damaged or destroyed its rivals. Its weapon was predatory pricing.

As argued below, there is considerable controversy and confusion as to what constitutes predatory pricing. Since we are concerned with explaining how Standard acquired and maintained its virtual monopoly, we will start with the traditional explanation of Predatory Pricing.

Start with two firms in a market, a predator and a prey. Initially they are engaged in some sort of price competition,¹⁵¹ but each would prefer to be the sole seller. The predator is willing to do something to bring this about. All of this occurs in period zero.

In period one, the predator implements the following strategy. It sets a price below each of their costs (to be defined below). It will maintain this low price until the prey exits the market.¹⁵² In period two, after the prey exits, the predator will set the monopoly price. If any other firm enters the market, it will again set a below cost price.

This strategy is clearly a losing one if the prey does not exit (we never get to period two). It will be profitable, although not necessarily optimal, if the discounted period two profits exceed the period one losses. The question then, is whether the prey will exit.

¹⁵¹ If the predator is first producing in the market and the prey wants to enter, we have an entry deterrent model. Many predatory pricing applications start with the prey already in the market, after which a large predator enters.

¹⁵² In the case of preventing the entry of a new rival by predatory pricing, and in the jargon of the literature, this is deterring entry. If the initial firm in the market decides to allow its rival to enter (or does not drive it out), this is accommodating entry. If entry is accommodated, each firm will want to charge high prices, which means that the initial firm wants to signal, before its rival enters, that it has high costs. If the initial firm wants to deter entry, it instead will want to signal that it has low costs (perhaps through predatory pricing), which implies future industry profits will be low if its rival enters. With slight qualification, this leads to the *puppy dog, top dog, fat cat, and lean and hungry look* dichotomy presented in Tirole, p. 327.

The “long purse” strategy claims the prey will have to exit, because its wealth and ability to borrow are less than the predators (this would be a condition for the strategy to work). The predator will force the prey to exit by eliminating its revenue for a sufficient length of time to drive it into bankruptcy. A somewhat more refined argument is that the predator need only indicate its ability and willingness to incur these expenses; the prey will avoid bankruptcy by exiting the market as soon as it perceives both the strategy and capability of the predator. It gains nothing by remaining in the market until bankruptcy.

The controversy with predatory price arguments is that the results of predatory pricing are very difficult to distinguish from the operation of the sort of price competition espoused in the antitrust laws. The advantage of competitive markets is that prices are driven to marginal cost via sellers undercutting each other’s bids. But how can firms compete if they can be prosecuted for destroying competition?

Suppose, for example, that the two firms each have constant (but different) marginal cost, with the predator’s cost below the prey’s. For fairly normal demand conditions, the predator’s best strategy is to price just below the prey’s cost,¹⁵³ thus capturing the entire market. Is this predatory pricing?

¹⁵³for as long as the prey is in the market. After the prey exits, the price is either the monopoly price, or the contestable market price, depending on the structure of the model. This is simply the Bertrand model with different marginal costs. See chapter 1.

Or suppose, as in the case of Standard Oil, we have one seller (the predator) in multiple markets, while its rivals are each located only in one market. Given similar demand conditions, we expect the price to be higher in markets without rivals than in markets with rivals *even if the predator is not trying to drive out its rivals*. If the predator's costs include a substantial fixed component, then the price in markets with rivals may even be below average costs (but not average variable costs); profits increase as long as only the marginal costs are covered. But these are exactly the circumstances that bring charges of predatory pricing.

The conclusion that Standard grew to its large size because of predatory pricing was seriously challenged by John McGee. He provided both theoretical and empirical evidence against the claim that Standard Oil acquired and maintained its monopoly by means of predatory pricing.

Extending some of the analysis of Aaron Director, he questioned if predatory pricing was a superior strategy to merger. A firm facing a below market price may anticipate that it is being preyed up in order to monopolize the market. This implies the price will rise in the future, and so profits are available if the prey stays in the market. Effectively, the tradeoff between present losses and future gains is equivalent for the predator and the prey.

Perhaps more importantly, the prey may reduce its short term losses by restricting output. But the predator, in order to maintain the low price, must expand output. And since each unit is causing losses marginally, this furthers the loss for the predator.

In order to raise prices above the competitive level, a would-be monopolist must both eliminate rivals and keep them out. Driving firms into bankruptcy is often insufficient, since rival's plant and equipment may be purchased by new firms seeking the above normal returns of monopoly. This may explain why Standard made it a point of dismantling many of the refineries of competitors it purchased.¹⁵⁴ The general point, which the courts have recognized, is that there must exist some "barriers to entry" that prevent new entry.

In cases where a large firm buys its rivals, as Standard bought competing refineries, predation is a poor strategy if the predator intends to retain its prey's former owners as managers. This was often the case with Standard, and as McGee notes, rivals who have been driven into bankruptcy are unlikely to make loyal employees.

¹⁵⁴ There are other explanations for this. Standard may have had the lowest refining costs, but it is the combination of demand and its rivals costs that limit what it can charge. Buy buying and dismantling rival's inefficient plants, it may raise its own price without fear of an inefficient rival reentering the market.

As compared to predatory pricing, McGee notes that the predator could instead buy out the prey, and that this strategy would cost each party less (except the customers), since it does not involve any period of below cost selling. It is also likely to make for better employees.

Empirically, McGee examined the trial records of Standard Oil, and found that there were few allegations of predatory pricing, and the few that did exist were unsubstantiated. Price wars did occur between Standard and its rivals, but they were often started by rivals trying to increase their market share. McGee does note that some refineries exited the market because Standard received a larger discount from the railroads, and that Standard would purchase their refineries.

McGee, who published his work in the late 1950's,¹⁵⁵ was quite successful in criticizing the prevailing understanding of predatory pricing. Given good information concerning demand and costs, predatory pricing does not seem to be a superior strategy to merger. The traditional "drive them into bankruptcy" strategy does not seem to be the method by which J.D. Rockefeller acquired his monopoly on refining. Initially Rockefeller did not have significantly more wealth than his rival Cleveland refiners, which makes the long purse story unlikely. It certainly seems to be that case that his rivals (and future partners) sold because he would price below their costs; but this was because the railroads charged Rockefeller less for transportation services, giving him a cost advantage.

¹⁵⁵McGee, 1958.

But this does not imply that predatory pricing is irrelevant to antitrust, or this case. Even if Rockefeller did not initially drive out competing refiners through predatory pricing, this does not mean Standard did not use it selectively once it had acquired a long purse, or that other firms have not used it in other circumstances.

For while merger might be a superior strategy, it is now illegal under circumstances equivalent to the Cleveland refining market in the 1870's.¹⁵⁶ While predatory pricing is also illegal under the Clayton and Robinson-Patman act, it is much more difficult to detect and prove. The Supreme Court has yet to even endorse a precise definition of what counts as predatory pricing, although necessary conditions seem to include structural barriers to entry and a price below average variable cost. Merger might have been a better strategy for Rockefeller, but because of changed merger laws and the difficulty of proving predatory pricing, it is not necessarily a superior strategy option now.

Recent game theory analysis has demonstrated that under certain circumstances of incomplete information, predatory pricing is an equilibrium strategy.¹⁵⁷ Alvin Klevorick,¹⁵⁸

¹⁵⁶See the appendix to this chapter for information on current merger law.

¹⁵⁷This criteria for the reasonableness of a strategy is somewhat murky. Using our standard notion of equilibrium (Nash), we simply mean that no agent has an incentive to play any move other than the one in our strategy if every other agent plays the strategy. In other words, no agent has an incentive to 'deviate' alone. Nash equilibrium was supposed to help us with predicting what agents would do. The problem is that many strategies are Nash

in a review of the recent economic literature on predatory pricing, identified three types of predatory pricing models. The first updates the traditional Long-Purse arguments. These have always depended on asymmetry between the firms, in that the predator had greater financial resources, either in corporate wealth or access to credit, and may therefore outlast smaller rival in their spending contest.

The asymmetry in access to credit was a source of criticism by McGee and other Chicago economists. After all, a firm facing predatory pricing might be a very good investment indeed; it suffers smaller losses than the predator during the conflict, and if the prey survives (as it will it has sufficient credit), it seems likely that the market price will rise in the future. This is particularly true if the predator acts as a dominant firm and later restricts output; in that case the smaller rival can increase sales at a higher price, while the dominant firm must reduce sales in order to raise the price (see chapter five for an explanation of the dominant firm model).

More modern models have supplied conditions under which there is a differential access to (or different prices for) credit. According to Klevorick, these models still depend on asymmetry, but it is generally asymmetry of information. The predator has better

equilibrium - so how do we decide if our prediction is the most useful one? This is what refinements are for; trying to sort between a multitude of equilibria, all of them Nash. My point is that merely being a Nash equilibrium does not mean that a strategy is the likely or reasonable one.

information as to its success than do banks or other lenders. Because of this lenders must charge a higher premium to the prey than experienced by the predator (particularly if the predator already has capital). A predator knows that it is investing in future profits; a banker only sees current losses.

The second type of predatory pricing model is one of reputation. Under many cost circumstances a predator is *able* to drive its prey into bankruptcy; the question has always been whether it is worth the expense of doing so. If it is not a profitable strategy, then predators are unlikely to follow through with threats of below cost pricing, which means that firms facing those threats need not take them seriously.

If the predator creates a situation in which it will suffer a large financial loss if it gives up a predatory price battle, then it will be willing to spend up to the amount of that loss in order to win the battle. That is exactly how a reputation for predation works. If a firm sells in many markets, and has unique rivals in each market, then driving one of them out of business may create a reputation for toughness.

If the predator gives up in one conflict, its rivals in other similar markets may interpret this as a sign of weakness (limited access to resources, etc.), which may lead to failure in all

¹⁵⁸Klevorick 1993.

markets. In this situation the predatory has a strong incentive to win whenever it starts a predatory contest, even if the losses exceed the gains in that market.

Klevorick's last group of economic models again depend on asymmetry in information, but in this case cost information. If a large firm can satisfy all of demand at a lower marginal cost than its smaller rivals, the small firms will probably exit the market. Predator's will make some effort to manipulate what the prey knows about its costs. In particular, the predator may charge a low price¹⁵⁹ in order convince the rival that its costs are low, and that the rival will lose money if it remains in the market.

Predatory pricing is now accepted by economists as an equilibrium strategy for some models. It is not clear how applicable these models are to particular antitrust cases. In the case of Standard Oil, it is difficult to believe that Rockefeller acquired his monopoly through predatory pricing, although he may have used it occasionally (the record does not show any really convincing cases) to eliminate new rivals.

The theoretical ambiguity is accompanied, not surprisingly, by legal ambiguity. The Supreme Court has not yet defined a single standard for predatory pricing, and lower courts use various standards. The Areeda-Turner test, which views marginal cost as the appropriate legal minimum, has been widely used. Since marginal cost is often impossible to measure,

¹⁵⁹ See Tirole p. 327.

Areeda-Turner suggest average variable cost as a legitimate proxy. This is a reasonable substitution if much of a firm's marginal cost curve is horizontal at minimum average cost.¹⁶⁰

But average variable cost is also often the wrong measure in predatory pricing. In a recent example, ValuJet, a discount airline, offered \$19 fairs.¹⁶¹ This amount is almost certainly below its average variable cost, and probably also below its marginal cost. Is ValuJet trying to monopolize the airline business? Are these predatory prices? Certainly not. One of ValuJet's planes crashed, killing all aboard, and the FAA grounded the airline for three months following allegations of poor maintenance. The low fairs were a method of convincing the public to resume flying ValuJet.

The Supreme Court has been clear about one of the requirements for predatory pricing; the Court has refused to consider predatory pricing an option if structural conditions, such as barriers to entry, do not make it possible for a predator to recoup its expenses after the price war.¹⁶² This is a rather minimalist test, and rules as to what counts as evidence of predatory pricing are likely to be a continuing topic before the court.

¹⁶⁰ See Stigler, p. 85.

¹⁶¹ Wall Street Journal, September 30, 1996, p. 1.

¹⁶² Klevorick, p. 164.

3. Merger to monopoly

The next explanation for Standard's success is merger. As asserted earlier, only the railroads were able to abstract quasirents, since at every other level of production there were multiple suppliers and inexpensive entry. Some economists have explained Standard's success as the result of changing the industrial structure.

Since the assets of the railroads had low salvage value, if all of the refiners were able to act in concert, they could extract many of the quasirents from railroads. Therefore the refiners' plants were more valuable under one management than under multiple competing ownerships. This is, of course, the other half of the profit equation. Economies of scale tell us how profits change through cost changes; revenue may also change with size via the mechanism of a reduced number of agents setting competing prices.

McGee argues that Standard succeeded because John Rockefeller was able to buy his competitors or give them stock in Standard in exchange for relinquishing control of their firms. Once he was in control of the refining stage of production, he could demand price reductions from the railroads. Rockefeller changed the industry structure from competitive refineries facing an oligopoly of three railroads, to three competing railroads facing a single monopsonist.

There are two difficulties with this explanation. First off, how did Rockefeller prevent new entry into the refining business? Clearly he received lower shipping rates than new entrants; but why did the railroads agree to these low rates, when they could have refused to lower their price and made profits from new entrants? The case history includes episodes in which the railroads stopped giving discounts to Standard. New refiners always appeared.¹⁶³

The other question is how did Rockefeller manage to so quickly buy all of his competitors in Cleveland. When Rockefeller started, his firm was only one of a number of Cleveland refiners. In the course of a year (1871), he became Cleveland's sole refiner. If we explain his ability to force discounts from the railroads based on his monopoly position in refining, then in turn we must ask how he acquired the monopoly position. Rockefeller may have been able to merge the competing refiners by exchanging Standard stock for control over their firms; but he should have experienced holdouts, since being a "fringe" firm (not a member of a cartel) is usually more profitable than being part of a cartel.

4. Standard as enforcement agent for RR collusion.

¹⁶³ As happened during the Empire Rate War. See Granitz & Klein, p. 33.

Ganitz and Klein have recently provided an alternative explanation for Standard's success.¹⁶⁴ They argue that the railroads created the Standard monopoly as a method of enforcing their collusive pricing agreement.

Competitors can increase profits by eliminating price competition. But once a price (or series of prices) is agreed on, each firm has some incentive to secretly undercut its rivals price in order to expand sales. While each firm is better off under the collusion than under full price competition, each is even better off if its rivals charge the high collusive price, and it expands its sales by selling slightly below this price.

A colluding firm then must weigh the benefit of "cheating" on the collusion against the cost. The benefit is the difference between the extra profits from undercutting its rivals, and the profits it would have made from maintaining the collusive price. The cost of cheating is the eventual breakdown of the agreement, as the firms' rivals discover that their lost sales were due to their rivals low price.

Anything that increases the benefits of cheating, or decreases the cost, will make the breakdown of collusion more likely. Factors that change these benefits and cost include the following:¹⁶⁵

¹⁶⁴ Their thesis is not original. See Williamson & Daum, p. 356.

¹⁶⁵ For other sources of factors thought to aid collusion, see Tirole p. 250 and Posner p. 55.

- a. Anything that makes the detection of price cuts easier increases the cost of cheating. If your rivals know instantly that you cut your price, they will immediately cut their own price. Industry groups that collect and disseminate price data are helpful in detecting cheating, as is the posting of prices in a manner available to everyone.
- b. If all customers make small orders, then the benefits of cheating are small. But if one customer makes a disproportionately large order, the benefits of cheating on that order increase. Therefore collusion is more likely in an industry characterized by many small customers instead of a few large ones.
- c. Variations in demand increase the benefits of cheating. If demand is above average now, then the benefits of cheating are relatively larger, and the cost of the agreement breaking down in the future is less.
- d. The cost of organizing a small number of collusive firms is generally less than organizing a larger number. Collusion in industries with a large number of firms is more costly to enforce, and therefore less likely.

The railroads had a history of forming collusive prices, followed by changes in management or other incidents leading to a renewal of price competition.¹⁶⁶ Given that history of cheating, each railroad knew that future agreements were likely to eventually break down, which provided further incentives to cheat now. What they needed was additional enforcement.

Ganitz and Klein argue that the railroads created the Standard refining monopoly as a method of enforcing their collusion. Standard, as the sole refiner, would divide its shipping between the three railroads. The railroads would be unable to cheat on the agreement individually, since Standard was the only customer. And Standard's monopoly position was entirely dependent on the rebates provided by the railroads, so it had an incentive to maintain the agreement. Since new entry into the refining stage was so easy, Standard would be unable to hold up the railroads.

This arrangement made collusion between the railroads easier by changing the benefits and costs. It made price detection easier by having only one buyer (*a* above). It eliminated the heterogeneity in purchasers (*b*) in the same manner. While it could not eliminate variations

¹⁶⁶Note the experience of the Southern Improvement Corp.

in demand (*c*), it could keep the number of railroads serving the oil industry small (*d*), since new railroads would not have existing customers to sell to.¹⁶⁷

It was Standard's control of transportation, via the railroads, that allowed it to limit the entry of competing refiners. Standard later gained control, through purchase and construction, of the pipelines that supplanted the railroads as transportation for both crude oil and kerosene. But while Standard developed an alternative to the railroads, it continued to ensure their cooperation through some shipping.

Ganitz and Klein provide considerable proof of this collusion between the rival railroads and Standard, including their attempt to form the Southern Improvement Corporation. This firm was explicitly incorporated for the purpose of enforcing the cartel as described above, but was disbanded (see below) before it began operations. Standard, which would have been the Cleveland agent of Southern, began its purchases before Southern's demise. Ganitz and Klein claim that Standard simply took on the role of the Southern Improvement Corporation.

While this explanation of Standard is admirable in its depth of historical detail (much like McGee), it also poses some puzzling questions. Trading the difficulties of arranging a cartel for the difficulties of selling to a monopsony buyer does not seem to be an improvement.

¹⁶⁷ Standard, as enforcer of the cartel, presumably would not purchase from any new railroads, since this would lead to a breakdown in the cartel agreement, and therefore to a loss of its monopoly buyer position.

Certainly there must have been pooling arrangements that both decreased the incentive to cheat and did not create a single buyer that could play the railroads off against each other. The railroads ended up with one buyer that they had to give discounts to.

So How Did Standard Gain its Monopoly Position?

We have four competing, but not exclusive, explanations of Standard oil's "Monopoly" position. Which one is correct? How does a firm become the sole seller of a product? What best explains Standard's success, in my opinion, is economies of scale; but not in the sense that there existed a decreasing average cost industry cost function that Standard was merely the first to exploit. Rather Rockefeller found a method of solving a contracting problem, and this made economies of scale in plant (refinery) size, as well as decreased shipping costs, available to him.

Throughout the 1860's both still and plant size had steadily increased, and by 1870 stills of approximately 500 barrels became most common (although there was some variation around this). This size still, and the technology appropriate for it, continued to dominate for the rest of the century. However since refineries could contain more than one still,¹⁶⁸ still size was not the determinant of refinery size. Instead economies existed because larger plants were

¹⁶⁸ Other important refinery equipment included tanks to store crude in, separate smaller stills for refining products other than kerosene, treatment tanks, and heating plants.

more flexible; multiple stills and tanks of differing sizes allowed a certain flexibility unavailable to smaller plants.¹⁶⁹

One element of proof that there existed unexploited refinery level economies of scale is simply that throughout the rest of the century refineries became larger and larger. This is also consistent with changes in technology - except that all of the basic technology of refining had been developed by 1870, and new technological changes (continuous flow, as opposed to batch processing) did not occur in the United States until the next century.¹⁷⁰

No party would be willing to incur the large sunk costs of constructing a plant when facing a monopoly (or cartel) supplier of transportation services, since the railroad could increase its tariff so that it, instead of the refinery owner, received the benefits of any cost reduction.¹⁷¹

Vertical integration did not seem to be the answer. Refineries did not buy railroads, presumably because oil transport was only a small portion of total railroad traffic. We know the railroads did not buy refineries,¹⁷² presumably because of the moral hazard problem. If

¹⁶⁹ Williamson & Daum, p. 283.

¹⁷⁰ Williamson & Daum, p. 253.

¹⁷¹ This argument requires some explanation of why a long term contract could not be signed before any costs were sunk. I believe the correct answer is that industry cost and demand conditions were too unstable to allow the correct fixing of a price *ex ante*.

¹⁷² This is not quite true. The Northern two railroads did occasionally own refineries, but they never were a large share of industry capacity. The Pennsylvania Railroad, through its

economies of scale in refining were to be taken advantage of, how were the refineries to protect themselves?¹⁷³

Rockefeller must have realized that if he was the sole purchaser of transportation services in Cleveland, the three railroads (two of which ran through Cleveland as part of their trunk line) would have a difficult time extracting all refining quasirents from him. Effectively Rockefeller was able to match monopsony against a small number cartel.¹⁷⁴

Rockefeller did not need to immediately monopolize the entire refining industry in order for this strategy to work. After purchasing his competitors in Cleveland, the railroads had the sunk costs of organization and track to Cleveland; Standard had its sunk investment in its plant and the purchase of its rivals. While the railroads could eliminate Standard's source of crude and access to market, Standard could play the three railroads off against each other. This was enough protection so that Standard was able to disassemble the high cost plants and expand the capacity of his new plants, taking advantage of economies of scale.

affiliate the Empire Transportation Company, owned a significant share of the industry's gathering lines (in direct rivalry to Standard), as well as refineries.

¹⁷³ Another solution to this problem of holdup is long term contracts. They were not used here, and in general are unlikely to be used when cost and demand conditions are subject to large variability.

¹⁷⁴ I am claiming that Standard's "protection" was its monopsony buying position, which is consistent with the merger to monopoly explanation. The "protection" could also be provided by the situation described by Granitz & Klein. See Snyder for a formal model consistent with this situation.

As the success of the arrangement became apparent, Rockefeller was able to both build larger plants and purchase former rivals (or trade stock). In this way the arrangement was self enforcing; since Standard's quasirents from its low cost plant were protected, its market share tended to rise. This gave it superior bargaining power, which allowed it to sink larger expenditures in even more efficient plants.

When technological changes allowed the use of pipelines, Standard either purchased (again often with stock) or built its own pipeline networks - for any other arrangement would have subject refinery investments to holdup. Rockefeller was able to purchase them, unlike the railroads, because they were dedicated assets - their only use was for transporting crude and refined products.

Rockefeller's success was due to economies of scale - but it was not simply economies in plant size. It was this combined with economies in ownership. Standard oil had lower costs than its rivals because it found an ownership structure that could exploit economies in plant size. As technological changes made pipelines the low cost alternative, Williamson's dedicated asset theory implies that refiners would own them.

Standard did receive "discounts" from the railroads; but this simply means that the railroads were unable to abstract quasirents from Standard, and were able to abstract them from (some) rivals. Rockefeller derived two cost advantages from his bargaining position with the

railroads: (1) lower shipping costs than his rivals, and (2) lower production costs that resulted from his ability to use large capacity plants.

Standard Oil eventually lost its monopoly position¹⁷⁵ as (1) demand exceeded the decreasing costs described above, and (2) new sources of supply made the sunk cost of refineries and pipelines irrelevant. The crude oil discovered in the Texas Gulf was not accessible to Standard's pipeline system, therefore the refinery/pipeline system would have to be duplicated - not expanded. Presumably Standard did not have any cost advantage in replicating itself; Standard had already exploited its economies of scale in ownership, and presumably moral hazard would have made any expansion increasingly expensive.¹⁷⁶

The development of rival integrated oil companies also changed the contracting problem. Once there were multiple competing companies, a refiner without its own pipeline system now faced a decreasing chance of holdup, since it could switch to rival pipelines (assuming pipeline systems are designed with capacity that exceeds their refineries).

This property rights explanation of Standard's business history has a number of advantages over the previous explanations. Unlike Chandler's economies of scale argument, both the

¹⁷⁵ I repeatedly use this term incorrectly. Standard was almost never the "sole seller". It did have very large market shares.

technological economies, and their relationship to firm size are demonstrated. Chandler seems to depend on economies of scale in production - but does not explain why Standard had multiple plants.

Although it is quite possible that Standard (or one of the local marketing companies partially owned by Standard) engaged in occasional predatory pricing, the conventional argument is an inadequate explanation. Since predatory pricing is a cost now in hopes of higher returns in the future, there must exist the expectation of industrial stability for the scheme to work. If the future industrial structure is highly uncertain, the discount rate is high - which makes the intertemporal tradeoff unappealing.

In fact Standard's market share (after 1877) was reasonably stable, with Standard retaining about 90 percent of the refining capacity and control over gathering lines. But the history of Standard is one of continuous conflict; the business environment never appeared stable.¹⁷⁷ If Standard's success was due to predatory pricing *without any accompanying cost advantage*, it was sheer luck. Standard bought most of its rivals (or traded stock), which created an

¹⁷⁶ By this point in the history of Standard oil, Rockefeller and his early associates had retired from active management of the company, and the ownership of Standard had become much more dispersed. This also increased the moral hazard cost of further expansion.

¹⁷⁷ Up until the Empire Rate War (1876-77), Standard controlled less of the transport of crude than the Pennsylvania/Empire combination. Immediately after its victory, Standard's control over transportation was threatened by the completion of a long distance pipeline built by independents. By the mid 1880's, Russian Oil threatened the existence of the

incentive for new entry - entrepreneurs built plants with the expectation of selling them to Standard. And Standard did price so as to minimize this cost some of the time. But this is not how Standard acquired its monopoly position.

The property rights explanation includes merger - but it is superior to the conventional merger argument because it explains how Standard maintained its market share. When there is a merger between competitors who were operating at the most efficient scale, the average cost of the new combination should be higher. This should attract new entry by firms of the "correct" size, leading to a gradual loss in market share of the merged firm.¹⁷⁸ Standard maintained and even expanded its capacity through merger, and when it eventually lost its monopoly position, its new rivals operated on a similar scale. This implies that Standard's purchase of its rivals decreased its costs - which is exactly the conclusion of the property rights argument. Merger alone is inadequate.

Finally, is the property right argument different from, and/or superior to the argument that Standard was merely an "evener" (enforcement agent of the railroad cartel)? It is certainly true that Standard was a party to collusive agreements which included a division of the market. It is certainly true that Standard received rebates, and various special concessions related to its ownership of gathering lines, tank cars, and shipping facilities. And it is true

export market (which exceeded the domestic market), and at the same time the Pennsylvania Oil fields started to become exhausted.

that Standard was a party to the Southern Improvement Corporation Railroad pool. But each of these facts are consistent with Standard's strong bargaining position resulting from its local monopolies.

What are the differences between these two theories? The two theories are not substantially different for any period after the early 1880's, since by that time the creation of long distance pipelines removed, with limited exception, that railroads' ability to extract refining quasirents. The key time period is from 1872, when both the Southern Improvement Company plan was attempted and Standard purchased the refining capacity in Cleveland, to 1880. This is also the time period in which Standard increased its market share in refineries from ten to ninety percent, and developed its extensive gathering network of pipelines.

The "evener" argument claims that the railroads effectively gave Standard its monopoly position through rebates. Standard was forced by the railroads to buy all of the rival refineries, so as to aid the enforcement of the collusion. Standard then should have expanded its output and market share almost entirely through the purchase of rivals, since its first goal was to eliminate rival buyers of transportation services. Standard should have continued to operate these existing refineries without changing their scale - since they were presumably already operating at the most efficient scale.

¹⁷⁸ This does not imply merger is not a profitable strategy.

As opposed to this, the property rights argument claims that Standard should have increased the scale of its plants, and decreased the number of them once it had secured its property rights in any given location. When it bought all of its rivals in Cleveland, it should have increased the scale of the largest plants, and dismantled the smaller (less efficient) ones. This is in fact what occurred. In 1872-73 Standard consolidated kerosene and naphtha refining in its half dozen largest plants, and dismantled the rest. Standard's remaining plants had the largest average capacity of any region in the country.¹⁷⁹ There existed increasing returns to scale in refinery plant operations, which Standard was able to exploit by its ownership of all of the refining capacity in Cleveland.¹⁸⁰

A final criticism of the "enforcement agent" argument is that Standard's market share was not sufficient to keep the Railroads from "cheating" on their rate agreements. The Pennsylvania Railroad in 1872-73 shipped 60 percent of the total petroleum traffic, most of which was supplied to and from independent refineries. It was not until after the Empire Rate War that Standard had control over 90% of refinery capacity.¹⁸¹ But by this time Standard's refinery market share was largely irrelevant, since equally effective control over the railroads could be maintained by Standard's virtual monopoly over feeder lines.

¹⁷⁹ Williamson & Daum, p. 293.

¹⁸⁰ Williamson & Daum, p. 366-367.

¹⁸¹ Williamson & Daum, p. 429.

Competition in the Political Arena.

In the process of Standard's rise most competing refiners either went out of business or were purchased by Standard. Many "independent" oil producers (independent of Standard and the associated railroads) competing with Standard found that they were unable to sell their product at a profit because of the low prices charged by Standard. Standard "won" most commercial competition, or at least made more money than its competitors.¹⁸²

But simply because rival firms were unable to compete on price does not mean that competition for the substantial wealth of the oil industry was eliminated. Rather competition shifted to margins where rivals believed they had an advantage. So while Standard and its rivals competed for wealth in the fields of refining and transportation, they also competed in both the courts and legislatures of various states.

The independent oil producers lobbied the various state legislatures for the passage of bills preventing the loss of their quasirents.¹⁸³ These bills usually threatened prosecution for "predatory" pricing or other forms of monopolization. While most states did in fact pass

¹⁸²Rockefeller's organization in 1870 controlled four percent of the U.S. refining capacity. After three months of buying Cleveland rivals, standard's market share rose to about 25%. By 1879 Standard Oil owned more than 90% of U.S. capacity. See Granitz p. 1.

antitrust statutes, Standard was generally able to protect itself through political contributions at the state level. For example independents repeatedly complained that Standard had “bought” the State Legislature of Pennsylvania, the state with the largest production of crude oil. In another example, Texas passed antitrust laws that were used to expel one Standard Oil company, but political contributions to legislators led to Texas allowing Standard to return.

Two important facts stand out in an analysis of state level political battles. The first is that they were never decisive. To this day the large oil companies (“Big Oil”) fight against organizations of independent oil producers/marketers for control of state legislatures.¹⁸⁴ The costs of protection from political rent seeking are never paid off. To own resources requires expenditures on protecting them.¹⁸⁵

The second is that while populist political support generally favors smaller competitors, the superior financial resources and more concentrated benefits of a successful firm like Standard make it unlikely that smaller firms will ever have decisive victories in the political arena.

¹⁸³Quasirents in the case of crude producers. Competing refiners who exited the market lost both quasirents and rents.

¹⁸⁴Atlantic Richfield Co. contributes the second largest amount to candidates for Washington State’s legislature; the largest is the state teachers union.

¹⁸⁵See Barzel 1994 for a more complete analysis of this point.

The other area of confrontation in which firms defend or appropriate rents is the court system. After repeated failure in the Pennsylvania legislature, the officers of Standard's opponents suggested state level antitrust action, either privately or through lobbying the appropriate government prosecutor.

There were many legal battles (and still are) over the wealth of the petroleum market. All of the early ones were in State courts, since there were few Federal "antitrust" laws. But at the state level there existed laws against "preventing competition" or "limiting trade", and cases against Standard were constantly in progress. They seem to have followed this common pattern:

1. With much hoopla, a private plaintiff or attorney general would claim that Standard was large, nasty, and in violation of the law. A complaint would be brought in a State Court, usually claiming that Standard restrained trade.
2. A year or two later, the case would go to trial. Standard would lose.
3. After many more delays, a usually impossibly difficult remedy (punishment) would be imposed. The politician would declare victory. Or the producer organization (private party) would celebrate and disband. Standard would appeal.
4. Several years later, and many years after the initial complaint, the case would go away. Either the state level politician would drop the appeal, the appeals court would overturn it, or the private plaintiff would go broke. In the host of antitrust cases, including perhaps the final

supreme court case that “broke up” Standard Oil of New Jersey, nothing relevant or timely ever happened.¹⁸⁶

All of these Nineteenth Century cases against Standard were filed in state level courts, although a number eventually reached the Federal Courts on appeal. After the passage of the Sherman Act in 1890, Standard continued to avoid Federal prosecution for the next sixteen years. But eventually Standard was prosecuted and convicted, and appeals lasted until May of 1911, when the Supreme Court found the transformation of Standard into a holding company in 1899 violated Sections One and Two of the Sherman Act.

Standard Oil of New Jersey was found guilty of violating the Sherman act, primarily based on driving its rivals out of the marketplace through “predatory” pricing and secret rebates. Since these actions lead to Standard’s status as sole supplier, and since Standard was unable to justify them based on efficiency considerations, the courts ordered the breakup of the Standard Oil trust.

Dissolution

The Court faced several difficulties in imposing a remedy to the Standard case. By the time the case was settled, the original parties had retired from the business, and much of the

¹⁸⁶This is the main conclusion of Bringhurst.

common stock had been divided or sold off to others. In addition, Standard was both facing new competition from Gulf coast refiners and was a successful business organization which provided valuable services in refining and distributing petroleum.

The Court's "solution" to the problem of a monopoly supplier of oil refining and transportation was to dissolve the company into thirty three independent companies. However since a refinery is a difficult asset to divide into parts, and since Standard had already distributed its refineries so as to minimize transportation costs, the court effectively created many regional monopolies.¹⁸⁷ It was many years before these descendants of Standard became the aggressive competitors they are today.¹⁸⁸

¹⁸⁷ Much like the break up of AT&T.

¹⁸⁸ Standard's descendants include Exxon, Mobil Oil, Standard Oil of California (Chevron), and Standard Oil of Indiana (Amoco). Atlantic Richfield (Arco), Continental Oil (Conoco), and of course Standard Oil of Ohio (Sohio) are also listed as having their origins in Standard. See Bringhurst, pp. 1-2.

CHAPTER 5: THE UNITED STATES STEEL CORPORATION

The United States Steel Corporation, with a capitalization of 1.4 billion dollars, was incorporated in 1901. It exceeded the size of the largest existing corporations, the Tobacco Trust and Standard Oil, and in doing so guaranteed itself a prominent position in the history of Antitrust.

The corporation resulted from the merger of nine large steel companies, most of which were formed from previous consolidations. The eighteen nineties was a period of large scale consolidation in many industries, including steel, and the US Steel corporation was the largest of these mergers.

The existing economics literature has focused on the way in which the merger changed the structure of the steel industry; the last decade of the nineteenth century witnessed large price fluctuations and failed attempts at collusion, while the first half of this century was a period of industry stability, steady prices, and few new entrants into the market. Economists, looking back over this century's cooperative pricing, have explained US Steel's formation as the event that brought about this change.

This method, explaining an event by means of its long term consequences, is often quite distorting. It implies that the merger's organizers were able to forecast changing supply and conditions decades into the future. In what follows, I argue that the prevailing explanation -- essentially that US Steel was formed to act as a dominant firm in Steel -- is incorrect. Vertical considerations better explain US Steel's formation and initial behavior. The merger occurred in order to delay entry into previously monopolized downstream industries.

Since my argument depends on various changes in production costs, technologies, and ownership patterns, let me begin with an extended section on industry development. In section two I will summarize existing explanations for the formation of US Steel, followed by arguments suggesting why my explanation is better.

I. Industry Background

The Technology to the civil war

Wood was the most common industrial material from the colonial period through the middle of the last century. The ships that the colonists arrived in were made of wood. Houses, barns, carriages, bridges and even the machines in weaving factories were all made of wood.

Iron was used in kitchen utensils, nails and other fittings,¹⁸⁹ and steel was reserved for cutting edges (saws, knives, etc.).

While steel had been used for a considerable portion of history,¹⁹⁰ it is iron that was the much more common metal. Steel is more useful, for it is relatively tough, hard and flexible, while iron is softer and more malleable. Steel was however much more difficult to make (see below).

Iron, of which the steel industry was a tiny segment, has been made in North America since colonial times. The process until the mid 1800's was unchanged. Iron ore, charcoal and limestone (a flux to remove impurities) were piled into a blast furnace¹⁹¹ and ignited. Cold air was forced into the furnace, which caused the charcoal to burn at a high enough temperature to melt the iron from the ore (smelting). The iron that formed at the bottom of the blast furnace was called pig iron. This material was too impure -- non homogeneous in properties -- for direct use.

¹⁸⁹ Think of the wagon the pioneers crossed the great plains in. It had wheels made of wooden spokes and rim, with an iron border or rim.

¹⁹⁰ Witness the famous steel swords of Spain, which were used to conquer the new world.

¹⁹¹ The first blast furnace in what would become the United States was built in 1645 in Massachusetts. Hogan p 1.).

In order to remove the many impurities, the pig iron was heated again, and then boiled, stirred, crushed, squeezed and rolled. This process was called *puddling*, and resulted in wrought iron, which could then be rolled or pounded into various shapes (plates, bars, etc.). Charcoal, the primary fuel for these processes, was made from local forests, and iron ore also tended to be locally supplied.

Steel made in the US, what little there was of it until the 1870's, was made by taking iron and adding carbon by a lengthy process of heating it in charcoal. This process created *blister steel* (called so because of the non-homogeneous blisters on the steel), which was inferior to the *crucible steel* imported from England.

For Western civilization, this was the status of material science from the end of the bronze age to the American Civil war; wood and stone were used for most construction, fittings and larger metal tools were made of iron, and a few weapons and cutting tools were made of steel.

How the dramatic increase in demand, brought about by the railroad, changed the industry.

What caused the first change in the iron industry -- really the formation of the modern iron and steel industry -- was the railroads. While the canals did not create any particular demand for iron, both rails and locomotives (and later bridges, rail cars and terminals) were

constructed of iron. As the total mileage of railroad track tripled in the 1860's,¹⁹² the small scale iron were entirely unable to supply this quantity of metal. While imported iron rails from England built most of the early track, given the increased demand and import duties,¹⁹³ investment in expanded domestic facilities proved profitable. Companies with such famous names as the Carnegie Steel Co., Bethlehem Steel Co., Wheeling Steel Co., and Jones & Laughlin all had their beginnings as iron producers in the 1850's and 60's. Thus we have the beginning of the modern iron industry.

The large quantity of iron demanded by the railroads was *not* made by a simple increase in the number of iron plants of the type that had been in common use since colonization. Two changes occurred. First, the new furnaces were of a substantially different design and capacity. Second, changes in the location of the raw materials for iron production eliminated the value of existing furnaces.

A typical iron production facility (a blast furnace) in 1860 produced perhaps ten tons per day of pig iron. It used local ore and fuel (charcoal or anthracite coal), and was constructed entirely of stone. The new furnaces constructed throughout the rest of the century had iron shells lined with firebrick. They were of substantially larger capacity (and increased in capacity as the century progressed), and used newly designed equipment that allowed

¹⁹² Hogan p. 111.

stronger and hotter blasts of air, which in turn allowed increased temperatures within the furnace.¹⁹⁴

These new types of furnace had a capacity of around 550 tons per week, and by 1878 one furnace produced 800 tons in one week. While even these furnaces were initially loaded with coal, ore and flux by hand, by the end of the century all phases of production were mechanized.¹⁹⁵ As the century came to a close the absolute number of blast furnaces actually declined (even though production doubled and doubled again); obviously the capacity of the largest furnaces continued to increase.

Upstream Changes

The increasing use of iron resulted in the exhaustion of most local ore by the 1880's,¹⁹⁶ and the general increase in population led to the decline of forests from which charcoal was made. The new furnaces built to supply iron tended to be built near large coal fields (much of it the bituminous coal or coke located in Western Pennsylvania) and used imported ore from the upper shores of Lake Superior. Neither the fuel nor ore was competitive with local

¹⁹³ For example the Morrill Tariff of 1861 placed a \$9 per ton duty on pig iron, and \$12 per ton duty on iron rails. Hogan p. 174.

¹⁹⁴ Hogan p 28.

¹⁹⁵ Hogan p. 216.

¹⁹⁶ Hogan P. 20

supplies when they existed, both because of the high transportation costs and special (and yet unknown) processing required for each.¹⁹⁷

The Lake Superior ore, discovered in the 1840's,¹⁹⁸ was initially very expensive to bring to market. It had to be transported from the mines to local docks, where it was stored until the opening of shipping season (the lake is not navigable in winter). It was then loaded by hand on wooden sailing ships. At Sault Ste. Marie it had to be removed from the ships, hauled overland one mile to Lake Huron, and then manually loaded onto other ships for the trip to Chicago or other lake ports.¹⁹⁹ Given that manual loading meant wheelbarrows and shovels, this cost was quite high.

The rest of the century saw a decrease in shipping costs as the result of specialized investment in transportation facilities. The Sault Ste. Marie canal was installed, allowing the entire water portion of the transit to take place in the same barge or ship. The canal was repeatedly enlarged to allow passage of larger and larger ships.

¹⁹⁷ The bituminous coal of western Pennsylvania, while large in quantity, was too impure for direct use. It had to be converted to coke by a process of heating. The first ores taken from Michigan were very suitable for iron and steel production, but the Mesabi ore used in the 1890's was dirt like, and required changes in furnace design. Hogan p. 20, 196.

¹⁹⁸ Hogan p. 18.

¹⁹⁹ Hogan p. 18.

The ships themselves went from generic freight carrying sailing ships to specialized steamers and barges designed for inexpensive loading and unloading by (eventually) mechanized means.²⁰⁰ With the development of the Mesabi Range in Minnesota the process became almost entirely mechanized. The ore was relatively soft, more of the composition of dirt than the previously mined rock. This ore was strip mined and extracted by steam shovel. Specialized railroad cars transported it to specialized docks, which could mechanically load it on to ore ships for transfer.²⁰¹ While the United States iron industry initially operated at a cost disadvantage compared to Britain, by the end of the century improvements in transportation infrastructure and processing plants allowed some international competition.

Changes in the use of iron and steel.

As described in previous chapters, decreased transportation costs resulted in much larger markets and more specialized production. Demand for iron not only increased in the transportation industry, but also from the industries to which the railroads provided transportation. In the area of machine tools, the Lincoln Mill (a plain mill) was invented in 1855. The turret lathe, first used by the Colt Firearms Co., was invented in 1856.²⁰² This

²⁰⁰ Hogan, p. 22.

²⁰¹ Hogan p. 198.

²⁰² Hogan p. 125.

same period saw the creation of famous machine tool companies such of Pratt & Whitney, and Brown & Sharp.

The use of pipelines in the petroleum industry (described in my chapter on Standard Oil) also sparked the demand for iron. Conversely, the introduction of petroleum lubricants made the use of high speed machine tools much more practicable.

Between the blast furnaces that made iron ingots and these final users of iron products, there existed another segment of the iron industry: fabricators of semi-finished products. These included tin plate (iron with a thin tin coating used in cans), railroad rails, structural pieces, wire, plate to be welded into pipe, etc. The production of almost all of these products involved squeezing the iron ingots between rollers with appropriate cutouts that gave the product its shape. The development of this industry followed the standard pattern: mechanization and specialization together resulted in plants that used more capital and less labor to produce at lower cost.

Following the rapid expansion of iron production in the 1860's and 70's, steel production also skyrocketed. Again it was the railroads that provided the initial demand. The iron rails in use in the 1860's had a life span of six months to a year on the most heavily used sections of the track. Newly developed English steel rails had an average life span of approximately

ten years under the same use.²⁰³ Once this was discovered, the railroads made a rapid switch to steel rails. Iron rail domestic output increased from 205,038 tons in 1860 to 905,930 tons in 1872 (its peak), and declined to 493,763 tons in 1880. Steel rail domestic output increased from 2,550 tons in 1867 (the first year of data) to 968,075 tons in 1880.²⁰⁴

Changes in Steelmaking Technology

Of course this quantity and quality of steel rail was not the product of the existing tiny blister steel industry. The increased demand for steel, combined with the technical expertise developed from large scale iron production, made it economically feasible to implement and refine a process developed in the 1840's and 50's named Bessemer steel making. This process consisted of forcing air through molten pig iron. The air mixed with and burned the carbon in the pig iron, removing both the carbon and impurities. Iron, carbon and manganese were then added to the decarbonized pig iron, in the proper amounts, and the resulting product was a good quality steel.²⁰⁵ The technique was the combination of several patented

²⁰³ Hogan p. 118.

²⁰⁴ Curiously, the *purchasers* of this steel rail lobbied congress for an increase in the duty on imported steel rails.

²⁰⁵ Hogan p. 33. I do not understand this process with any level of precision. Carbon is first removed from pig iron. Then carbon is added back in. Why? Hogan seems to imply that pig iron had the wrong amount of carbon, as well as various impurities. The decarbonization process removed all of the carbon and impurities, but the resulting metal was useless. It was only by adding back a limited amount of carbon that steel was created.

processes (one by the Englishman Bessemer), and in 1866 the patents of three inventors were pooled to form the Pneumatic Steel Association.²⁰⁶

An early steel plant consisted of the following equipment and processes. First, pig iron (smelted in a furnace at another plant) was melted into molten form. It was then poured into a sphere shaped container called a Bessemer converter. This container was heated, and hot air was forced into the molten metal, resulting in the conversion described above. Once the iron was converted to steel, the sphere was rotated, allowing the molten steel to pour out into a ladle, which then poured the steel into ingot molds.²⁰⁷

The discovery that it was *possible* to make steel by this process was not equivalent to the successful implementation of the process. The technique required considerable experimentation and innovation in process and equipment. The metal was both boiled and poured, processes involving tremendous amounts of heat. Molten metal is very difficult to contain, for it is both very heavy and hot enough to either melt or burn almost all containers. The bottoms of the Bessemer converters, to which heat was applied, would last only for a few batches. When they gave out, liquid steel would pour onto the floor, and if it

²⁰⁶ Hogan p. 34

²⁰⁷ Hogan p. 219.

encountered moisture, explode.²⁰⁸ The early pioneers in steel making were true risk takers, both financial and physical.

By the end of the century, Steel plants took on a fairly standard form. The invention of a *Hot Metal Mixer* allowed pig iron from a blast furnace to be stored in molten form. By grouping blast furnaces, hot metal mixers, and Bessemer converters together at the same facility, the process by which pig iron was cast into ingots, then remelted for the Bessemer converter was eliminated. The Bessemer converters could be loaded with liquid pig iron that had been melted in the blast furnace. The converters were grouped in pairs (two converters per plant), and while one was being loaded or unloaded, the other would undergo the actual conversion (blowing the steel).

The size of the converters resulted in plant level economies (and diseconomies) of scale. Paired converters were optimum, because the amount of time required to blow the steel in one converter was also the required time for loading, unloading, and repairing the other. In this way, expensive peripheral equipment, such as the blowing engine and ladles, could be kept in continuous operation. Obviously the blast furnace and hot metal mixer then also had an optimum scale -- one that just supplied the correct amount of pig iron to the converters. What changed over the century was not the number of converters per plant, but instead the size and speed of operation of the converters. When first popularized in the early 1870's

²⁰⁸ See for example the extended quotation in Hogan p. 34.

most converters were between six and seven tons; by the end of the century the normal converter was closer to 10 tons.

The greatest increase in plant capacity (something like a tenfold increase)²⁰⁹ was not converter size however, but an increase in the number of “heats” (batches) per day. The size of the converters was in fact relatively small (see below) as compared to the cost of the equipment. The only way in which this fixed cost could be recovered was through streamlining production to a quick and continuous process.

A steelmaking plant was therefore a busy place. The blast furnace was in continuous use, with constant top loading of ore and coal. The hot metal mixer ran back and forth on tracks between the furnace and Bessemer converters, loading and unloading liquid iron. Teams of men alternated unloading and reloading one converter, while the other boiled iron into steel. It was only by perfectly coordinating these processes for continuous operation of the equipment that US steel companies could produce at a lower cost than their English rivals (who had lower raw material and labor costs).²¹⁰

²⁰⁹ Hogan p. 220.

²¹⁰ It was only at the end of the century that the tariff on steel and rails became irrelevant in preventing British imports. Up until this time, the British cost advantage even exceeded the transatlantic shipping costs, and often exceeded the substantial tariff as well.

Bessemer plants created most of the steel produced in the US during the 19th century. But by the end of the century the *Open Hearth Process* began to supplement Bessemer steel, and by 1908 became the dominant method of steel production.²¹¹

Whereas a Bessemer converter worked by bubbling heated air through a relatively small container of molten pig iron in a short time period, an open hearth furnace used a heated mixture of flammable gas and air to heat a very large container for an extended time. The operator of a Bessemer converter had to estimate when the steel was converted by the color of the flame extruding from the container. An open hearth furnace allowed samples to be removed and tested, and the process could be adjusted until each batch was complete. Each process removed impurities and carbon from the iron by oxidation; open hearth furnaces performed this process over a longer time period, allowing slightly different chemical reactions to take place.

The advantages of the Open Hearth process, which resulted in its ascension over the Bessemer process, are as follows:

1. Ore and Phosphorus. High quality steel must have a very low phosphorus content. While some iron ore contained low enough concentrations of phosphorus, higher phosphorus ore was more common. The Bessemer process was unable to eliminate

²¹¹ Hogan, p. 413.

phosphorus from pig iron, and thus was restricted to a reduced supply of ore. The slower open hearth process did eliminate phosphorus from steel, and was therefore able to use lower cost ore.

2. Scrap. Open hearth furnaces had removable roofs. As a result of this, large pieces of scrap steel (rails, machinery, etc.) could be loaded into the furnace and melted before the addition of liquid pig iron. This resulted in a considerable savings in ore.
3. Homogeneity in output. Batches were relatively large, making sampling feasible. And since the process could be adjusted based on these samples, the output was not dependent on variations in the inputs. This resulted in more homogenous steel.²¹²

Summary of Changes in Technology that resulted in Differing Plant Level Economies

Changing technology, and the resulting changes in cost, led to different tasks being performed at steel making plants. These are economies of scope, and to some extent scale, and their relationship to economies of scale in ownership will be explored below.

Iron blast furnaces were originally located in their own plants, and the resulting ingots were shipped to other plants for rolling into semi-finished forms (see above). I have already discussed how the invention of hot metal mixers gave Bessemer steel plants with attached furnaces a cost advantage over those that purchased iron ingots.

²¹² Hogan, p. 403.

Coke, the major fuel for iron and steel making, is formed by heating coal. This process gives off various byproducts, and the early beehive ovens (named so because of their shape) allowed these gaseous byproducts to evaporate into the atmosphere. These ovens tended to be located at coal mines, for coking is a reducing process. It was cheaper to ship coke instead of the more voluminous coal. The ovens were also relatively inexpensive to construct, so once a given mine depleted its coal supply, the oven could be abandoned with the mine.

The byproduct oven, which slowly replaced the coke oven, was located with the furnace. As the name implies, a byproduct oven captured the gases given off in coking, which were used as fuel for open hearth furnaces (and other uses). These ovens were most economically located at the steel plant, since this would save the transport of gas to the steel plant, and also allow steel makers to mix coal from different mines to form coke appropriate to a specific furnace. These coke ovens were many times more expensive than beehive ovens, and therefore were not abandoned with a mine.²¹³

Another technological improvement that increased the scale of plants was air conditioning. The amount of moisture in air changes with the season (summer air is more humid), and

²¹³ Hogan, p. 378. One of the major changes that brought about the switch to byproduct ovens was the increasing demand for other elements (besides the fuel) given off by coal.

since a tremendous amount of air is blasted into a furnace, this also means that large amounts of water are also added when the air is moist. Blast furnaces could be run more efficiently (less fuel and smoother operation) if the air was first dried by refrigeration, but since the facilities for this process cost approximately \$400,000 they were unlikely to be used in “small” plants (creating an economy of scale).²¹⁴

Changes in Ownership Patterns.

The businesses of the 40’s and 50’s were all small partnerships, the most successful of which evolved into closely held corporations in the 60’s and 70’s. Bonds were a major source of funding throughout the century, although many firms expanded their capital investments (which became increasingly large) through retained earnings.²¹⁵ While all of these partnerships started in one segment of the industry (e.g., rail plants), a combination of internal expansion and merger resulted in firms that were increasing in both size (horizontally) and scope (vertically).

While the iron business of the 1860’s was dominated by partnerships possessing one plant, by the end of the century the more typical (as measured by output)²¹⁶ firm was much more

²¹⁴ Hogan p. 400.

²¹⁵ See Hogan, p. 100 and Meade 1901 for various examples.

²¹⁶ At the time of the US Steel merger, most steel companies (of which there were hundreds) were non-integrated; they purchased steel for fabrication into products. As indicated below

vertically integrated, much larger (employing tens of thousands of workers and multiple plants), and owned by a larger number of shareholders (many of whom were originally the partners in merged or purchased firms). These trends may be illustrated by the most successful and famous pre-US Steel company, the Carnegie Steel Company.

The Carnegie Steel Company began as a partnership of three businessmen in 1861. The partners found themselves in conflict almost immediately, and the business was reorganized in 1863 with the substitution of one partner. This partnership also was not amenable, and in resolving the partner's dissension, Andrew and Thomas Carnegie entered the business with a financial interest. The firm, located near Pittsburgh and known originally as the Iron City Forge, had a mill for creating rolled iron products.²¹⁷

In 1865 the firm merged with another rolling mill, in which Andrew Carnegie also had an interest, and formed the Union Iron Mill. Iron City Forge had been profitable throughout the early 60's because of large orders related to the civil war. By 1870 the new firm had built the first universal mill (which could both reduce the thickness of a plate and roll the edges) west of the Allegheny mountains. In 1870 Carnegie and several other partners formed another business to build a modern blast furnace (which established output records almost immediately upon its completion) to supply Union Iron with pig iron. In 1873 Carnegie

by US Steel's market share, most of the heavy steel output was produced by large integrated companies.

formed another partnership, which was dissolved in 1874 in order to form a limited liability corporation.

This corporation was named the Edgar Thomson Steel Company, named after the president of the Pennsylvania Railroad, who purchased \$100,000 in bonds issued by the corporation. It built a Bessemer steel plant and a rail mill. In 1882 Carnegie exchanged a six percent interest in his steel company for a 29½% interest in H.C. Frick & Company, a leading supplier of Connellsville coke, and in 1883 purchased control over the strife-torn Pittsburgh Bessemer Steel Company, a firm that had originally been formed as an alternative supplier of steel to customers of the Edgar Thompson Steel Company. A third Steel rail plant, the Duquesne mill, was purchased in 1890 when it was beset with labor problems.

In 1892 Carnegie reorganized his various holdings as the Carnegie Steel Company Limited, with a capitalization of 25 million dollars, over thirteen of which was owned by Carnegie. H.C. Frick, with under three million dollars in stock, was the operating head of the company. Under Frick's management the company integrated upstream, purchasing Mesabi ore land, leasing transportation on Rockefeller's railroad and steamship lines, and purchasing a railroad to transfer the ore from the lakes to the steel plants.

²¹⁷ This section is drawn from Hogan p. 98-100 and 243-254.

Carnegie's company was considered the most efficient in the industry. The firm was constantly implementing new technology, building larger and more automated plants. Expansion came about through a combination of increasing plant size and purchase of poorly run rival firms with modern plants. Carnegie was also known as a cutthroat competitor, undercutting the prices arranged through pools and gentlemen's agreements.

By 1901, when it merged with the other components of US Steel, the Carnegie Company was not only the most efficient, but also the largest firm in the steel industry. All of its manufacturing properties were concentrated in the Pittsburgh area. It was integrated into upstream transportation, and coal and ore mines. It produced 18% of the US ingot output, and was concentrated in "heavy" steel products, which were purchased by downstream producers of semi-finished steel goods.

Carnegie's firm illustrates many of the ownership changes that transpired in the steel industry. It began as a partnership owning a single plant, and expanded both by purchase of rivals and internal expansion. As demand for its product expanded, more specialized resources were developed (particularly in transportation of inputs), which resulted in vertical integration. Carnegie used the sale of corporate shares to finance expansion upstream into railroads, ore carriers, and coal mines. The industry had made a transition from partners owning plants, to shareholders owning integrated firms.

The Formation of US Steel

On February 25, 1901 a charter was obtained for the US Steel Company. It included the following companies:²¹⁸

Carnegie Steel Company

National Steel Company

National Tube Company

American Steel and Wire Company

American Sheet Steel Company

American Hoop Steel Company

American Tinsplate Company

Federal Steel Company.

American Bridge Company.

In 1902 it had 168 thousand workers, \$561 million in sales, \$140 million in accounting profits, and was capitalized at \$1.4 billion. This was approximately 6.8% of the GNP in 1901.²¹⁹ To place this in perspective, Microsoft's market capitalization (the value of its shares) in 1997 was approximately \$ 120 billion, which was 1.5% of nominal GDP.

²¹⁸ Hogan p. 471.

²¹⁹ McCraw & Reinhardt Table 2.

Over time, its percentage of total output changed as follows:²²⁰

Table 1: US Steel Market Share Over Time

	1901	1911	1919	1927
Iron Ore ²²¹	45.1%	45.8	42.1	41.4
Blast-Furnace Products	43.2	45.4	44	37.7
Steel Ingots and Castings	65.7	53.9	49.6	41.1
Steel Rails	59.8	56.1	62	53.3
Heavy Structural Shapes	62.2	47	43.8	38.8
Plates and Sheets	64.6	45.7	44.3	36.5
Wire Rods	77.6	64.7	55.4	47.4
Wire Nails	65.8	51.4	51.9	42
Tin and Terne Plate ²²²	73	60.7	48.4	40.5

²²⁰ McCraw & Reinhardt Table 1.

²²¹ US Steel Corporation did not sell any ore to other firms until 1940. Hines, p. 651.

²²² Tin Plate was used in making cans. Terne Plate, made by finishing works known as dipperies, was a plate coated with a mixture of tin and lead used in roofing. See Lamoreaux, p. 19.

It should be noted that while US Steel's market share decreases over time for all of the products listed, its absolute production increases in all categories, since total US production increased throughout the period. For example its ingot capacity was 9.4 gross million tons in 1901 and 23.2 gross million tons in 1927.²²³

Furthermore, US Steel's capacity increased smoothly over time (as capacity tends to do when individual plants are not too large a percentage of the total). While capacity did decrease during a few years (e.g. 1912 = 18.8 million tons, 1913 = 18.5 million tons), the trend for the first thirty years of its existence is almost monotonically increasing.²²⁴

Profitability is another matter. The steel industry is very pro-cyclical. In times of expansion, for example the first world war, profits were very large -- almost four times their pre-war levels. During the great depression US Steel lost money.²²⁵ For the period 1901 to 1930 US Steel had an average profitability (as a percentage of gross fixed assets) of 12.6%, while the six next largest producers had profits of 10.3%, 10.1%, 8.2%, 16.3%, 16.9%, and 13.7%.²²⁶ US Steel was a profitable company, at a level comparable to its leading rivals.²²⁷

²²³ McCraw & Reinhardt Table 2.

²²⁴ McCraw & Reinhardt, Table 2.

²²⁵ McCraw & Reinhardt, Table 2.

²²⁶ McCraw & Reinhardt, Table 3.

²²⁷ See also George Stigler, "The Dominant Firm and the Inverted Umbrella," *Journal of Law and Economics* 8 (October 1965), pp. 157-71 for questions concerning relative profitability.

Industry Price and Industry Quantity Over Time

Industry output was less stable after the formation of US Steel. Between 1887 and 1901 the average of yearly deviation from a five year moving average was 7.7% for Pig Iron, 10.5% for Steel ingots and castings, and 8.5% for rolled iron and steel. These same figures for the 1902 to 1922 period are 13.5%, 14.7%, and 14.9%.²²⁸ If the merger had any effect on industrial stability, it was to make the quantity produced less stable.

Industry price was another matter entirely. A comparison between the periods of 1898 to 1901 and 1902 to 1914 shows the greatest variation from the average price was 31.3% for the early period and only 19.3% for the latter period for finished steel products. Prices were generally more stable after the merger than before it.²²⁹

Legal History

Following the 1906 antitrust cases against Standard Oil and Tobacco Trusts, the Justice Department filed an antitrust suit against US Steel in 1911. The Supreme Court completed its final ruling by 1920, and unlike the previous cases, the Court ruled in favor of US Steel.

²²⁸ Berglund 1924, p. 614-15.

²²⁹ Berglund 1923, p. 14. See also McCraw & Reinhardt, p. 602.

Both the oil and tobacco cases involved complaints on the part of rival producers (so called predatory behavior). But the creation of US Steel involved the elimination of the most competitive producer (Carnegie, who retired), and the introduction of an almost passive giant. Thus rival producers were not willing to press for a dissolution.

The court did find evidence of attempted price fixing -- a series of dinners hosted by Judge Gary, the Chairman of US Steel -- but these had occurred in the past and had been voluntarily discontinued. The generally advancing technology led to lower prices of steel goods, leaving few customers willing to bring suit. Since there was no motion to dissolve, and no current business behavior to discipline, the Court ruled in favor of U.S. Steel, including the statement, "the law does not make mere size an offense, or the existence of unexerted power and offense. It ... requires overt acts."²³⁰

II. Economics and Antitrust Questions

Existing Explanations

The economics literature has provided a number of explanations for the merger that led to the birth of US Steel. George Stigler, for example, cites US Steel in his exposition on the

²³⁰ This hardly ended US Steel's legal troubles; an FTC investigation began the next year.

Dominant Firm Model.²³¹ A number of features of this model dovetail nicely with the facts of the case, including market share, pricing, and exports.

Since the residual demand curve is more elastic in the long run than in the short run, a dominant firm's market share should decline over time. As reported in table 1 above, US Steel's market share went from approximately 65% at its formation to 45% by the late 1920's. Declining market share over time is consistent with the model.

The residual demand curve is constructed on the premise that, for any given price, the fringe will supply a certain quantity. The rest of the quantity demanded will be provided by the dominant firm, making it a price searcher. For this reason the dominant firm may ignore the pricing actions of the fringe.

For much of the period in question, US Steel announced its price at the beginning of every year for various products, and then maintained those prices even when being undercut by rivals. This is exactly the sort of pricing predicted by the model (although I will return to this topic).

²³¹ Since Stigler specialized in homogenous goods industries, it is not surprising that he often cites US Steel. See "Monopoly and Oligopoly by Merger." One article, "The Dominant Firm and the Inverted Umbrella" would appear, from the title, to primarily concern itself with the dominant firm model. While the subject of the article is the US Steel merger, it is not an application of the D.F. model to this case. Instead he is rejecting another explanation; that US Steel was formed to "bilk" untutored shareholders through watered stock. See his initial credits for the type of statement feared by all graduate students.

Finally, since the dominant firm is acting as a price searcher, it is setting a quantity where marginal revenue equals marginal cost. The price taking fringe expands output until price equals marginal cost. Since for the dominant firm marginal revenue is below price, this implies the dominant firm will have a lower marginal cost.

US Steel was the only firm able (or willing) to export steel at the lower world market price. US tariffs prevented profitable imports of foreign steel, allowing a divergence between the domestic and international price. Other US producers claimed to be unable to profitably sell steel on the world market -- which is entirely correct when they could act as price takers in the higher priced domestic market. Only US Steel would be operating at a marginal cost that would allow profitable export.²³²

That application of the dominant firm model to US Steel is somewhat more complicated than Stigler presents, however, and that complication removes some of the force of his argument.

For example Stigler is illustrating a situation in which demand is static, and capital investments may, at best, be withdrawn only slowly. With the onset of the merger, new capital stock enters the industry. Once sufficient entry occurs for a return to competitive

²³² Actually US Steel had between 74.07 and 107.42 percent of total iron and steel exports between 1902 and 1911, with an annual average of 83 percent.

pricing, the industry is faced with losses due to excessive capital investment. This follows from the fact that the pre-merger industry was operating at zero profit, and any new entry must force a competitive industry into an equilibrium in which the marginal firm is suffering losses.

Stigler notes that merger may still be profitable in present value, since present profits may offset future losses. He further states that if either specialized resources are not indestructible or demand increases, future losses will be reduced or eliminated. What he does not analyze is the case of US Steel, whose ingot capacity increased from 9.4 million gross tons in 1901 to 27.8 million tons in 1932.

Why would a dominant firm expand production and capacity? The answer, of course, is that demand increased throughout the period, so that US Steel was able to increase production three fold, yet lose half of its market share. This essentially dynamic problem is more complicated than the of straight forward application of marginal revenue / marginal cost analysis.

McCraw and Reinhardt (1989) argue that from its formation to 1907, US Steel did not act as a passive dominant firm -- at least in steel ingots. Compared to the rest of the industry, it added relatively large amounts of capacity, and operated its plants much closer to their capacity limits. What changed, in 1906, was the filing of antitrust suits against American Tobacco and Standard Oil. McCraw and Reinhardt claim that US Steel's Chairman, Judge

Gary, perceived that a passive strategy would allow the merger to survive judicial review -- and that he was right.

US Steel pursued a strategy of price stability and high capacity utilization, which resulted in declining market shares, declining share of capacity, large swings in output, but also large profits. George Stigler also makes this argument, claiming that US Steel was attempting to form a monopoly before 1906, but was content with an oligopoly after 1906.

Parsons and Ray (1975) add to Stigler's analysis, again by arguing that the merger changed the structure of the industry in a manner that aided collusion. They point out, for example, that US Steel did not bear a disproportionate share of the decline in output during periods of recession -- and that in the sharp recession of 1921 US Steel's market share actually rose. They prefer to characterize the industry as having a "dominant cartel".²³³ In 1903 the four largest steel makers owned 70.19% of industry capacity (54.5% for US Steel). The largest eight had approximately 80% of capacity, and the largest twenty had almost 90%. As witnessed by the Gary Dinners, these firms were willing to meet and discuss methods of eliminating "destructive" competition.

The source of US Steel's market power -- in this case its ability to maintain cartel pricing -- was ownership over iron ore. The other major raw input, coal, was used in many other

²³³ Parsons and Ray, p. 206-208.

industries, leaving US Steel and the cartel in a poor position to control a significant share of production. But iron ore was only useful to the iron and steel industry. If the “dominant cartel” could control access to ore, they could delay or prevent new entry.

US Steel, at its formation, owned ore holdings of 700 million tons (44 times its 1902 production rate), and increased its holdings in 1902 and 1904. In 1907, it entered into a long term lease of the Great Northern Railroad’s ore holdings, and purchase the Tennessee Coal, Iron and Railroad Company, tripling its initial holdings.²³⁴

While the size of US Steel’s ore holdings was impressive, it was not directly relevant in preventing entry -- that required limiting other firms from acquiring ore. US Steel’s market share of ore production (see table 1) is easy to measure and report, but it also is not the correct number. What is important was US Steel’s ownership, perhaps on a percentage basis, of the ability to produce ore -- of ore deposits. That number is much more difficult to estimate precisely.

One reason for this is that the quantity of potential deposits depends on price of ore. Iron ore is spread throughout the earth’s surface, in varying concentrations. As the price of iron rises, lower concentration ore becomes economical to mine. Other complications include the possibility of new discovery of ore deposits, or technological innovation that decreases the

²³⁴ Parsons and Ray, p. 199

cost of using existing ore. The discovery of large deposits in Cuba, which were believed to be of size comparable to the Mesabi range, created the possibility of a large increase in supply. Likewise, the shift from the Bessemer to the Open Hearth method of steel production allowed the use of the much more common high phosphorus ore.

Because of these complications, it was, and is, very difficult to measure US Steel's control over ore deposits. Ray and Parsons report that the Commissioner of Corporations estimated US Steel's ownership at 75% of the Lake Superior district, a figure they consider an upper bound. A US geological survey estimated US Steel's holdings at 50% of those commercially viable. Ray and Parsons also note that US Steel owned or leased all three of the railroads dedicated to removing ore from the Lake district.

As well as ownership of ore, the cartel maintained its pricing by the purchase of new entrants. While US Steel did not make major purchases (other than those noted), the other large steel producers (e.g. Bethlehem, Youngstown, etc.) increased their capacity and market share through both internal expansion and purchase of rivals. Weston reports that 40 percent of additions to ingot capacity by the major producers were through acquisition. These mergers allowed the leading producers to maintain their combined market share, even though US Steel faced declining market share

This summarizes, then, the existing literature on the US Steel merger. The prevailing explanation seems to be that US Steel was formed to act as a dominant firm in the steel

industry. The firm may or may not have acted as a passive giant in its first few years of existence, but certainly the merger did lead to price stability, either through maintaining its own price or facilitating oligopolistic pricing. The US Steel merger was brought about in order to produce long term changes in industry structure.

A Better Explanation for the Formation of US Steel

While I believe that the existing literature is broadly accurate in its portrayal of the industry, it is incorrect on a number of narrower issues. In particular, US Steel was not created in order to form a dominant firm in any of the downstream industries (pipes, nails, rails, etc.). Nor was the merger intended to form a dominant firm in the steel ingot industry. Instead, US Steel was created as a method of preserving downstream profits by delaying entry into downstream monopolized industries.

It is easy to demonstrate that the US Steel merger did not create monopolies in the downstream industries (pipes, nails, etc.). An examination of the component downstream firms should be evidence enough.

The American Steel and Wire Company, an 1899 consolidation of most US wire firms, had a virtual monopoly on wire and rod production, and was an important purchaser of heavy steel

products.²³⁵ Earlier attempts by J.P. Morgan to organize a wire trust had failed, and this organization was brought about by John W. Gates (who had connections to both Morgan and Gary). The company had some steel making capacity, but was dependent on heavy steel companies for most of its inputs.

The National Tube Company, which formed in 1899 from the merger of 25 small pipe makers, owned about 75% of the industry capacity.²³⁶ J. P. Morgan had a role in this merger. The new company inherited some steel making capacity, but purchased much of its steel from Federal and Carnegie.²³⁷

The American Bridge Company, incorporated in 1900, was the result of the merger of 27 bridge erection firms. It was another Morgan merger, and had approximately 90% of the bridge tonnage (both girders and erection) in the US. The firm did not have any steel making capacity, and purchased most of its steel from Carnegie.²³⁸

The American Tin Plate Company was formed in 1898 from the merger of 38 tin plate firms. The organizer, W. H. Moore, was able to purchase the most “complete” monopoly of any

²³⁵ Hogan p. 263.

²³⁶ Hogan p. 275.

²³⁷ Meade p. 538.

²³⁸ Hogan p. 272.

steel merger, with virtually every tin plate plant in the US.²³⁹ This evidently was easy (or relatively inexpensive), for fierce price competition had preceded the merger. American Tin did not make any of its own steel, but instead purchased it from Federal and Carnegie. Moore then went on to organize the National Steel Company (described below) as an alternate supplier of steel ingots.

The American Sheet Steel Company has an almost identical history to the American Tin Plate Company. It also was organized by Moore (in 1899), but had only about 70% of industry capacity.²⁴⁰ Moore's fourth combine was the American Steel Hoop Company (1899), which produced steel bars, hoops, bands, cotton ties, and iron skelp. It too purchased all of its steel from the National Steel Company, and was made up of previously competitive firms.²⁴¹

All of these mergers occurred before 1901, therefore the US Steel merger did not create any downstream monopolies -- these industries were already monopolized. How do we know that these 1899 mergers created monopolies? McCraw and Reinhardt display changes over time in the price of three downstream industry outputs; plain wire and wire nails, beams and bars, and tin plate. All of them display large price increases in 1899/1900, followed by an

²³⁹ Hogan p. 290.

²⁴⁰ Hogan p. 296.

²⁴¹ Hogan p. 299.

extended period of stable but decreasing price. The evidence, once again, demonstrates that these industries were “monopolized” before the formation of US Steel.

This leaves the alternative that US Steel was created in order to form a dominant firm in steel ingots. Since ingots are the most basic form of steel, they are the output normally cited when analyzing the industry. In the previous review of the existing literature, I raised criticisms of this idea. US Steel did not act as a dominant firm in the first few years of its existence, for it expanded capacity and operated at a capacity utilization at a level above industry average. And there are additional reasons to doubt that J.P. Morgan was trying to create another dominant firm.

Ray and Parsons argued that US Steel, in combination with other large steel producers, limited entry by restricting access to iron ore. This is probably correct -- but this only describes how these firms were able to limit access years after the merger. At the time of the merger, prospects for limiting access to ore were quite poor.

The authors quote, at great length, various statements by Charles Schwab in 1899 (when he worked for Carnegie) indicating there existed good prospects for further *decreases* in the price of ore. The supply of ore appeared to be expanding dramatically for two reasons.

First, with the increasing use of the open hearth method, large deposits of high phosphorus ore were now usable in making steel. Combined with decreases in transportation costs (see above), the ore from the Upper Great Lakes region seemed to supply an inexhaustible source.

Next, discoveries in Cuba promised equal expansions in supply. The Cuban ore would be transported by sea going vessels, and the international nature of this shipping make it less susceptible to monopolization than the dedicated transportation network developing in the northern states and great lakes.

In fact these Cuban ore supplies did not fulfill expectations, and the Pennsylvania Steel Company, which owned them, was purchased by Bethlehem Steel. But these developments were unforeseen in 1901. Limiting entry into steel ingot production by means of monopolizing sources of ore would not have appeared a successful strategy at the time of the merger.

There is another reason not to attempt to create a dominant firm in steel ingots. It is unlikely to be a successful strategy when the existing firms have older, higher cost technology. But this was exactly the situation in 1901. The existing firms had a mixture of Bessemer and open hearth plants, and the latter was slowly replacing the former for reasons of cost. To restrict output (or the growth rate of output) was to invite new entry by lower cost rivals with new open hearth plants, placing the dominant firm at a cost disadvantage. The open hearth

method of production was also widely cited as suitable for smaller plants, lowering the required scale of production.

Finally, why would the organizers of US Steel wish to include previously formed monopolies? In the case of horizontal merger, the resulting monopoly is worth more than the constituent firms, since the new firm may sell at the monopoly price instead of the competitive one. But this does not apply to previously formed monopolies -- which are already charging the monopoly price. So what is the benefit of including them in the merger?

If Morgan really was trying to organize a dominant firm in the heavy product industry, one would think that he would have purchased only heavy product producers (and more than the 45% of blast furnace output controlled by US Steel). Why should he include the downstream finished goods producers? They were monopolies, and he therefore would have to pay the monopoly price for them. Morgan must already have been suffering financial constraints by assembling a \$1.4 billion dollar merger; he certainly did not need to increase his expenditures on firms that did not assist in control of heavy steel production.

As a final critique of this idea, why did US Steel fail to modernize or streamline its corporate organization? McCraw and Reinhardt note that the company retained the structure of a loose holding company, which resulted in poor control over costs, needless duplication of sales, etc. These authors attribute this inefficiency to the fact that US Steel's leadership was

dominated by financiers and lawyers, instead of “production” managers. But this is to claim that Gary and Morgan were not maximizing -- after all, the new streamlined management forms were already developed.

The strait forward application of the dominant firm model, to either the steel ingot or finished product market, does not seem correct. I believe that it may only be applied in this manner.

When a dominant firm is created, the profitability of the merger depends on how quickly the fringe is able to expand, both in number of firms and output per firm. The high price creates incentives to enter and expand, but numerous factors limit the rate that fringe output increases.

Entry is impeded by the cost of acquiring knowledge. Outside firms must discover that it is profitable to enter, and then acquire technical production knowledge. I would expect that firms able to acquire this knowledge at a lower cost would enter first.

In the case of one of the downstream dominant firms, for example the National Tube Company, there were two obvious sources of increased fringe supply. Successful firms in related industries, like wire production, could add tubing to their output. Each product was created using similar (but not substitutable) rolling types of machinery. Firms producing wire knew the price of tube inputs, the cost of rolling the product, and (to a lesser extent) the

price of the finished product. Given monopoly pricing in tubes, I would expect entry by firms in related fields such as wire, plates, etc.

The other major source of entry is the upstream producers of steel ingots. These firms supplied National Tube with inputs, and therefore were familiar with the productive process. Prices of various inputs and finished goods were more readily acquired by steel makers than unrelated firms. Steel ingot makers had to decide what degree of vertical integration maximized profits, and an increase in downstream output prices increased the probability that one more step in the vertical chain could be included.

This is exactly what happened. Carnegie Steel, which supplied much of the steel to National Tube, placed an order for a plant to make tubes. National Tube threatened to increase its own production of steel ingots and cut off purchases from Carnegie. This event was not unique; the same change in relation occurred between Carnegie and American Steel and Wire Company. The organizers of these downstream mergers did raise prices and profits, but in doing so created powerful incentives for new entry.

My hypothesis is that in order to prevent entry into the industries of these downstream monopolies, their organizers (Morgan, Moore, Gates, Gary) identified all of the likely entrants, and merged them into one firm -- US Steel. All of the downstream monopolies had

to be included, for each was a likely entrant into the markets of the rest.²⁴² Upstream producers of steel ingots, especially those that were disproportionately likely to integrate downstream, had to be included. This is the reason “rationalization” of the corporate form did not occur; the entire purpose of the merger was to delay entry by one firm into the markets of the others. Rationalization did nothing to further this (short term) goal.

My explanation changes the focus of the analysis of the merger from a horizontal to a vertical combination. Given that the vertical relation is more important, I must rule out competing arguments for vertical integration.

One possibility is that US Steel was created to eliminate double monopoly markup.²⁴³ It might be the case that with the monopolization at different levels in the productive chain, the final output price exceeded the monopoly price. Total industry profits might have increased by vertical integration, which would allow transfer of inputs at the efficient price.

This explanation does not adequately fit the facts. While the downstream industries were certainly monopolized, it is unlikely that they paid the monopoly price for steel ingots. A

²⁴² Another reason for including downstream monopolies is that expansion was often financed through retained earnings, and these new firms were making substantial profits. This also made them more likely to enter each other’s markets.

²⁴³ This is the first real “efficiency” argument I have presented.

review of the three largest steel ingot producers, all of which joined US Steel, should make my point.

I have already described the Carnegie Steel Company. Because of its size, efficiency, and threat of vertical integration, it had to be included in the merger. The organizers of US Steel explicitly demanded the retirement of Andrew Carnegie, who had spent much of his career undercutting rivals' prices and stealing their markets. Carnegie did not aid cooperative industry pricing in steel ingots.

The Federal Steel Company was much like Carnegie's company in its size and degree of vertical integration, except that it was the product of mergers of successful companies. Its organizers, J.P. Morgan and Judge Gary, were associated more with the financial end of the business.²⁴⁴ J.P. Morgan, in particular, was unlikely to form an organization that would raise the input price to the downstream monopolies he had created.

The National Steel Company, which formed in 1899, was the consolidation of eight other steel companies, and was also well integrated vertically.²⁴⁵ It was created by W.H. Moore, and shared the same board of directors with the three downstream monopolies Moore had previously organized. The double monopoly markup argument does not work, for National

²⁴⁴ Hogan p. 257.

²⁴⁵ Hogan p. 284.

Steel supplied steel to these downstream firms, and all four firms were controlled by the same individuals.

The heavy product producers that I would expect *not* to be included (and more than half were not, since US Steel had only 45% of blast furnace output), would be those firms in *other* downstream industries that had integrated upstream in order to provide their own raw materials. The other group of heavy steel producers I would expect to be excluded consists of heavy steel producers that, because of poor facilities and management, could not significantly expand their output in the face of higher prices.

Let me now review the major firms left out of the deal, and note why some of them should have been included in a heavy products merger, but would be excluded from a merger designed to protect downstream industries. These firms grew to be US Steel's largest rivals.

Bethlehem Steel Corporation. This firm's chief output was armor plate for the War Department. After the US Steel merger, Bethlehem was purchased by Charles Schwab. He was another hard driving production man working first for Carnegie, and at the time of his purchase of Bethlehem, he was chairman of US Steel. He quickly sold Bethlehem to J.P. Morgan, but bought it back after leaving US Steel (under questionable circumstances - Judge Gary eventually replaced him).²⁴⁶ When he took control he began an extensive expansion

²⁴⁶ McCraw p. 595.

plan, building new steelworks directly in competition with US Steel. Bethlehem, because of its poor financial and organization situations, as well as its specialization in armor plate, did not appear to be a potential threat to US Steel's downstream monopolies.²⁴⁷

The Pennsylvania Steel Company. This firm was associated with the Pennsylvania railroad; it was started by the railroad as a supplier of rails. Rails and railroad bridges continued to dominate its output. It did produce its own steel, which is worthy of note in that it used Cuban ore (instead of Great Lake ore), and was therefore free of the domestic limitations. Pennsylvania was unlikely to enter the pipe or nail businesses, because it both was specialized for railroad production, and also suffered a series of financial problems related to the Spanish American War. It was purchased by Bethlehem in 1915.²⁴⁸ Because of its (potentially) large ore holdings, it should have been included in a steel ingot dominant firm merger.

Republic Steel Corporation. This firm specialized in merchant bar iron and steel (a type of rolled product). It had earlier agreed to exit from the production of sheet metal, and exchanged its plants for the American Sheet Steel company's merchant bar plants. Republic was in poor financial condition, and had many antiquated plants at the time of the US Steel

²⁴⁷ Hogan p. 537.

²⁴⁸ Hogan p. 549.

merger,²⁴⁹ but was able to expand under Robert Gates, a man associated with one of the firms that merged into US Steel. Republic was again a specialized downstream producer unlikely to compete with US Steel's monopolies.

Jones & Laughlin Steel Corporation. This firm formed in 1901 from the merger of two older steel making firms. It expanded its output substantially from 1901 to 1905, particularly in the area of semifinished goods. Its expansion was limited however, by the fact that there was not any room at any of its Pittsburgh plants for additional production. It did not have any facilities for tubular, tin mill, or wire products, and it took until 1905 for the firm to find a suitable location for plants to compete with US Steel's downstream dominant firms.

Inland Steel Company. This firm, founded in 1893, grew rapidly in the first half decade of this century. However at the turn of the century it, "Could hardly have been considered a significant factor in the steel industry."²⁵⁰ It had a very small facility of inferior equipment used for making rolled semi-finished goods, and did not have any steel making facility until the end of 1901. It did expand dramatically, but only by building new plants.

The Armco Steel Corporation. This firm was incorporated in 1899, and unlike most of the new incorporations, was not the result of the merger of previous firms. Armco built entirely

²⁴⁹ Hogan p. 562.

²⁵⁰ Hogan p. 601.

new plants, of modern design, and could produce both steel and semifinished goods. Armco was however a startup company, and at the time of US Steel's formation, was not yet of any significance in the steel market.²⁵¹

These firms eventually developed the capacity to enter US Steel's markets - as witnessed by US Steel's decline in market share. The key to my argument is that they were excluded because they would take a relatively long time to enter, whereas the included firms had capital and management capable of entering in a shorter time period. Both the "Moore Group" and "Morgan Group" had steel making and rolling capability. Carnegie was already integrating downstream. Monopolizing the steel ingot market, as opposed to protecting the downstream monopolies, would have led to a different distribution of included firms in US Steel.

Summary and Conclusion

I have argued that the existing economics literature is more or less correct in its analysis of how the steel industry changed after the US Steel merger. But I have argued that these changes, while *ex post* profitable for the industry, would not have appeared so at the time of the merger. The correct reasons for the merger were of a shorter time horizon.

²⁵¹ Hogan p. 615.

Because of the change in technology from Bessemer to Open Hearth, and because of the difficulty of controlling ore given the new discoveries in Cuba, I have argued that US Steel was not formed in order to create a dominant firm in steel ingots. It also did not act like a dominant firm in its capacity utilization or expansion during 1901-1906.

Likewise, the downstream rolled products industries were not monopolized by US Steel, for the simple reason that they had been monopolized a few years before. US Steel was created in order to delay entry into these markets.

The dominant firm model does explain the mergers in the finished product industries, and for many decades has been the standard explanation for the formation of US Steel. If my analysis is correct, the standard analysis is also mistaken. JP Morgan formed US Steel to protect dominant firms, not to create them.

CHAPTER 6: UNITED SHOE MACHINES

On February 18, 1953 the District Court of Massachusetts found the *United Shoe Machinery Corporation* liable, under Section 2 of the Sherman Act, of monopolizing the shoe machinery market.²⁵² By May the Supreme Court, without opinion, upheld Judge Wyzanski's District Court opinion and decree. The government complaint was essentially that United Shoe Machinery Corporation was a monopolist in the shoe manufacturing machinery market, and that it maintained this market position via contractual restrictions that exceeded legitimate business practice.

In this chapter I will propose and evaluate alternative theories explaining these contractual restrictions. In particular, did United use leasing contracts because they best motivated the efficient production of machines and knowledge, or instead, did these contracts allow United to increase its profit at the expense of efficiency?

The chapter is broken down into four sections. The first provides a summary of the historical development of the shoe making industry. The second describes the industry at the time of the antitrust cases. Section three, the core of the chapter, presents important economic

²⁵² USM was also found liable of monopolizing the market for shoe machinery supplies, a related topic I will not address. Kaysen 301.

explanations of United's use of leasing. The final section evaluates the relevance of these alternative explanations.

I. A Brief History of Shoemaking in the United States.

There are four basic processes in making leather shoes; cutting, fitting, lasting, and bottoming.²⁵³ Cutting involves, as its name implies, cutting leather into the various pieces of the upper, insole, and sole of the shoe. Fitting means sewing the various pieces of the upper together. Lasting involves pulling the upper over a foot-shaped object called a last, to which the insole was previously attached, and attaching the two. Bottoming is attaching the outer sole to the rest of the assembly.

With the low level of urbanization and specialization during the initial colonial period (1600's), most shoes were made by individuals for member of their own families (home production). The leather was a byproduct of the slaughter of the family cow, and the quality of construction and individual fit was obviously dependent on the skill of the farmer. Only in the most urban areas (Boston, Philadelphia) did the size of the market suffice for specialized shoemakers of the sort found in Europe.

²⁵³ Hazard, p. 3.

By the middle of the eighteenth century the population had increased sufficiently so that the majority of shoes were made by specialized shoemakers. But production was still in essence custom; shoes were made only at the request of specific customers, either in the shops of master shoemakers in towns, or by traveling journeymen. These shoemakers still made complete shoes individually, without specialization in process.

Gradually this form of organization began to produce non-custom shoes.²⁵⁴ Apprentice shoemakers were kept busy in periods of low demand with the buildup of inventory. Once a market for ready made shoes developed, further specialization developed by the entry of businessmen (usually shopkeepers) who took on wholesale and retail tasks. This was the “putting out” system; a central agent delivered raw materials to craftsmen, who constructed shoes in their own establishments.²⁵⁵ This change was again driven by market conditions, including the enlargement of the market via inter-colony trade, then later the demand for footwear by the continental army.²⁵⁶

The shoe wholesalers and their “central shops” gradually took on more than distribution tasks; they also moved into those processes in which they had a comparative advantage. Cutting, for example, yielded different amounts of scrap leather depending on the skill of the

²⁵⁴ Hazard p. 11

²⁵⁵ Shoemakers usually banded together to share shop space. These shops were generally known as “ten footers”. See Mulligan, p.59

²⁵⁶ Hazard p 26.

craftsman. Since the central shop owner also owned the leather, but the craftsman owned the scrap (which was valuable), craftsman had an incentive to cut as few shoes as possible out of each hide. The cutting of the leather was one of the first tasks that was taken into the central shop.²⁵⁷

Between the 1850's and 1870's more and more tasks were performed within the central shop, which evolved into the factory. While machines began to be used for various tasks in the production of shoes as early as the 1820's (there were no specialized machines for shoe production before that time), the development and usage of power machines seemed to follow, rather than precede the centralization of production under one roof.²⁵⁸ The basic pattern is identical to that described by Adam Smith in *The Wealth of Nations* (his pin maker example); the degree of specialization was limited by the extent of the market. With an expanding market, the shoe trade changed from one in which craftsmen made entire shoes to one where workers performed much more specialized skills (such as only cutting leather) at far greater speed.

Mechanization and the Formation of United.

²⁵⁷ Note that comparative advantage may be a question of incentive, not just skill.

²⁵⁸ Mulligan, p. 61.

This factory type of production, where each worker only performed a few specialized tasks, was a precondition for the development of shoe making machines. Before specialization in task, each craftsman was a specialist in making shoes - but a generalist at each particular sub-task. Once workers specialized in particular tasks (e.g. punching holes for eyelets) the invention of machines to perform those tasks became, for several reasons, more advantageous. First the task itself became simplified. No longer was a shoe making machine required, but instead an eyelet punching machine. Complex operations were reduced to their component steps - and the cost of designing a machine is fundamentally related to the complexity of the task.

Next, since workers performed the same operation multiple times (perhaps hundreds or thousands of times per day), the relatively fixed cost of designing or purchasing a specialized machine was divided over many units. Finally, the sheer repetition involved in performing the same task often made innovation easier. The worker himself experimented with different hand techniques for increasing speed. This knowledge was useful in sparking creative ideas on how a machine could perform the task.

By the late 1850's the sewing of uppers was generally performed on sewing machines, a task that had been previously performed by the wives and daughters of shoemakers.²⁵⁹ By

²⁵⁹ Hazard. P. 109. Mary Blewett notes that women never were artisans in the shoe industry. They performed most of the hand sewing of uppers before the application of sewing machines, but the skill they used were common to almost all women of the period. They

1858, Lyman R. Blake patented a machine (named the McKay machine after the purchaser of the patent) for sewing the sole of the shoe to the upper.²⁶⁰ The McKay machine did not see widespread use until the civil war, when the large orders from the Union Army for fairly homogeneous boots made the installation of the complex machine profitable. 1875 saw the introduction of another method of attaching the sole to the upper, the Goodyear Welt machine.

As compared to other industries, this conversion to both the factory system and machine operation was relatively late.²⁶¹ Explanations for each are probably related to the characteristics of the primary raw material - leather. For unlike products such as iron or steel, leather is remarkably non-homogeneous. It varies in thickness, and designers of shoe making machines have always had difficulty constructing machines that are capable of operating consistently with so variable an input. Furthermore factory production is fundamentally one of standardization. All economies of specialization arise out of performing the same task with great speed. With a non-homogenous input, standardization is difficult. Thus shoemaking has always been a labor intensive task, and single craftsmen that

did not therefore have craftsman status. With the introduction of mechanization, women eventually worked in all phases of shoe construction, but this time as specialized machine operators. Again they did not achieve independent artisan status.

²⁶⁰ Hazard p. 245.

²⁶¹ Hazard p. 115.

performed all stages in the production of a single pair of shoes were better able to adjust one process as the result of inconsistencies in previous stages.²⁶²

While sewing machines (for uppers) seem to have been purchased outright, this is not the case for the more specialized shoe making machines. McKay, who first introduced the machines for “bottoming” (sewing the sole to the rest of the shoe), used a royalty system and did not sell machines. He provided three reasons for exclusively leasing shoe machines. First, the machines were not yet perfect - his firm was still working out the bugs. With leasing, machine errors resulted in less loss of goodwill than did outright sales. Next, McKay claimed that if he tried to sell his machines, only the wealthiest shoemakers would be able to purchase them.²⁶³ He was attempting to finance poor shoe manufacturers. Finally, he believed that revenue would be maximized by steady income, as opposed to one time purchases.²⁶⁴

Throughout the rest of the century a number of shoe machine companies formed, each tending to specialize in either a separate task of shoe making or alternative processes. For example, the *Goodyear Shoe Machinery Corporation* had a 60 percent market share in welt

²⁶² The highest quality shoes are still hand made. See Boon p. 54.

²⁶³ Of course McKay, as the monopoly supplier of these machines, preferred a competitive downstream market as a way of maximizing his rents.

²⁶⁴ As per our general suspicion of surveys, there is no particular reason to believe McKay's listed reasons are in fact the ones that motivated him. The early lease terms were one to five cents per pair of shoes stitched. Hazard p. 121-122.

inseamers and sewers, and a 10 percent share in lasters. The *Consolidated and McKay Lasting Machine Company* had a 60 percent share of lasting machines. The *McKay Shoe Machinery Company* had 70 percent of the heeling market, and 80 percent of other metallic fastening devices. The *Eppler Welt Machine Company* had 10 percent of inseaming and welt sewing, and the *Davey Pegging Machine Company* had 7 percent of pegging machines.²⁶⁵

In 1899 these five listed companies merged into the *United Shoe Machinery Company* (later Corporation). Between this date and 1911 United made 59 acquisitions, only one of which was significant in size.²⁶⁶ United had only one plant, in Beverly, Massachusetts, but 27 branches in 13 states for sales and service. United provided service without charge on all of its machines. United also had a large research department, which produced about 150 new machines between United's formation and 1914.²⁶⁷

In 1911 the government filed its first antitrust suit against United, claiming that the merger was a combination in restraint of trade (Section 1 of the Sherman Act), and that the leasing contracts, with no additional charge for service (tying), were violations of Section 2. The government lost the case, with the court (upheld by the Supreme Court) ruling that the initial

²⁶⁵ Kaysen, p. 7 footnote 13. The welt is a strip of leather attaching the upper and lower sections of the shoe.

²⁶⁶ United paid \$6 million for the Thomas G. Plant Shoe Company, which had recently integrated upstream into shoe making equipment.

²⁶⁷ Kaysen, p. 8-10.

merging companies had complementary systems, and were not therefore competitors, and that the leasing system was a valid application of United's patent rights.

Shortly after the 1914 passage of the Clayton Act, the government again filed suit, claiming that United's leasing policy violated Section three of the Act, which forbade contracts that restricted the purchaser's right to use goods supplied by rivals (tie ins). The district court (again upheld by the Supreme Court) found some provisions of United's leases illegal, but upheld others included in the Department of Justice's complaint.²⁶⁸

In particular, the Court found that lease provisions that tied bundles of United machines together (if the shoemaker used one machine, he had to use them all), or tied the use of the machine with supplies provided by United, violated section three of the Clayton Act. Lease provisions that were not in violation included full capacity clauses (a machine had to be used to full capacity for any work for which they were suitable), the tie in to parts and service (which had to be purchased from United), and clauses allowing United to remove machines that were excess capacity.²⁶⁹

In order to comply with the courts decree, United negotiated with the National Shoe Manufacturers Association, and formed new contracts excluding those provisions the court

²⁶⁸ Kaysen, p. 15.

²⁶⁹ Kaysen, p. 15-16.

found unacceptable. United still exclusively leased some machines (but the term was shortened from the previously used 17 years), exclusively sold others, and offered the option for the rest.

In 1947 the United States Government again brought suit against United under Sections 1 and 2 of the Sherman Act.²⁷⁰ The government again complained that United monopolized the shoe machine industry via its leasing contracts,²⁷¹ and the court's ruling set the important legal precedent for interpreting the legality of leasing.

II. Industry Structure at the Time of the Case

The Structure, Conduct, Performance Paradigm and the Initial Analysis of the Case.

The foremost source of information relating to the United Shoe Machinery antitrust suit is the doctoral dissertation written by Carl Kaysen.²⁷² Charles Wyzanski, the presiding judge, appointed Kaysen, who was a graduate student at Harvard, as a "law clerk" - really an on staff economist. Kaysen used the material presented by the contesting parties to perform an

²⁷⁰ As stated in the first paragraph of this chapter, United lost this case. While this is the last United case analyzed in this chapter, it is not the last time United found itself before the Supreme Court. See, in the last chapter, the *Hanover Shoe* case.

²⁷¹ And a host of other practices, most of which were dismissed by the court.

²⁷² This was later published. See the reference section at the end of this chapter.

analysis in a form suggested by the then newly emerging “Structure, Conduct, Performance” paradigm.²⁷³

This method of industrial analysis is best understood in contrast to the prevailing practice. Since antitrust violations were violations of the law, most previous cases had been prosecuted in the same way as other “crimes” (e.g. murder). The law stated that some conduct was illegal (usually because it was immoral), and prosecuting a case involved proving that the defendant in fact committed an act that violated the law.

The new paradigm of analysis did not look for “guilt” per se. Instead it proposed that the conduct of a firm was derivable from the industry structure (market share, demand, barriers to entry, etc.). If an economist knew this industry structure, then he would also know what actions firms would take, and those actions could be evaluated using efficiency criteria (this is the performance element). According to this method, a prosecutor should not look for illegal actions, but instead for illegal market structures.

Market Structure

²⁷³ The economists Joe Bain and Edward Mason are most commonly associated with this form of industry analysis. George Stigler is the best known critic of this method.

The first element of a “market power” test is defining the market. What products count as part of the relevant market? United was clearly a monopolist over new units of its own production, and almost certainly had insignificant market power in the “machine tools” market. United made hundreds of different types of machines, most of which performed unique functions. Kaysen decided that the relevant market was for machines of the type produced by United, and did not include any machines not produced by United. For example the Singer Sewing Machine company produced most of the machines used in sewing uppers. Kaysen did not believe that sewing machines for uppers were part of the relevant market, since United no longer produced machines of this type.²⁷⁴

Kaysen was quite aware how potential competitors - firms that would enter the market if the price rose sufficiently - could theoretically constrain United, and would therefore have to be included within the bounds of the market. However he denied the potential of entry, citing the specialized knowledge required for shoe making machine design.

Given that the relevant market was coextensive with United’s products, Kaysen found that United had over 90 % of the market for “major” machines, and somewhat less (60-80 %) for “minor” machines, with a weighted average of between 81 and 86 %. Campo, United’s

²⁷⁴ United had produced machines for sewing uppers, and Singer had produced machines of the type produced by United. After a period of some competition, the two firms evidently decided to split the market, a charge explicitly denied by United. Kaysen p. 360.

largest competitor, had approximately 3.4% of the market for major machines.²⁷⁵ As opposed to this, the shoe manufacturing industry had thousands of firms of differing sizes. Some shoe manufacturers were of the same size (presumably in dollars of sales) as United.²⁷⁶

As explained in chapter five,²⁷⁷ the ability to raise price above marginal cost (market power) is dependent on the elasticity of residual demand. Residual demand, in turn, is the difference between the demand for the product (as measured by the size of the market) and supply by other producers. Even if market demand is inelastic, a firm may not have market power as long as the competitive fringe has relatively elastic supply.

Kaysen argued that United had market power. Market demand was relatively inelastic, since the market was limited to shoe machines, for which there were few good substitutes. And the fringe supply was inelastic, since (1) other firms would not enter due to the specialized knowledge requirements, and (2) the existing competing suppliers did not have the financial resources to expand production.²⁷⁸

²⁷⁵ Kaysen, p. 47, 52. The distinction between major and minor machines was drawn from the government's complaint. Major machines are those with poor substitutes, such as hand labor. Kaysen p. 32.

²⁷⁶ Kaysen, p. 27, claims that in 1947 there were 1650 shoe factories operated by 1460 firms, and one or two were the same size as United.

²⁷⁷ See also the appendix on the Dominant Firm.

²⁷⁸ Kaysen cites total assets of United vs. its competitors. P. 53.

Conduct and Performance

Given Kaysen's finding of market power, he naturally expected to find various forms inefficient behavior. He was not disappointed.

He did not find "coercive" behavior, at least toward shoemakers. The government was unable to find hostile shoe manufacturers to complain about United's abuse of its monopoly position. Shoe manufacturers were usually quite grateful to United, for it often assisted them in starting their business with both the equipment (leased, which did not require up front payments) and expertise. United maintained cordial relationships with shoemakers, charged prices that they considered low, and provided excellent service.²⁷⁹

The inefficiency that Kaysen found was instead widespread price discrimination. While United did not offer different prices to different customers (the normal method of price discrimination), it failed to charge higher prices in the face of higher costs - which is equally valid a method of price discrimination. For example, by not charging marginally for service United provided differing levels of service for the same price. This seems to have been Kaysen's major criticism; that United was not using the prices it charged to lead customers to minimize the total "social" cost.

²⁷⁹ Kaysen, p. 202.

Kaysen's arguments and analysis fundamentally depend on the conclusions (1) that there were substantial barriers to entry in shoe machines, and (2) that leasing was the significant barrier.

Leasing

We then come to the heart of the case, both legally and economically. Why did United's lease many of its major machines? Monopoly explanations include leasing as a barrier to entry, or leasing as an aid in price discriminating (or preventing unprofitable price discrimination). As opposed to "market power" explanations, some authors have argued that leasing is the efficient contract; that it optimizes various tradeoffs in manufacturer and customer performance. Before presenting these alternative explanations for leasing, allow me summarize United's leasing terms.

Contract Types

Masten and Snyder²⁸⁰ provide the following useful breakdown of contract types that existed at the time of United's trial. Each number represents the number of machine types under each contract type:

²⁸⁰ Masten & Snyder, p. 51.

Sale only: 42

Option sale or lease: 122

Lease with monthly rental charges only: 88

Lease with monthly rental charges plus unit charges: 85

Lease with unit charges only: 6

Revenue for United was then broken down into four categories: the direct sale of machines, monthly rental charges, unit charges (i.e. a fee for each operation a machine makes), and “deferred payments” (this was a lump sum payment due at the expiration of the lease).²⁸¹

The duration of the leasing contracts steadily declined over time. Initially duration was 17 years, but as the result of negotiations between United and a shoe manufacturer trade group (following the 1914 case) duration was reduced, and then reduced again in 1922 and 1939. By the time of the case, lease duration was ten years.²⁸² The average life of a lease, counting

²⁸¹ Masten & Snyder, footnote 70.

²⁸² There is some confusion on this point of fact. Masten & Snyder, p. 57 cite 10 years. Kaysen, p. 64, also cites 10 years. But then Kaysen writes, “On expiration of the original lease, the lessee and United may sign a renewal lease on the same terms as the original one, which, since 1939, has run for five years, and from 1922 to 1939 ran for 10 years.” Does the term he mention refer to the original or renewal?

renewals, was somewhere between eight and ten years, with a fairly large dispersion around the mean.²⁸³

Upon the return of a leased machine, the shoemaker would have to make the deferred payment, a return charge for shipping and pay for the repair of damaged parts. Furthermore, if the machine was returned early, the customer was obliged to pay 25 percent of monthly rental and 50 percent of the minimum unit charges left in the contract. However, United voluntarily (not as part of the lease contract) credited four percent of all unit and monthly charges to the account of customers "in good standing" - meaning those that had not otherwise violated the lease contract. This credit often fully paid for any shipping and repair charges.

III. Explanations for United's Use of Leasing

Why were leases used? Economists have offered a number of explanations for leasing, five of which I have summarized. Most of these are fairly complex, game theoretic models, and they require substantial effort to understand the various tradeoffs. All of them attempt to provide valid maximizing explanations of leasing *given certain underlying conditions*. I have provided a chart summarizing these conditions, with more detailed (but still simplified) explanations in the section that follows.

²⁸³ Kaysen p. 65.

Table 2: Alternative Explanations for Leasing

<i>Explanation</i>	<i>Leasing Functions As A(n):</i>	<i>Required Conditions</i>
Masten & Snyder	Efficient Contract	<ol style="list-style-type: none"> 1. Moral hazard in the provision of durability. 2. Moral hazard in the provision of technical knowledge. 3. Double moral hazard in the provision of service.
Kaysen	Barrier to Entry	<ol style="list-style-type: none"> 1. No close substitutes. 2. Myopic buyers.
Aghion & Bolton	Barrier to Entry	<ol style="list-style-type: none"> 1. No alternative suppliers. 2. Uncertainty as to when new suppliers will enter. 3. Any new rivals will have substantial cost advantages.

Coase	Aid to Price Discrimination	<ol style="list-style-type: none"> 1. Short term leases (rental contracts). 2. Durable goods. 3. Heterogeneous consumers.
Waldman	Aid to Price Discrimination	<ol style="list-style-type: none"> 1. Low valuation consumers who are not efficiently served. 2. Inability to commit to repurchase durable good. 3. Depreciation in value of durable good.

Masten and Snyder - an Efficiency Explanation of Leasing

Scott Masten and Edward Snyder argue the shoe machine industry had a number of contracting problems, and that the features of United's leasing contracts best "solved" these problems. Leases best motivated the production of quality machines, associated

technological knowledge, and continuing maintenance. Most importantly, their explanation of leasing describes why we observe the variability in contract type.

The services that United sold to shoemakers were a complex bundle of equipment and knowledge. Shoemakers wanted machines that would function under diverse circumstances, some of which were not easily predictable. The machines had to be durable, in the sense that they would continue to operate as expected far into the future. Furthermore, United provided not just machines, but also continuing service. The machines had to be maintained and adjusted to operate efficiently, and had to be capable of upgrade as new shoe making innovations came on the market. Finally, United provided information, both training in how to use its machines, and training in general shoe factory layout and management.

This bundle of parts, labor, and information could be broken down into many different packages, and priced on many margins. Following Masten and Snyder,²⁸⁴ I will separate various margins of opportunistic behavior, and analyze how competing contracts minimized the associated dissipation. The final contracts, I will argue, balanced the resulting tradeoffs optimally.

Assuring Machine Quality

²⁸⁴ The contents of this section, and to some degree its format, follow that of Masten and Snyder.

Shoe machines are durable goods, with the characteristic that, at the time of contract, the benefits of owning the machine extend far into the future. The degree of durability, which is of course one of the most significant factors in the value of the machine to the shoemaker, is dependent on both how the machine was made and how it is operated and maintained. For now, let me focus on how initial design and construction affect durability.

Shoemakers' valuation of a machine varied directly with how long it lasted. But at the time of purchase, the measurement of this attribute was costly. United presumably had the most control over its construction, and therefore was more likely to know how long it would last, but United also had an incentive to overstate a machine's durability, since how much it received as payment depended on it. This was one margin for opportunistic behavior; United had an incentive to under provide²⁸⁵ in machine durability when durability was sufficiently expensive to measure, since once the machine was sold, the new owner bore all of the variability in machine life.

²⁸⁵ The optimal durability is of course when the present value of another unit of future machine life is just equal to the present value of the cost of providing it. United pays the cost; the owner of the machine receives the benefit.

There are several solutions to this durable goods problem.²⁸⁶ One contracting solution is warranties. If a machine is relatively valuable as compared to the cost of enforcing the warranty,²⁸⁷ and if it is relatively inexpensive to prove²⁸⁸ that machine did not work as specified, warranties work well in assuring optimal durability.

Another solution is reputation.²⁸⁹ Even if courts are not able to enforce warranties, firms often find it advantageous to provide the contracted amount of durability in order to protect rents from future sales. This is probably the most important function of brand names; if a company with a national reputation makes low quality goods and sells them as high quality, other customers are likely to refuse to pay for high quality in the future. Of course brand names are inconsistent with perfectly decentralized markets, which have the characteristic of producer anonymity (under what conditions is the name of a farmer growing wheat important information in its sale?).

²⁸⁶ It is a problem common to any “experience” good, where the purchaser finds it very expensive to ascertain quality at time of purchase.

²⁸⁷ An example of a good for which warranties do not work is candy. The packaging of some brands of chocolate candy offers to refund the full purchase price of any candy that is not fresh. But since the purchaser must send the unused portion back to the manufacturer, and only receives the price of a candy bar, the warranty is close to worthless.

²⁸⁸ Warranties require third party enforcement - the shoemaker must be able to prove to a court that the machine did not last as long as specified. This knowledge does not seem difficult in the case of durability; the contract could specify that the machine would work for a certain length of time or number of operations. Time is inexpensive to meter; and per operations counter were a feature of many of these machines. The more expensive element to prove is what it means to “work”. Does work mean no errors in operation ever? Or how many errors are allowed? Who counts the number of errors?

Neither solution is optimal when we add another source of variability; how a shoemaker uses and maintains the machine also changed durability. Obviously machines may be damaged by careless use or poor maintenance, and if these factors are expensive to measure (and prove in court), a warranty will not result in optimum levels of durability. Likewise, reputation does not function optimally, since other shoemakers do not know if the machine did not work because United constructed it poorly, or if instead the shoemaker misused it.²⁹⁰

Short term rental contracts minimize these margins for opportunism. By changing the transaction to one of repeated interaction, each party knows it will lose future rents by short term cheating. If United made a machine that had below optimum durability, then United would lose, since it owned the machine - the shoemaker would simply return it to United. If the shoemaker misused the machine, United would refuse to continue renting to the shoemaker.

Of course under this arrangement the shoemaker had to be monitored to guard against misuse or poor maintenance (or United might perform the maintenance). But United did not have to prove misuse in a court; it knew the degree of durability built into the machine, so if the

²⁸⁹ See Klein and Leffler for a more complete treatment of this topic.

²⁹⁰ Suppose your next door neighbor never cuts his grass or otherwise maintains his house. If he complains that his car stopped working after only seventy five thousand miles, will you refuse to buy that brand of car, or assume he never changed the oil?

machine failed prematurely it knew who was responsible, and could simply refuse future rentals. The key element is that either party must be able to end the relationship without substantial cost if the other party misbehaves.

While short term rental agreements motivate durability and the provision of service, longer term leases also motivate this if payment is based on output (i.e. the number of shoes produced). If United leased a machine that broke down, it would not generate output, so United would not be paid. However these leases did not motivate innovation in process; once United had sunk research and development costs, it would not find it profitable to incur additional research costs (even when otherwise efficient) if it was able to force shoemakers to continue to use existing machines under long term leases. This problem may be reduced by contracting terms that allow the switching to new technology.

Which margin did these lease contracts measure and price? Rentals could be by the month (time) or per operation. Time is less costly to measure, so if all other factors are the same, we expect contracts with a fee per unit time. But if durability is correlated with the number of times a machine is used, then a per unit time contract leaves this as a free good, which will lead to inefficient use on the part of the shoemaker. In cases where the valuable attribute is durability as measured by number of times the machine is used, we expect contracts based on how many times the machine is used.

Contracting for Knowledge

For a number of reasons, United Shoe had a comparative advantage in the provision of information related to the efficient operation of their machines. United acquired much of this knowledge in the design and production of the machines, and instructing shoemakers also gave United further feedback that was useful in the upgrade of existing machines and design of new ones.²⁹¹ This information was broader in scope than simple directions on any given machine, and included the integration of various machines and general shoe making know how.

This information, much like machine durability, was an experience good. Shoe manufacturers found the assessment of the quality of the information expensive (if they already knew the information, why would they buy it?). Therefore if the information was simply sold outright and up front, United had an incentive to under provide in its quality. The same considerations as above limited the ability of warrantees and reputation to solve this contracting problem.

If payment was based on use of the machines in short term rental, with an implied premium for the tied in knowledge, then the problem was decreased, just as it was with the durability

²⁹¹ As noted by Yoram Barzel, working with many machine users allowed United to acquire knowledge as to a machines generic opposed to the idiosyncratic problems.

problem. One difficulty with this was that while United could take its machines back if a shoemaker refused to pay his bill, it could not repossess the knowledge it sold. If United provided most of the knowledge at the beginning of the relationship - and was paid for it through the short term rental contracts - a shoemaker had an incentive to rent the machines, acquire the knowledge, and then replace United machines with other machines that did not have an included premium for knowledge.

Let me summarize the tradeoffs. If United sold its know-how with a cash payment up front, then it had an incentive to under provide knowledge, since the shoemakers obviously did not know the value of the product. If instead United was paid for the success of its information - via payments included in the rental of machines - then shoemakers had an incentive to acquire the knowledge with the first period of machine rental, and then substitute to rival's machines which did not include an information premium.

By having a lease, not a short term rental agreement, with payments based on output (per unit or operation), United has the proper incentive to provide knowledge. Shoe manufacturers were not able (without paying a penalty for breaking the lease) to avoid payment for the knowledge.

The Optimal Lease

The ideal lease contract attempts to balance the tradeoffs from each type of contracting problem. A long term lease, with payments based on output, motivated durability and proper maintenance on the part of United. Including a premium for an initial knowledge transfer motivated the efficient production of knowledge and training.

Further contractual constraints were required. Leasing did not motivate the development of new types of machines, so the contract allowed switching to new machines. However switching to machines without the information premium had to be prevented, so we expect restrictions to bring this about. Finally, having per operation charges lead shoemakers to substitute to machines with time delineated contracts. We therefore only expect per unit charges on machines without ready substitutes, and we expect other contractual restrictions to prevent this type of switching.

Finally, we would not expect leasing on all types of machines. Leasing required United to monitor equipment, both to provide maintenance and to ensure that inefficient switching did not occur. For some types of machines, particularly those with ready substitutes, these monitoring expenses may have outweighed the advantages of leasing - and therefore pure sales contracts were used. Because of these considerations, Masten and Snyder explain not only why leases were used, but also why there was variance across machine type in the form of contract.

The Evidence

Masten and Snyder examined how the contract type varied with machine type. They found that complex, important machines without ready substitutes tended to be leased with a per shoe fee. Less important and complex machines were leased with monthly fees. Machines with the most ready substitutes were sold outright.

There were exceptions to their predictions. For example the contracts uniformly required the shoe factories to provide maintenance (contra the District court's findings). But in fact United provided the maintenance free of charge. United's contracts did not allow shoemakers to switch to new machines, but again United did not enforce the provisions if switching was to another United machine. By differentially applying the terms of these contracts, United motivated continued machine development, while preventing switching to competitors machines in order to avoid payment for information. So even in cases where the contracts varied from their predictions, the actual practice corresponded.

Monopoly Explanations of Leasing

Leasing played a prominent role in United's case (and loss), yet at the time monopoly explanations of leasing were not well developed. This created something of a demand for arguments explaining how leasing supported monopoly, and almost all of the resulting papers

cite the United Shoe case. While most of the models are correct, in the sense that if the underlying conditions exist, leasing supports monopoly, few make any attempt to identify if these conditions existed in the case of United. In what follows I both present simplified versions of four models, and analyze if they are appropriate for United Shoe.

Leasing as a Barrier to Entry

Kaysen's Argument.

Kaysen suggested, as a remedy for correcting the market structure, that United be forced to exclusively sell its machines. His argument for why leasing is a substantial barrier to entry runs as follows.²⁹²

A shoemaker who *owned* a United machine would switch to a competing machine if the net benefits of the machine (the difference in performance between the new machine and the United one) exceeded the cost of switching (the difference between the price of the new machine and the resale price of the United machine). Or

$$\text{Value New} - \text{Value Old} > \text{Price New} - \text{Resale Old} = \text{Switching Cost Under Sales.}$$

²⁹² Kaysen's argument was not specified even to this trivial degree of completeness.

As opposed to this, if the industry used leasing, the shoemaker would switch if the net benefits of the new machine (same as above) exceed the cost of switching (which in this case is the lease price of the new machine plus any penalty payments owed to United). These penalty payments include the 25% of the monthly rental and 50% of the minimum usage charges, as well as the loss of the four percent credit. The deferred charges do not count as part of the penalty for switching, since they occur whether or not switching occurs (the only qualification is a present value question). Or

$$\textit{Value New} - \textit{Value Old} > \textit{Price New} + \textit{Penalty} = \textit{Switching Cost Under Lease}$$

Given that the benefits of switching are constant (the machines are the same in each case), a comparison of leasing versus selling is a matter of comparing the cost to the shoemaker. In each case, a new machine must be purchased, and absent other considerations, we may assume this cost will be the same (even though one is a sale price and the other is a series of payments made under lease). Since this leaves only the resale price of the owned United machine (a positive amount) compared to the return charges of a leased machine (a cost or negative amount) switching will occur more often under the sales system than the leasing system.²⁹³

²⁹³ This result derives from the fact that when a machine is purchased, its cost is sunk and does not compute in future switching calculations.

Kaysen's argument was not convincing to many economists and jurists, particularly those associated with the University of Chicago.²⁹⁴ These scholars criticize Kaysen's argument from the basis that United's customers, some of whom were as financially large as United, would be unlikely to sign contracts that created barriers to entry. Given that alternative supplies did exist (perhaps with somewhat inferior machines), and new entry did periodically occur, if the most significant result of the leasing contract was to exclude less costly suppliers, the shoemakers would not sign the contracts.²⁹⁵

Aghion & Bolton

Aghion & Bolton directly respond to that criticism. They demonstrate under what conditions it would benefit a shoemaker to sign a contract with United that excluded future entry.

The intuition behind their model is this: when United's monopoly position is usurped by the entrance of a lower cost rival, a shoemaker only benefits to the extent that the price is

²⁹⁴ Aghion and Bolton cite Richard Posner and Robert Bork in this context.

²⁹⁵ I am also not sure if the reasoning is correct. With the advent of a new superior machine on the market, the value of existing United machines will drop to reflect their inferiority. With a sales contract, this loss is born by the purchaser (since it changes the resale value). With short term rental contracts, this loss is born by United. Leases are, in this respect, like sales contracts. The question of entry needs a bit more work given this context.

lowered to United's cost (think Bertrand competition) - not to the cost of the new monopolist. This new monopolist gains the difference between United's cost and its own.

If the existing monopolist (the incumbent) and the buyer sign a contract with a "switching fee" - a penalty for using the entrant - they may together receive some of the benefit of the entrant's lower cost. The incumbent receives the switching fee if its rival enters the market, and the buyer receives a lower price from whichever supplier makes the good.

But why would a shoemaker sign a contract with a switching fee? It would not, if it knew with certainty that new entry would occur. But if there is the possibility that new entry will not occur (because the entrant's costs are not yet known), United may bribe the customer to sign by offering a contract that has the same expected value to the shoemaker as not signing.

Aghion and Bolton first illustrate their model with the following numerical example. There are three players in this model; a monopoly supplier (the incumbent), a single buyer (which has no bargaining power; it only accepts or rejects contracts), and a potential rival supplier - the entrant. The entrant is only a potential rival, because only it knows its cost of production, and it can not credibly reveal this to the buyer. The distribution of its expected cost is known to all parties, and is distributed uniformly over the range $[0,1]$. The monopolist's cost is known, and is equal to $\frac{1}{2}$ ($c = \frac{1}{2}$). The buyer has a reservation price of one, and buys zero or one units.

The model has two stages. At stage one, the incumbent makes a take it or leave it offer to the buyer, who accepts or rejects the contract based on the option that maximizes his expected profit. After the contract has been signed (or rejected), the entrant decides whether or not to enter the market. At stage two, production and trade occur. If entry occurred, trade takes the form of Bertrand competition.

If there were not a potential entrant, the incumbent would not offer any sort of contract, but would sell, in stage two, at $p = 1$. This is the buyer's reservation wage, and leaves the buyer without profit. If there was a potential entrant, but binding contracts could not be signed, the lower cost producer would supply the buyer, and the price would be equal to the cost of the high cost supplier (the Bertrand result).

If the buyer refuses to sign a contract, his expected payoff is equal to the probability that entry occurs, times his profit from entry, plus the probability entry does not occur, times his profit without entry. Since the probability that entry will occur is simply the probability that the entrant has a lower cost than the incumbent, it is equal to the probability that the entrant's cost is less than one half. Given the specification described above, this is $\Pr(c_e < \frac{1}{2}) = \frac{1}{2}$.

Or:

$$\text{Expected Profit to the Buyer} = \frac{1}{2} (1 - \frac{1}{2}) + \frac{1}{2} (1 - 1) = \frac{1}{4}$$

If the buyer refuses to sign, the incumbent's expected profits are analogously,

$$\text{Expected Profit to the Incumbent} = \frac{1}{2}(0) + \frac{1}{2}(1-1/2) = 1/4$$

Any contract offered to the buyer must leave him with at least $\frac{1}{4}$ in expected profit, or he would refuse to sign it. Likewise, the incumbent must receive at least $\frac{1}{4}$ in expected profit from a contract, or it would not offer one.

One contract that satisfies these criteria is this. The incumbent offers a price of $\frac{3}{4}$. If there is entry and the buyer breaks the contract, he must pay the incumbent a switching fee of $\frac{1}{2}$. Since the buyer must pay a switching fee of $\frac{1}{2}$, the entrant must charge a price not more than $\frac{1}{4}$, or the buyer will instead buy from the incumbent at $\frac{3}{4}$. Since the entrant now may at most receive $\frac{1}{4}$ (remember that without a signed contract, the two suppliers engaged in Bertrand competition, and the incumbent's cost of $\frac{1}{2}$ would have been the price the entrant made), the probability of entry is now reduced to the probability that the entrant's costs are less than $\frac{1}{4}$ (which in this case is also $\frac{1}{4}$).

The buyer's expected profits are now the probability of entry times his valuation less the price and switching fee, plus the probability of no entry, times the profit he receives facing a price of $\frac{3}{4}$. Or:

Expected Profit to the Buyer from Signing: $\frac{1}{4}(1-\frac{1}{4}-\frac{1}{2}) + \frac{3}{4}(1-\frac{3}{4}) = \frac{1}{4}$

This of course means the buyer is indifferent between signing and not signing. The incumbent's expected profit is the probability of entry times the switching fee, plus the probability there will not be entry times the profit from selling at $\frac{3}{4}$. Or:

Expected Profit to the Incumbent from the Contract: $\frac{1}{4}(\frac{1}{2}) + \frac{3}{4}(\frac{3}{4}-\frac{1}{2}) = \frac{5}{16}$

This contract is both profitable for the incumbent, and it decreases the probability of entry by a rival with lower costs.

Does this model apply to the facts of United Shoe Machines? First of all, entry was never strictly blocked; it was possible to equip a shoe factory entirely with non-United machines. But if we instead focus on particular types of machines which United had the largest market share (presumably because they also had the best machines), it is possible to envision that United feared entry by new machines of superior technological capability, and used leases as a way to minimize the possibility of entry.

However the model still has problems in its application to United Shoe. The possibility of entry is reduced because, at the time that the rival is able to enter, the customer(s) has already signed the lease with United. But what if new entrants knew that some customers would

always be available? Then, depending on the fraction available, the range of parameters in which entry is deterred is reduced. In the case of United, some customers were always available because (1) leases were of limited duration, and (2) new entry into the shoe making field was on the order of ten percent per year.²⁹⁶ If, following the model, an entrant was able to produce a machine with a significant cost advantage (so that it could unambiguously win any price competition with United), it would capture all of the new shoe sellers and all of the renewals. The loss in United's market share would be described as dramatic. For this reason, if no other, the model is unsatisfactory as applied to United Shoe.

Leasing as an Aid in Price Discrimination

At the very least, the above arguments prove that under *some* circumstances leasing acts as a barrier to entry. But this is not the only reason a durable goods monopolist might select leasing. Profits are increased not only by making demand less elastic (via eliminating rivals), but also by improving price discrimination. One particular twist of the following models is that leasing increases profits to a monopolist - but it does so by *reducing* the incentive to price discriminate.

Coase and Time Consistency

²⁹⁶ Kaysen, p. 55.

One explanation of leasing that is widely cited, but always for its inapplicability to this case, is a problem of time consistency. In the durable goods case, it is often called the *Coase Conjecture* after its author, Ronald Coase. While not important in explaining leasing in this case, the idea is relatively simple, and the tradeoff is worth understanding.

Suppose a monopolist faces a downward sloping demand curve, and the good he produces has the property that customers only buy once (or the good lasts for a sufficiently long time that we may model it as a once and for all purchase). Initially the monopolist sets marginal revenue equal to marginal cost, and sells that quantity. For example with a linear demand curve and zero marginal cost, the price and quantity are half of the intercepts.

In the next period, only those customers who had low evaluations are left in the market. In the example of linear demand this would be the bottom half of the demand curve. Now the monopolist knows that he could sell also to these customers; he will again set marginal revenue equal to marginal cost, and sell to half of the remaining customers - but this time at a lower price. Given perfect continuity, this process will continue forever, with the price ever dropping.

Except that this will never work; the customers who purchased in the first period knew that the monopolist would drop the price in the second period. The marginal customer - the one with a valuation just equal to the price charged by the monopolist - received a surplus of zero

and so had nothing to lose by waiting. If he waited until the second period he would have the gain from a lower price, and only lost the benefit of having the good one period later. As long as his discount rate is not too high, many first period buyers will wait until the price drops.

The Coase conjecture (later proven by other economists) was that the durable goods monopolist, under certain parameter values, would be forced to drop his price in every period to close to his cost, and that he would make larger profits if he could *commit* to not dropping the price. However it is difficult to form a credible commitment to not reduce the price, since sales that occurred in the past no longer enter into the monopolists benefits and costs.

One way to commit to not reducing the price is through short term leasing; if all leases are up for renewal every period, the demand curve is the same each period (the high value customers are not removed by previous sales), so the monopolist never finds it advantageous to drop his price. While this explains why a durable goods monopolist may wish to only lease, it does not explain United's policy, for the simple reason that the lease contracts were too long. With 17 year (later 10 year) lease contracts, the first customers to sign leases are effectively removed from the demand curve.

Waldman & Price Discrimination

Michael Waldman provides an explanation for leasing that also involves a time inconsistency and an attempt to *prevent* price discrimination. In the case where the value of a durable good depreciates rapidly over time, the existence of the used good on the secondhand market constrains how much a monopolist is able to charge for new units. Given certain parameter values, it may be worthwhile to eliminate the secondhand market by refusing to sell machines, and leasing exclusively.

Waldman's model is easiest to understand by means of a simplified numerical example. Suppose United's machines last two periods, but they experience rapid depreciation in value in the second period. Let the value to shoemakers (high value shoemakers, see below) in the first period be 120, while their value in the second period (when the machine is used) is 60.²⁹⁷ The marginal cost of producing a machine is 50 each period, and we will eliminate present value complexities by using a discount factor of one.

Clearly in this example it is efficient to have the high valued shoemakers rent a new machine every period; the increase in marginal value to a shoemaker of using a new machine (instead of a used one) is $120 - 60 = 60$. The cost of producing a new machine is only 50. The marginal cost of used machines is, of course, zero once they have been built.

²⁹⁷ This numerical example is provided by Waldman. He also provides more general conditions (constraints on parameter value) under which his results are valid.

By leasing machines with a one period rental contract, United would be able to extract all of the rents from these machines. It could charge a rental price of 120 to each shoemaker each period, and always provide a new machine. It is easy to see why United would never find it worthwhile to rent old machines to high value shoemakers.

Now I will complicate the model slightly by adding low value shoemakers, who value both old and new machines at 15. Assume that United is unable to prevent arbitrage, and that there are 10 low value shoemakers and 15 high value shoemakers. United still would not lease out used machines; the most a low value shoemaker would pay for a used machine would be 15. High value shoemakers would obviously substitute to these old machines at this price (since they are receiving no surplus on the new ones), and even adding the 10 rentals to low valued shoemakers, this results in less profit for United.²⁹⁸

This result, that United is able to extract all of the rents from its high valued customers, does not work if it makes outright sales. After all, once it had sold new units to shoemakers in the first period, they would own them in the second period. This reduces the price it is able to charge in the second period.

²⁹⁸ You may check that this really does reduce profit; profit in the second period without renting old machines is $70 \times 15 = 10,500$; profit in the second period with renting old machines is $25 \times 15 = 375$.

In fact, the highest price that United would be able to charge the high valued shoemakers in the second period would be 75. At this price they would be indifferent between buying a new machine and keeping their old one (the benefit from the old one is 60, the benefit from selling the old one for 15 and buying a new one for 75 is also 60). Since the high valued shoemakers are receiving a machine which they are able to later sell, they are in fact willing to pay more for a new machine (in fact $120 + 15$), but the combined prices for new machines each period ($135 + 75 = 205$) is still less than United would have received by leasing ($120 + 120 = 240$).

Waldman proves that United could duplicate the profits from leasing with a sales contract - by committing to a *repurchase* price of 60, and charging a new price of 180. But this is not a credible offer, for after selling new units for 180 in the first period, United finds that it is in fact better not to offer the repurchase price of 60 - it can do better by setting the price of new units at 75. Its commitment to repurchase at 60 (which duplicates the leasing contract) is not credible, for the simple reason that when it is actually in the second period, United would be better off doing something else. Since the repurchase price of 60 is not credible, high value customers are not willing to pay 180 in the first period, and the ability to duplicate the leasing contract is lost.²⁹⁹

²⁹⁹ Commitments become credible when the penalty for violating them exceed the benefits of "cheating". Well specified contracts, enforced by a third party with sufficient power, are one way to achieve credibility - but often the required conditions are not met. Converting transactions from one interaction to repeat interaction over time also supports credible

Waldman of course provides more general conditions (constraints on parameter values) under which leasing provides the largest profits to a durable goods monopolist. Descriptively, when the number of customers who are not profitably served is small, the monopolist may find it useful to eliminate any service to them because if they are supplied, higher paying customers will substitute to the low valued service.

Does this model explain why United chose to lease many of its machines? The model predicts that we will see leasing in cases where there is both depreciation and a small number of low valuation potential customers. The model does not specify whether contracts will be of long or short duration; the only important element is that the machines still have some value at the end of the lease, and that the monopolist remains owner of that value. Customers who do lease machines return them at the end of every contract.

United's shoe making machines did not experience important amounts of depreciation. Shoemakers were charged for any damage to the machine; presumably this enabled United to lease used machines. However the model only claims that the value to shoemakers of an old machine will be less than a new machine; if we interpret the process of aging as becoming

commitments, as does the related method of reputation (see Klein Leffler). These forms of interaction are also likely to result in one or both parties acting as price searchers.

technologically outdated, the model may apply in this regard. Shoe machines did experience technological improvements over time.³⁰⁰

While Waldman does demonstrate that United may have increased profits by eliminating the secondhand market for some types of machines, his model does not seem like a very satisfactory explanation of contract form. United sold some machines, gave the option for others, and strictly leased still others. Leases included per operation fees, monthly fees, or a combination of the two. Lease terms included tie ins on some machines, and none on others. Waldman's model is silent on explaining the variation in contract type, and is therefore incomplete as an explanation of United's leasing policy.

IV. An Evaluation of Monopoly Versus Efficiency Explanations of United's Leasing Contracts

While some of the monopoly explanations for leasing do seem valid in themselves, it is difficult to believe that United's contracting practices arose from these considerations. I have already detailed why these arguments are not applicable to the case of United Shoe. But there are further reasons to doubt that leasing was primarily exclusionary in intent or consequence.

³⁰⁰ Waldman does not identify the small set of potential low value shoemakers. As stated earlier, the shoe manufacturing industry was composed of a large number of heterogeneous firms.

Leasing was a standard practice from the very beginning of the shoe machinery industry. It was used by the rival companies that eventually merged to form United. Leasing was used not just by United, but also by its much smaller rivals. It is difficult to understand why Campo, with a market share in the single digits, would also use leasing contracts if their intended purpose was exclusionary. Was Campo planning to exclude United from the market?

Furthermore, as stated earlier, these monopoly explanations of leasing have little to say as to contract form. They do not predict which margins will be priced, or which machines will be sold. They also do not seem intellectually powerful; it is hard to believe that the exclusionary considerations raised would be sufficient to prevent entry (or expansion) by rivals over long periods of time if United raised its price substantially.

The fact that over ten percent of shoe machinery contracts were available to rivals (due to expiration or new entry) each year means that if there existed substantial rents in the shoe machine industry, new entry would have occurred. Leasing contracts could not bind new shoemakers.

Finally, shoemakers were not powerless. They had an industry trade organization that did in fact bargain with United. Individual shoemakers were as large as United. If United had been extracting significant rents from the shoemakers, they could have signed long term contracts

with rival shoe machinery producers, or created an organization to integrate vertically in competition with United. But the shoemakers did not do this; they generally appreciated United's service and prices.

Kaysen's analysis was correct in at least one respect. He claimed that once United was no longer able to use leasing contracts, its dominant position relative to the United States shoe industry would decline. The decree forbidding leasing was issued in 1955; by 1963 United had lost over a third of its market share.³⁰¹ But as usual, there is more to the story. At the same time that United's market share was declining, the share of foreign made shoes imported into the U.S. market was rising. By 1964 the share of foreign made shoes rose from almost none to twenty five percent, and today almost all shoes are imported.

Masten and Snyder claim it is "overly heroic" to attribute the decline of the U.S. shoe industry to the elimination of leasing; the reason is clearly the rise in relative wage rates in the U.S., which priced U.S. firms out of this labor intensive market. Yet when we examine the history of shoemaking, the dominant pattern is the replacement of hand labor by complex machinery. With the elimination of leasing, this substitution stagnated, and as the marginal product of labor rose in the United States (due to improvements in technology in other industries), increasing wages drove shoe production to lower wage countries.

³⁰¹ Masten & Snyder, p. 67.

If the efficiency explanations of leasing are correct, leasing's elimination surely was the one factor most important in the loss of U.S. shoemaking capability. Rather than protecting competition in the shoe industry, this decision may very well have destroyed it.

CONCLUSION

As recently noted by Oliver Williamson,³⁰² much of property rights theory was developed in reaction to antitrust excesses of the 50's and 60's. In the friction-less world of perfect information - the prevailing mode of analysis - it was very difficult to explain both changes in firm size and the nature of contractual restriction without some appeal to market power.

Economist of that era, correctly attempting to develop a positive science, focused on those choice variables that were easy to analyze and measure - price, quantity, and market share. This focus often resulted in the legal condemnation of business practices tailored to other considerations, such as quality and rate of innovation.

Economists since that time have extended the maximizing logic of price theory to the analysis of these other factors. The result was transaction cost or property rights economics, which proposes to demonstrate, for reasons other than simple monopolization, why firms engage in these previously condemned practices. But is any of this new analysis useful to antitrust?

That has been one of the goals of this work; to demonstrate that transaction cost economics is useful in antitrust analysis. While I do not wish to summarize the entire dissertation here

(see the introduction for that), allow me to point out a number of examples where I have used transaction cost analysis.

The second chapter, which analyzes the switch in business organization from sole proprietorships to corporations, attributes this change to the standard property rights topics of information costs and wealth constraints. Thus the corporation - the target of antitrust - originated because of property right considerations.

In the third chapter, which centers on the development of the canals, I address the question of the cost of alternative ownership arrangements, especially government versus corporate ownership. This topic is motivated by, yet distinct from, Yoram Barzel's recent work analyzing the size and scope of the state. It attempts to use efficiency arguments to explain what form of organization will own different types of monopoly assets, and is therefore a property rights argument.

In my chapter on Standard Oil, I present an entirely new (to this case) transaction cost explanation for Standard Oil's monopolization of the petroleum industry. I find it superior to existing explanations, be they predatory pricing, merger to monopoly, or Standard as a cartel enforcement agent.

³⁰² At a session organized by Douglas North at the AEA convention in December of 1997.

Finally, in my last chapter on United Shoe Machines, I defend a measurement cost argument developed by Masten and Snyder. In doing so, I argue that while various monopoly models of leasing are internally consistent, none fits the facts of the case - even though all are widely cited in that context.

Should property rights theory replace monopolization in antitrust cases? This is equivalent to asking if we should abandon antitrust law, since property rights arguments generally describe behavior as efficient reactions to contracting problems. As I argued in my chapter on United States Steel Corporation, monopolization is sometimes the motive for changes in ownership, and much of antitrust law may be viewed as an attempt on the part of consumers to prevent inefficient transfers of wealth.

Property right theory is therefore an integral part of antitrust, and does not exclude the possibility of monopolization as the sole end of a business practice. But it does demonstrate that analysts often do not know why a particular action takes place - even when it is efficient - and that analytic humility is often the most honest approach.

BIBLIOGRAPHY

Chapter I.

Allen, Douglas W., "What are Transaction Costs," *Research in Law and Economics*, Vol. 14, 1991.

Barzel, Yoram. *Economic Analysis of Property Rights*. Cambridge: Cambridge University Press, 1989.

-----, "The Capture of Wealth by Monopolists and the Protection of Property Rights," *International Review of Law and Economics*, 14 (1994).

Coase, R. H. "The Nature of the Firm," *Economica*, 4 (1937).

Demsetz, Harold. *Ownership, Control, and the Firm*. Oxford: Basil Blackwell Ltd, 1988

Klein, B., and Leffler, K. "The Role of Market Forces in Assuring Contractual Performance," *Journal of Political Economy*. 81 (1981): 615-641.

Leffler, K. and Ferguson, J. "The Organization of Production: The Case of Franchising Contracts," Mimeo 1986.

Tirole, Jean. *The Theory of Industrial Organization*. Cambridge: The MIT Press, 1988

Williamson, O. *Markets and Hierarchies: Analysis and Antitrust Implications*. New York: Free Press, 1975.

Yu, Ben T. "Potential Competition and Contracting in Innovation," *Journal of Law and Economics*, 24 (1981):215-238.

Chapter II

Bork, Robert H. *The Antitrust Paradox*. New York: Basics Books, 1978.

- Chalmers, David M. *Neither Socialism nor Monopoly: Theodore Roosevelt and the Decision to Regulate the Railroads*. J.B. Lippincott Company, 1976.
- Chandler, Alfred D. *Scale and Scope: The Dynamics of Industrial Capitalism*. Cambridge, MA: Harvard University Press, 1990.
- Destler, Chester M. *Roger Sherman and the Independent Oil Men*. Ithaca: Cornell University Press, 1967.
- Frey, Robert L. (ed.) *Railroads in the Nineteenth Century*. Bruccoli Clark Layman, 1988.
- Letwin, William. *Law and Economic Policy in America: The Evolution of the Sherman Antitrust Act*. New York: Random House, 1965.
- Miller, George H. *Railroads and the Granger Laws*. Madison: The University of Wisconsin Press, 1971.
- Mountfield, David. *The Railway Barons*. London: Osprey Publishing Limited, 1979.
- North, Douglass C. *Growth and Welfare in the American Past*. Englewood Cliffs: Prentice-Hall, 1966.
- _____. *The Economic Growth of the United States 1790-1860*. Englewood Cliffs: Prentice-Hall, 1961.
- O'Brien, Patrick. *The New Economic History of the Railways*. London: Croom Helm, 1977.
- Porter, Glenn. *The Rise of Big Business 1860-1920*. Arlington Heights, Illinois: Harlan Davidson, 1992.
- Seavoy, Ronald E. *The Origins of the American Business Corporation, 1784-1855*. Westport: Greenwood Press, 1982.

Chapter III

- Cranmer, H. Jerome. "Improvements Without Public Funds: The New Jersey Canals." In *Canals and American Economic Development*. Goodrich, Carter (Ed.). New York: Columbia University Press, 1961.

- Fairlie, John A. "The New York Canals." *The Quarterly Journal Of Economics*, Vol. 14, No. 2 (Feb., 1900), 212-239.
- Goodrich, Carter. *Government Promotion of American Canals and Railroads 1800-1880*. New York: Columbia University Press, 1960.
- Grossman, Sandord and Hart, Oliver. "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration." *Journal of Political Economy*, 1986 vol. 94, no. 4.
- Segal, Harvey H. "Cycles of Canal Construction." *Canals and American Economic Development*. Goodrich, Carter (Ed.). New York: Columbia University Press, 1961.
- Spence, Michael. "Product Differentiation and Welfare." *American Economic Association*. May, 1976.
- Rubin, Julius. "An Innovating Public Improvement: The Erie Canal." *Canals and American Economic Development*. Goodrich, Carter (Ed.). New York: Columbia University Press, 1961.
- Rubin, Julius. "An Imitative Public Improvement: The Pennsylvania Mainline." In *Canals and American Economic Development*. Goodrich, Carter (Ed.). New York: Columbia University Press, 1961.
- Tirole, Jean. *The Theory of Industrial Organization*. Cambridge: The MIT Press, 1988.

Chapter IV

- Barzel, Yoram. "The Capture of Wealth by Monopolists and the Protection of Property Rights." *International Review of Law and Economics* 14, 1994
- Bringhurst, Bruce. *Antitrust and the Oil Monopoly: the Standard Oil Cases, 1890-1911*. Westport, Conn.: Greenwood Press, 1979.
- Chandler, Alfred D. *Scale and Scope; the Dynamics of Industrial Capitalism*. Cambridge: Belknap Press of Harvard University, 1990.
- Destler, Chester McArthur. *Roger Sherman and the Independent Oil Men*. Ithaca: Cornell University Press, 1967.
- Granitz, Elizabeth and Klein, Benjamin. "Monopolization by 'Raising Rival's Costs': The Standard Oil Case." *Journal of Law and Economics* (forthcoming).

- Klevorick, Alvin K. "The Current State of the Law and Economics of Predatory Pricing." *AEA Papers and Proceedings*, May 1993.
- McGee, John S. "Predatory Price Cutting: The Standard Oil (N.J.) Case." *Journal of Law and Economics* 137 (1958).
- Miller, George H. *Railroads and the Granger Laws*. Madison: The University of Wisconsin Press, 1971.
- Posner, Richard A. *Antitrust Law: An Economic Perspective*. Chicago: The University of Chicago Press, 1976.
- Snyder, C. M. "A Dynamic Theory of Countervailing Power." *The RAND Journal of Economics* 747 (Winter 1996).
- Stigler, George J. *The Organization of Industry*. Homewood, Ill.: Richard D. Irwin, 1968.
- Tirole, Jean. *The Theory of Industrial Organization*. Cambridge: The MIT Press, 1990.
- Williamson, Harold F. & Daum, Arnold R. *The American Petroleum Industry 1859-1899*. Evanston: Northwestern University Press, 1959.

Chapter V

- Blair, Roger D. and Kaserman, David L. "Vertical Integration, Tying, and Antitrust Policy," *The American Economic Review*, 1978.
- Berglund, Abraham. "The United States Steel Corporation and Industrial Stabilization," *The Quarterly Journal Of Economics*, Volume 38, Issue 4 (Aug 1924), 607-630.
- . "The United States Steel Corporation and Price Stabilization," *The Quarterly Journal of Economics*, Volume 38, Issue 1 (Nov., 1923).
- Commons, John R. "The Delivered Price Practice in the Steel Market," *The American Economic Review*, Volume 14, Issue 3 (September 1924), 505-519.
- Hines, Gregory L. "Price Determination in the Lake Erie Iron Ore Market," *The American Economic Review*, Volume 41, Issue 4 (September 1951), 650-661.

- Hogan, William T. *Economic History of the Iron and Steel Industry in the United States*. Volumes I and II. Lexington Books, 1971.
- Lamoreaux, Naomi R. *The Great Merger Movement in American Business, 1895-1904*. Cambridge University Press, 1985.
- McCraw, Thomas and Reinhardt, Forest. "Losing to Win: US Steel's Pricing, Investment Decisions, and Market Share, 1901-1938," *The Journal of Economic History*, Volume XLIX, No. 3 (September 1989).
- Meade, Edward Sherwood. "The Genesis of the United States Steel Corporation," *The Quarterly Journal of Economics*, Volume 15, Issue 4 (Aug 1901), 517-550.
- Parsons, D.O. and Ray, E.J., "The United States Steel Consolidation: The Creation of Market Control," *Journal of Law and Economics* 18 (April 1975), 181-220.
- Robinson, Maurice H. "The Gary Dinner System: An Experiment in Cooperative Price Stabilization," *The Southwestern Political and Social Science Quarterly*, 7 (Sept. 1926), 137-61
- Stigler, George. "The Dominant Firm and the Inverted Umbrella," *Journal of Law and Economics* 8 (October 1965), 157-71.
- Stigler, George. "Monopoly and Oligopoly by Merger," *The American Economic Review*, Volume 40, Issue 2, Papers and Proceedings (May 1950), 23-34.
- Weston, J. Fred. "The Role of Mergers in the Growth of Large Firms", *The Journal of Law and Economics*, 23 (1953).
- White, Alice Patricia. *The Dominant Firm: A Study of Market Power*. UMI Research Press. Ann Arbor, Michigan, 1983.

Chapter VI

- Aghion & Bolton. "Contracts as a Barrier to Entry," *American Economic Review* 77 (1987), 388-401.
- Alchian, Armen & Allen, William. *Exchange & Production 3rd Edition*. Belmont: Wadsworth Publishing Company, 1983.

- Blewett, Mary H. "Mechanization and Work in the American Shoe Industry: Lynn, Massachusetts: Discussion." *The Journal of Economic History* March 1981, 64.
- Boon, Gerard K. *Technology and Employment in Footwear Manufacturing*. Alphen aan den Rijn, the Netherlands: Sijthoff & Noordhoff 1980.
- Coase, Ronald. "Durability and Monopoly" *Journal of Law and Economics* 15, 143-149.
- Hazard, Blanche Evans. *The Organization of the Boot and Shoe Industry in Massachusetts Before 1875*. Cambridge: Harvard University Press, 1921.
- Kaysen, Carl. *United States v. United Shoe Machinery Corporation*. Harvard University Press, 1956.
- Klein, B., and Leffler, K. "The Role of Market Forces in Assuring Contractual Performance." *Journal of Political Economy* 1981, 615-641.
- Masten, Scott E. and Snyder, Edward A. "United States Versus United Shoe Machinery Corporation: On the Merits." *Journal of Law & Economics* April 1993, 33-69.
- Mulligan, William H. "Mechanization and Work in the American Shoe Industry: Lynn, Massachusetts, 1852-1883." *The Journal of Economic History* March 1981, 59-63.
- Tirole, Jean. *The Theory of Industrial Organization*. Cambridge: MIT Press, 1988.
- Waldman, Michael. "Eliminating the Market for Secondhand Goods: An Alternative Explanation for Leasing." Mimeo, 1994.

APPENDIX A: EVIDENCE FROM AN ORIGINAL SOURCE

In the previous section entitled *Moral Hazard and Firm Size in Shipping Firms*, I make the claim that because of large communication costs, and heterogeneity in transactions, the captains of shipping vessels would not be paid employees, but instead residual claimants. This led me to predict small shipping firms, with few multiship firms.

Subsequent to making this prediction, I have read *Two Years Before The Mast*, by R. H. Dana, published initially in the early 1840's.³⁰³ Dana was an undergraduate at Harvard in the mid 1830's, and after deteriorating vision left him unable to study, he joined a merchant vessel as a crewman.³⁰⁴ With great detail he describes the ship's passage South around the dangerous Cape Horn, and North to (then) Mexican California. Dana's ship (and one he later transferred to) spent a year sailing up and down the California coast, purchasing and processing beef hides, which would be processed into leather on the ship's return to Boston.

The book's setting contained all of the elements used in my prediction of small firm size. Transit time between the ship's owners in Boston and the coast of California was a matter of

³⁰³ My copy was published in 1909 as part of the "*Harvard Classics*", a series with the subtitle "The Five Foot Shelf of Books", which is a true claim.

³⁰⁴ The title derives from the sleeping quarters of the crew, which was located before the mast. The ship's officers were lodged behind the mast.

months, and was subject to large random variation because of weather, pirates and mishap (both of the ships Dana served on eventually were eventually lost at sea). The trade itself also varied; Dana's ship found an unexpected decline in the number of hides supplied, which required a longer stay on the coast in order to fill the ship. The outcome of a several year voyage seemed subject to large random variation, and the captain seemed to be the party with the most control.

Given these facts, I fully expect the captain to be residual claimant. Yet my prediction did not (at first) appear to be correct. The ship was owned by a partnership, with senior and junior partners. This firm had at least three ships on the California coast, as well as one agent who oversaw the purchase (but not processing or transportation) of hides. The captains of these ships, and all others mentioned, were paid employees. All had worked their way up through the ranks, and none had ownership in the firm. The owners of the shipping firm had insurance, but there is no reference to restrictions imposed by the insurer.

A careful reading however does demonstrate that our theories of residual claimancy applied to this setting. What follows is an extended quote. Dana is, at the time, throwing hides off of a cliff onto the beach below.

“There I stood again, as six months before, throwing off the hides, and watching them, pitching and scaling, to the bottom, while the men, dwarfed by the distance, were walking to and fro on the beach, carrying the hides, as they picked them up, to the distant boats, upon the tops of their heads. Two or three boat-loads were sent off,

until, at last, all were thrown down, and the boats nearly loaded again; when we were delayed by a dozen or twenty hides which had lodged in the recesses of the hill, and which we could not reach by any missiles, and the general line of the side was exactly perpendicular. and these places were caved in, and could not be seen or reached from the top. *As hides are worth in Boston twelve and a half cents a pound, and the captain's commission was two per cent., he determined not to give them up;*³⁰⁵ and sent on board for a pair of top-gallant studding-sail halyards, and requested some one of the crew to go to the top, and come down by the halyards”.

No, the captain was not an owner of the ship, for it was unlikely that an individual with the wealth to own a significant share of a ship would also have a comparative advantage in operating the ship. For comparative advantage is determined by opportunity cost, and the opportunity cost for a wealthy man to spend several years of discomfort and misery sailing to the remote ends of the earth is quite large. The captains were instead individuals with low opportunity costs; yet in order to minimize losses from moral hazard, they were paid a share of the ship's profits, which decreased their incentive to shirk.

³⁰⁵Italics mine.

APPENDIX B: TECHNOLOGICAL INFORMATION RELATED TO CRUDE OIL AND REFINED PRODUCT

Crude oil is a mix of various chemicals, each with its own properties. The important chemicals, listed in ascending order of boiling point, are Butane, Pentane, Gasoline, Naphtha, Benzine, Kerosene, various heavy oils, and wax.

Kerosene was the most important product (by volume and sales), and was widely used for lighting. Proper kerosene has an ignition temperature of well over one hundred degrees, which means that it must be heated in order to burn. Common lamps allowed a cloth wick, which was lighted at the top, to descend into a reservoir of kerosene, which was absorbed by the wick. As the wick burned, it heated the absorbed kerosene to its burning point, and it was the ignition of the kerosene that produced light.

Kerosene could have three faults, all related to poor refining technique. First, when it was refined (see below) too much of the heavier waxes could remain in the fuel. This wax would both create smoke and soot when burned, and would also condense on the wick as the lighter kerosene was burned off. This in turn required wicks to be trimmed.

The next potential problem occurred when too much of the lighter chemicals remained. Since they both evaporated at room temperature and had a much lower ignition point, they

could condense in the lamp and then explode when the lamp was lighted. This was obviously a somewhat more serious problem.

Finally, crude also contained various sulfuric elements, and if they were not removed they would give burning kerosene a “rotten egg” smell. If they were left in lubricants, they would rust machinery.

The process of refining crude oil had two basic steps. First, the various chemicals were separated by distillation, then the resulting products were chemically treated to remove sulfuric compounds.

The distillation process was much like that of making “moonshine”; the crude was heated in a container with a large, coiled copper pipe exiting the top. As the temperature of the crude increased, the chemicals with the lowest boiling point evaporated first. These condensed in the copper tube, and returned to liquid form. By monitoring the temperature of the crude (and examining the condensing fluid), the refiner was able to determine which fluid was evaporating and condensing, and was therefore able to separate the various elements.

The process was somewhat complicated by two factors. First, some of the elements have ranges of boiling points, making a precise “cut” between components difficult. The second complication was that the heavier elements (oils and waxes) are composed of large organic elements, and if these were heated to a high enough temperature, they “cracked”, forming

additional kerosene. This process (cracking) was beneficial in that it increases the yield of kerosene, the product in highest demand. But the cracking process also tended to create and release additional sulfur elements, which made the next stage more difficult.

Once the crude was distilled, the sulfur had to be removed. This occurred through chemical treatment; various acids were added, which reacted with the sulfur and formed new chemicals that separated out from the kerosene. In some refining processes the end product was distilled once more to eliminate additional sulfur.

Refiners obviously wanted to maximize yield (the percentage of kerosene in crude). Yield could be increased by including both additional lighter elements, like gasoline, heavier elements, like paraffin oil to the kerosene. This obviously had the effect of reducing the quality and safety of the final product.

APPENDIX C: A CHRONOLOGY OF STANDARD OIL

1859. Oil Discovered in Pennsylvania.

1863. 26 year old John D. Rockefeller invests in Samuel Andrew's Cleveland Oil refinery.

1865. Rockefeller/Andrew's refinery has a daily rate of 500 barrels per day.

1870. Formation of the Standard Oil Company of Ohio, with 10% of U.S. refining capacity.

Standard is not legally able to hold stock in affiliated out of state companies, so stock is held by trustees.

1871. Formation, and failure, of the Southern Improvement Corporation. Purchase of all major Cleveland refineries by Standard Oil. Merger with New York Southern Improvement Corporation partners. Standard has 25% of U.S. refining market.

1874. Secret merger of Standard with its Southern Improvement Corporation partners. Standard has 40% of U.S. refining market.

1875. Completion of the Columbia Conduit Company pipeline to Pittsburgh, and the inclusion of the B&O/Conduit combination the railroad cartel.

1876-77. Empire Transportation Company Rate War.

1876-79. Standard bought or leased 108 independent refineries, which included most of the refiners in Pittsburgh, Philadelphia, New York and the Oil Region.

1879. Completion of the Tidewater pipeline to Williamsport, with a capacity to deliver six thousand barrels per day. Beginning of the Standard Oil Trust; all stock from all companies

is held by three minor officials. Standard is investigated by the Hepburn Committee of the New York State Assembly.

1880. Inclusion of the Tidewater in the Railroad Cartel.

1882. Reformation of the Standard Oil Trust; stock in 40 companies held by 41 investors is exchanged for trust certificates. Standard Trust is now the largest and richest manufacturing organization in the world. Shift to new headquarters in New York. Rationalization of Standard companies, and \$30 million dollar investment in pipelines (assets were \$3 million). Investment in national and international sales office, and centralized crude purchasing.

1884. Completion of Standard's own long distance pipeline network.

1885. Formation of Sun Oil in the Lima Field of Northwestern Ohio.

1888. Standard is investigated by both the New York Senate and the United States House of Representatives.

1889. Beginning of construction of Standard's Whiting, Indiana refinery.

1890. Beginning of Ohio vs. Standard Oil.

1892. Loss in Ohio vs. Standard Oil. Standard Oil of Ohio ordered to separate itself from the trust. The Standard Oil Trust enters state of perpetual dissolution.

1893. Completion of competing United States Pipe Line from the oil region to the coast.

1894. Beginning of State of Texas antitrust cases against Standard.

1895. Three largest independent refiners sell their firms and shares in U.S. Pipeline to Standard.

1897. Victory for Standard in first set of Texas antitrust cases. New Texas suit filed. J.D. Rockefeller withdraws from day to day management of Standard.

1899. Another Ohio Suit against SOHO for failure to dissolve. Change from the Standard Trust to a New Jersey holding company, Standard Oil of New Jersey.
1900. Dismissal of Ohio case by attorney general. Loss of Second Texas case. Dissolution and reincorporation of Waters-Pierce Oil Company (the Standard company in Texas). Release of damaging report by United States Industrial Commission.
1901. Discovery of oil in Texas and formation of Gulf Oil and Texaco.
1906. New Texas antitrust case against Standard (Waters-Pierce). Standard refines about 83% of United States crude oil, markets about 80% of refined products in U.S., and controls 85% of export business. First Federal suit against Standard Oil under the Sherman Act.
1907. New Texas antitrust case against other Standard Texas companies. First Federal suit suspended.
1909. Standard loses Texas antitrust case (Waters-Pierce). Waters-Pierce driven out of Texas. Other Standard companies also lose case. Their assets are purchased by another Texas Standard Oil company. Resumption of Federal Antitrust suit.
1910. Standard is now one of eight integrated oil companies among the U.S. two hundred largest industrial enterprises. The other seven are: The Texas Company (Texaco), Gulf Oil, Associated Oil, Union Oil of California, Shell Oil, Tide Water Oil, and Sun Oil.
1911. Standard loses appeal before the Supreme Court. Standard Oil of New Jersey order dissolved. No criminal charges were filed. Stock was distributed pro-rata, effectively only changing the form of the organization, not its ownership. The court divided Standard along functional lines, and only Standard Oil of New Jersey and Standard Oil of California remained fully integrated.

1917. Five of the sixteen major companies spun off of Standard were among the U.S. top 200 largest industrial companies. Eight of the new Standard companies became vertically integrated. All integrated companies began moving into crude production.

Sources: All legal references, especially those involving court cases, come from Bringhurst. Economic facts are found in either Granitz, or one of its sources, Chandler.

Only the dates of beginning and final conclusion of state level antitrust cases are included, with all intervening dates suppressed. Also suppressed are state level antitrust suits for states other than Texas and Ohio.

APPENDIX D: MERGER LAW

If today a firm the size of Standard Oil in 1870 tried to buy its local rivals, it is likely either the Justice Department's Antitrust Division or the Federal Trade Commission would challenge the merger/purchase in court. The standard these agencies use in evaluating a merger is to measure the change in industry concentration before and after the merger. If the concentration exceeds a certain level, or increases by more than a certain level, the merger will be challenged.

The measure of industry concentration used is called the Herfindahl-Hirschman Index ("HHI") of market concentration. The HHI is calculated by summing the squares of the individual market shares of all the firms in the market. For example, before Standard purchased its rivals in Cleveland, it was about as large as the next three largest rivals. Suppose³⁰⁶ 24 Cleveland³⁰⁷ refineries produced 100 units of kerosene in total, with Standard producing 30 units, the next three largest refineries producing 10 units each, and the remaining 20 refineries producing two units each. The market share of each firm would then be as follows:

³⁰⁶These numbers are for illustrative purposes only.

³⁰⁷This assumes that Cleveland is the relevant market; that crude oil producers would not sell their crude elsewhere if one party controlled all of the Cleveland refiners and lowered the

Standard:	Each of the Next Three:	Each of the Next Twenty:
30/100	10/100	2/100
=30%	=10%	=2%

In order to compute the HHI, square each of these (for each firm), and add them together:

$$(30)^2 + 3(10)^2 + 20(2)^2 = 900 + 300 + 80 = 1280$$

The Justice Department would report this as 1280, using percents instead of fractions.

After the merger, the Cleveland market would have one firm, with 100% of the market. So the HHI would be $100^2 = 10,000$. The HHI would have increased by 8720!

The current standards for mergers are, to quote the Justice Department:

“a) Post-Merger HHI Below 1000. The Agency regards markets in this region to be unconcentrated. Mergers resulting in unconcentrated markets are unlikely to have adverse competitive effects and ordinarily require no further analysis.

b) Post-Merger HHI Between 1000 and 1800. The Agency regards markets in this region to be moderately concentrated. Mergers producing an increase in the HHI of less than 100 points in moderately concentrated markets post-merger are unlikely to have adverse

price it paid to refiners. See chapter xx for a more complete analysis of the relevant market.

competitive consequences and ordinarily require no further analysis. Mergers producing an increase in the HHI of more than 100 points in moderately concentrated markets post-merger potentially raise significant competitive concerns depending on the factors set forth in Sections 2-5 of the Guidelines.

c) Post-Merger HHI Above 1800. The Agency regards markets in this region to be highly concentrated. Mergers producing an increase in the HHI of less than 50 points, even in highly concentrated markets post-merger, are unlikely to have adverse competitive consequences and ordinarily require no further analysis. Mergers producing an increase in the HHI of more than 50 points in highly concentrated markets post-merger potentially raise significant competitive concerns, depending on the factors set forth in Sections 2-5 of the Guidelines. Where the post-merger HHI exceeds 1800, it will be presumed that mergers producing an increase in the HHI of more than 100 points are likely to create or enhance market power or facilitate its exercise. The presumption may be overcome by a showing that factors set forth in Sections 2-5 of the Guidelines make it unlikely that the merger will create or enhance market power or facilitate its exercise, in light of market concentration and market shares.”

These are the standards by which the government will challenge mergers, but not necessarily the standards the courts will enforce. They also change periodically, both with additional understanding of oligopolies and political inclination of the current administration. Changes

in these standards may be found on the World Wide Web at <http://www.usdoj.gov>, which also includes a number of documents and reports on current antitrust actions.

APPENDIX E: THE DOMINANT FIRM MODEL

It is understandable why a firm would prefer to be a monopolist - it would then be able to set the monopoly price unconstrained by other sellers. But US Steel was a merger of either firms that were already monopolies (combinations in specialty areas of production) or for which rivals definitely existed. Why would US Steel be interested in having a large market share that was still below 100%? In general, why are antitrust authorities concerned about mergers that do not yield monopoly (sole seller) status?

In the case of Standard Oil, the allegation was that Rockefeller purchased enough industry capacity that he could drive competitors out of the market by charging a below cost price (predatory pricing). But this allegation did not appear in the case of US Steel; rival steel producers generally appreciated US Steel's pricing policy, which consisted of announcing prices at the beginning of the year, and maintaining them, even when undercut by rivals. What did Gary, Morgan and Associates hope to gain by forming this gigantic industrial enterprise?

The traditional answer is that owning a sufficiently large share of industrial capacity does in fact confer some temporary monopoly power. This answer is best understood via the *Dominant Firm Model*.

Suppose that for consumers steel is entirely homogeneous - they buy it from the seller with the lowest price.³⁰⁸ The market demand is strictly a function of price $d = d(p)$. Assume that there are many small sellers (the fringe), each of which acts as a price taker. And there is one large seller (the dominant firm) with a "significant" (to be defined later) share of the industry production capability.

Since the fringe is composed of price takers, the "fringe supply curve" is made up of each firm producing that quantity which equates marginal cost to the market price (I have not yet said anything about how this market price is determined). Geometrically, I now have two lines (functions) on my diagram, the industry demand curve (price taking consumers) and the fringe supply curve (price taking suppliers).

³⁰⁸ Conditions under which this is a reasonable assumption include no shipping costs (or everything net of shipping costs), a product that is homogeneous and easy to measure in quantity and quality, and complete production before the transaction is completed. These are the costless information assumptions of Neo-Classical Economics.

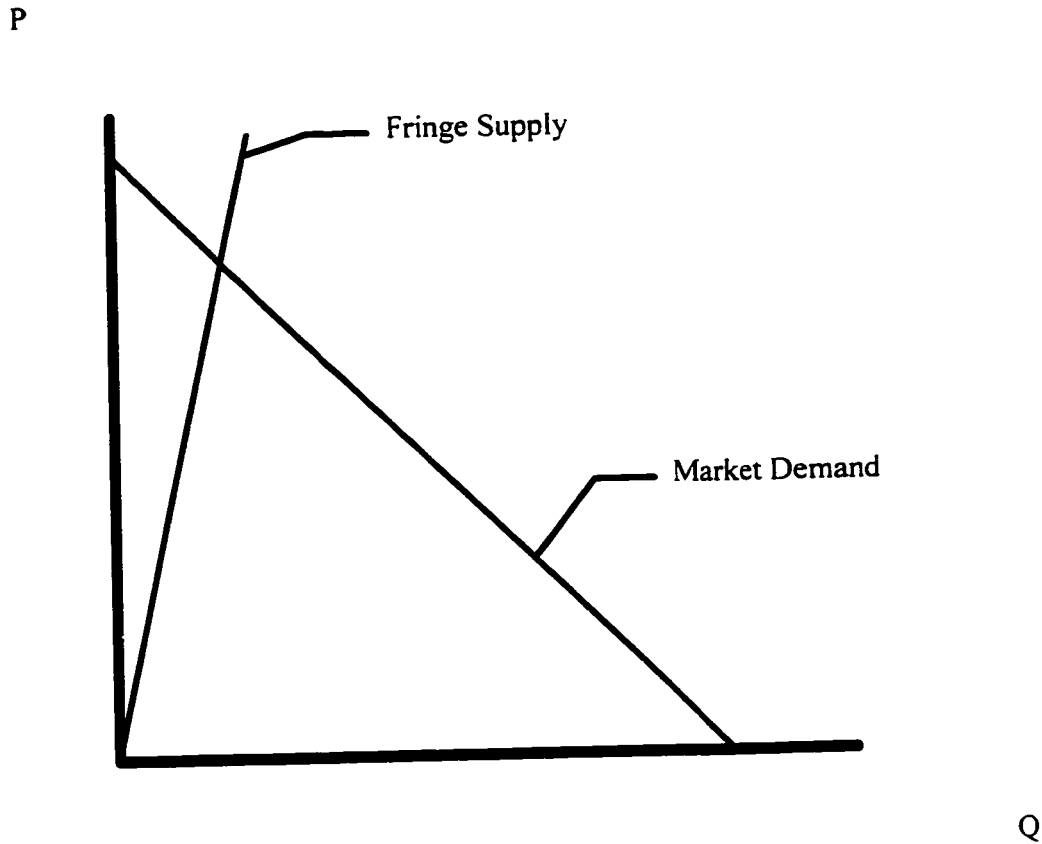
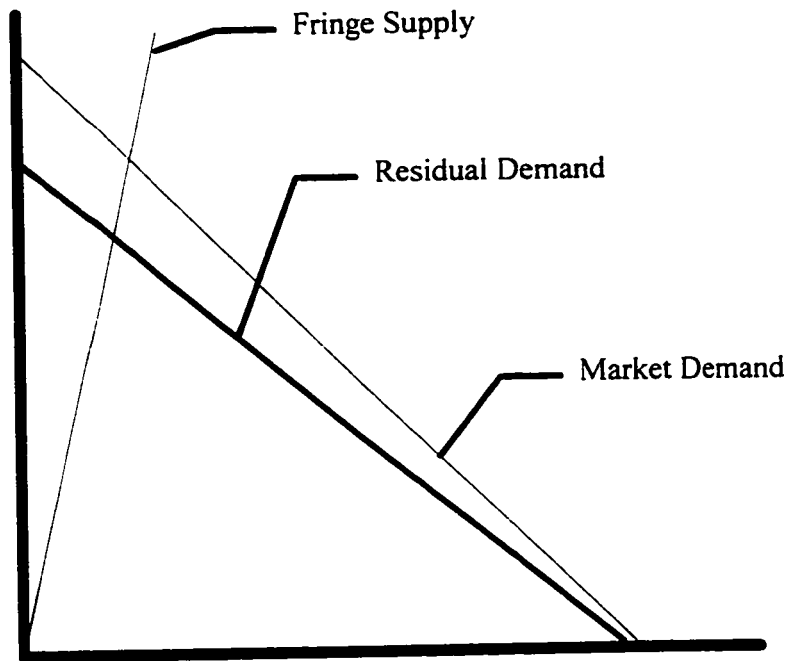


Figure 3: Demand and Fringe Supply

Given the market demand curve and fringe supply curve, the dominant firm faces a “residual demand”. For any price set by the dominant firm, there will be a corresponding quantity demanded. The fringe will supply the quantity for which marginal cost is less than the price, and the remainder will be supplied by the dominant firm. At the price in question, none of

the other suppliers will wish to either increase production or decrease price - this is what the fringe supply curve shows us. The dominant firm may vary the quantity it sells by varying its price, and is therefore a price searcher.

P



Q

Figure 4: Residual Demand

The dominant firm will act as a pure monopolist over this residual demand, setting marginal revenue (of the residual demand curve) equal to its marginal cost. The only change from the

standard Joan Robinson type of monopoly is that the residual demand curve is more elastic than the market demand curve.

Several obvious questions arise. First, what market share is required (or what is the tradeoff) for a firm to profitably act as a dominant firm instead of a mere oligopolist? Next, what will happen over time to the profits of a dominant firm?

These sorts of questions are best answered using our standard tools, particularly elasticity of demand and supply. Symbolically, the residual demand curve (demand facing the Dominant firm) is the difference between the market demand curve and that quantity supplied by the fringe. Or:

$$D_{df} = D_{mkt} - S$$

Just as portrayed diagrammatically, each of these is a quantity that is a function of price. We are interested in elasticity (rates of change), so we must take derivatives.

$$dD_{df}/dP = dD_{mkt}/dP - dS/dP$$

Multiplying each side by P/D_{df} .

$$(dD_{df}/dP) * (P/D_{df}) = (dD_{mkt}/dP) * (P/D_{df}) * (D_{mkt}/D_{mkt}) - (dS/dP) * (P/D_{df}) * (S/S)$$

$$E_{d-df} = E_{d-mkt}(D_{mkt}/D_{df}) - E_{s-f}(S/D_{df})$$

The dominant firm's market share (MS) is (D_{df} / D_{mkt}) . Since the quantity supplied by the fringe is equal to the quantity demanded by the entire market, less the quantity supplied by the dominant firm,³⁰⁹

$$E_{d-df} = E_{d-mkt}(1/MS) - E_{s-f}(1/MS - 1)$$

This gives us the elasticity of demand facing the dominant firm as a function of its market share, the elasticity of market demand, and the elasticity of the fringe supply. All of these, to varying degrees of success, may be investigated empirically.

But how does knowing the elasticity of demand facing the dominant firm help us? The Lerner Index relates the idea of how much a firm is able to raise its price above its marginal costs to the elasticity of demand it faces. It says:³¹⁰

³⁰⁹ The necessary substitutions are easy but not obvious - students should perform these steps themselves.

³¹⁰ The Lerner index may be derived as follows. A monopolist sets marginal revenue equal to marginal cost. Total Revenue = $P*Q$

$$(P-MC)/P = 1/E$$

Where P is the profit maximizing price. If demand is very elastic - for example if it approaches infinity because of competition, the firm is only able to charge a price that is equal to marginal cost (a price taker). As demand becomes less elastic, the firm is able to charge some price above marginal cost. For example a firm is able to price ten percent above marginal cost if the elasticity of residual demand is ten; if the elasticity of residual demand is four, the firm is able to set its price twenty five percent above its marginal cost.

Now let me combine these two ideas in order to answer our questions about market share. I will begin by making assumptions about the elasticity of market demand and fringe supply, and then see how changing market share results in different markups over marginal cost.

$$MR = dTR/dQ = (dP/dQ * Q) + P$$

Set this equal to Marginal cost

$$(dP/dQ * Q) + P = MC$$

Rearrange

$$P - MC = -(dP/dQ) * Q$$

$$(P - MC)/P = -(dP/dQ) * (Q/P) = 1/E$$

remembering that elasticity of demand is dQ/dP times P/Q .

If $E_{d-mkt} = -1$, $E_{s-0} = 1$

Then

$$E_{d-df} = 1 - 2/MS$$

If the dominant firm's market share is small - say 5%, then the elasticity of residual demand it faces is -39, and its markup over marginal cost is approximately 3%. If its market share is 50%, then its elasticity of residual demand is -3, and its markup is 1/3. With a market share of 75%, its elasticity of residual demand is 1.67, and its markup is 60%.

Of course all of these numerical examples are based on assumptions as to elasticity of market demand and fringe supply. If there are relatively good substitutes (elastic demand) or other suppliers are willing to substantially increase output based on a price increase, then the elasticity of residual demand becomes more elastic, and the Lerner index decreases. So what market share allows a firm to become "dominant"? To answer this question an investigator must first assess the elasticities of demand and fringe supply.

Other Implications

What will happen to profits over time? The dominant firm model is static - it does not explicitly model change over time. But we do know that both demand and supply are more elastic over longer periods of time. Information concerning substitutes generally is not free to consumers, but the cost of acquiring it decreases with time. Likewise it is more expensive to build a factory in a week than in a year.

Because of this, we expect the price charged by a dominant firm and its market share to decrease over time. Given that there are not any permanent limits to entry, a dominant firm's "monopoly power" will be eliminated by the entry of new rivals.

Other implications also follow from the model. The fringe suppliers act as price takers, and therefore keep producing until their marginal cost is equal the price set by the Dominant firm. They are uninterested in additional sales at the prevailing price. But since the dominant firm is a price searcher, its marginal cost is less than the market price. It would find additional sales profitable if they did not depress the market price.

One way to have these additional sales is by export to other markets. If the price in other markets remains above the dominant firm's marginal cost, it may be able to increase sales without lowering the domestic price by means of exports. The rival fringe sellers will not be able to export, since they are already producing every unit for which the marginal cost is less than the price.

VITA

Timothy Dittmer

University of Washington

1998

Home:
1541 NE 88th St.
Seattle, WA 98115
Tel.:(206) 522-7956
e-mail tdittmer@u.washington.edu
Fax: (206) 522-3342

EDUCATION

University of Washington, Seattle, WA
Ph.D., Economics, 1998
Fields: Industrial Organization
Public Finance

Wheaton College, Wheaton, IL
B.A. Cum Laude, Philosophy, 1988

DISSERTATION

A Property Rights Approach to Antitrust Analysis

PAPERS AND WORKS IN PROGRESS

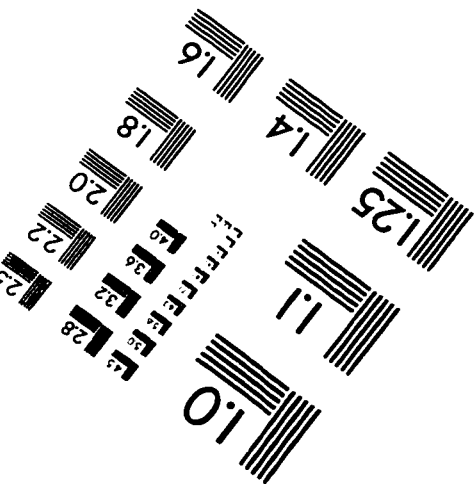
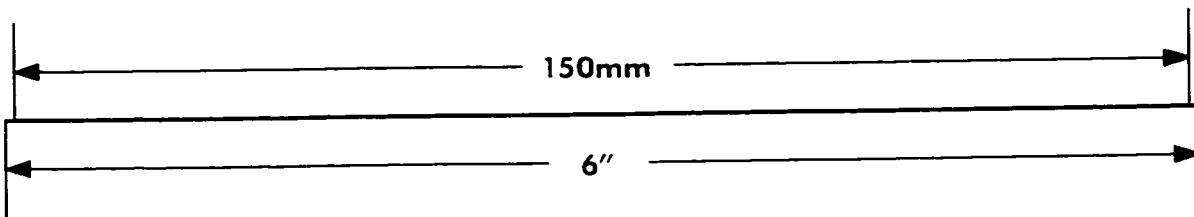
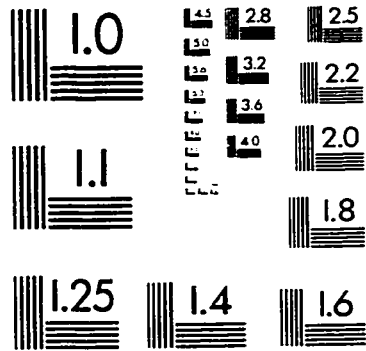
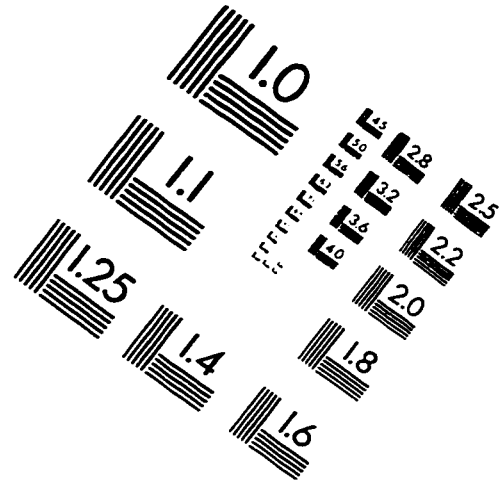
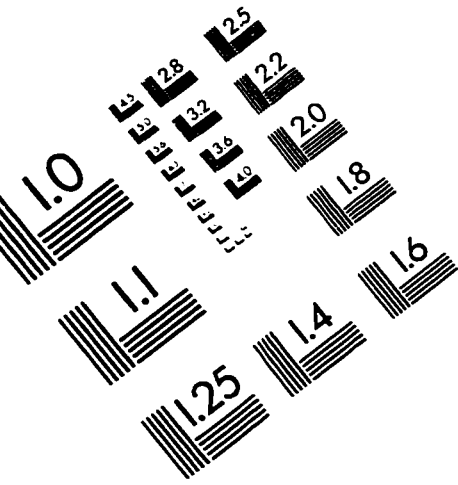
“Substitutes for the Sword: A Measurement Cost Explanation of Voting,” presented at the 1997 Western Economic Association International Conference.

“Rebel Oil: Why Opportunity Cost is the Correct Legal Standard” (with Keith Leffler), in progress.

“Illinois Brick and the Extension of the Direct Purchaser Rule to the Clayton Act” (with Keith Leffler), in progress.

“Reputation and Predation: An Analysis of Strategic Interaction in Retail Gasoline Markets” (with Keith Leffler), unpublished manuscript, August 1995

IMAGE EVALUATION TEST TARGET (QA-3)



APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989

© 1993, Applied Image, Inc., All Rights Reserved

