The Active and Passive Voice are Equally Comprehensible in Scientific Writing

by

Susan Rhodes

A thesis submitted in partial fulfillment
of the requirements for the Degree of

Doctor of Philosophy

University of Washington

1997

Approved by Philip S Dale
Chairperson of Supervisory Committee

Program Authorized
to Offer Degree PSYCHOLOGY

Date DEC. 15, 1997

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 also 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 12-17-97

University of Washington

Abstract

The Active and Passive Voice are Equally Comprehensible in Scientific Writing

by Susan Rhodes

Chairperson of the Supervisory Committee:
Associate Professor Philip S. Dale
Department of Psychology

The passive voice is typically thought to be wordy, impersonal, and difficult to read. Although early studies examining the passive voice appeared to show that passive voice sentences are harder to process and less preferred than active voice sentences, later studies pointed out a number of communication situations when passive sentences were not more difficult to process and were indeed preferred to their active counterparts.

The present investigation focused on the use of passive voice in scientific reports. Three studies were performed. Study 1 focused on the incidence of the use of passive voice verbs in empirical reports in fields of varying object-orientation (psychology, botany, and chemistry/physics) and across article report sections (Introduction, Method, Results, and Discussion) of varying objectivity. The incidence of passive voice verbs was greatest in the most object-oriented field (chemistry/physics), second highest in a somewhat less object-oriented field (botany), and lowest in the least object-oriented field (psychology). Similarly, the objective article sections (Methods and Results) had a higher percent of passive voice verbs than the interpretive article sections (Introduction and Discussion).

Study 2 was an empirical investigation designed to assess reading speed and comprehension for "high-active" vs. "high-passive" versions of scientific reports. Two experiments were conducted, each of which included two articles that varied solely in voice, the "high-active" version having 100% active voice verbs and the "high-passive" version having 42% passive voice verbs. Subjects read one high-active article

and one high-passive article. After reading each article, they answered 16 multiple-choice questions designed to test their understanding of the article. Large samples were used to ensure highly sensitive experiments. For both experiments, there were no significant differences in either reading speed or reading comprehension due to differences in voice.

Study 3 compared student lab papers rated by instructors as weak versus strong to determine which factors, including voice, might account for the differences in instructor ratings. The analysis revealed differences in content development, paper length, incidence of violations of American Psychological Association guidelines, organization, and degree of sentence cohesion, all in favor of strong papers. However, it did not reveal any significant differences in percentage of passive voice verbs used. Additionally, the pattern of usage in the lab reports was similar to that observed for the published articles examined in Study 1.

These results have several implications: (a) passive voice verbs are widely used in scientific writing (Study 1); (b) passive voice verbs are generally used appropriately, primarily for presenting objective information (Studies 1 and 3); (c) rhetorically-appropriate passive voice verbs do not elicit lower reading comprehension scores nor longer reading times than active voice verbs (Study 2), and (d) the incidence of passive voice verbs in student lab reports did not affect instructor ratings of paper quality (Study 3). This study demonstrates that passive voice need not create the kind of problems with which they are commonly associated. It also demonstrated the usefulness of evaluating linguistic devices such as the passive voice in stand-alone documents (as opposed to isolated sentences or short passages) and in specific rhetorical contexts (scientific reports).

Table of Contents

Acknowledgments

I would like to express my warmest appreciation for the assistance given to me by Norman A. Thompson, whose ideas are liberally reflected in the final version of this thesis.

Glossary

The glossary below defines frequently-used terms. Please note that the definitions are intended to be operational—to explain how the terms are used in this thesis.

Active vs. Passive Voice

Most English sentences use a SVO (subject-verb-object) structure. In **active voice** sentences, the subject corresponds to the actor and the object corresponds to the action receiver. In **passive voice** sentences, the subject corresponds to the action receiver and the actor is (optionally) included in a *by*-phrase as a sentence modifier:

subject (actor) — verb (action) — object (action receiver)
Red Adair capped these oil wells.

subject (action receiver) — verb (action) — prep. phrase (actor)
These oil wells were capped (by Red Adair).

Actor, Action, Action-receiver

Words can play three key semantic roles in a sentence: actor, action, and action-receiver. These roles are sometimes confusing because different authors use different terms to describe them. The **actor** is the intentional entity who causes the **action**, also known as the agent or logical subject. The **action receiver** is the entity that the action is directed toward; it is also known as recipient, acted-upon, or logical object (see example under *Active vs. Passive Voice*, above).

Reversible vs. Non-reversible Sentences

A sentence is **reversible** if the actor and action receiver can change positions without the production of an anomalous sentence:

Reversible pairs:
Day follows night.
Night follows day.
Night is followed by day.
Day is followed by night.

Non-reversible pairs:
People read books.
Books read people. (anomalous)
Books are read by people.
People are read by books. (anomalous)

Full vs. Truncated Passive Voice

Full (agentive) passives retain the optional *by*-phrase; **truncated** (agentless) passives do not.

These oil well were capped by Red Adair. (full)
These oil well were capped. (truncated)

Please note that the *by*-phrase must be agentive, i.e., it must specify the actor (agent) who does the action, not the instrument used to do the action, i.e., *These oil well were capped by hand* is not a full passive.

It is dangerous to write about good writing.

— Hebb & Bindra, *American Psychologist*, 1952

Introduction

> A style is a response to a situation. When you call a style bad, . . . you ought to make sure that you understand the situation it responds to. You may be objecting to the situation, not to the style invented to cope with it.
>
> — Richard Lanham, Style*: An Anti-Textbook*, p. 58

The passive voice has long been a controversial sentence construction. Linguists have varied opinions about its general merit as a rhetorical device and about when and how it should be used. For example, the venerable Strunk and White (1979) write that "the active voice is usually more direct and vigorous than the passive" (p.18). Flesch and Lass (1947) assert that "the passive voice makes sentences weak" (p. 171). However, not all stylists have adopted such a negative view. Kolln (1991) asserts that "the passive voice has a place in every kind of writing" (p. 96); Wilkinson (1992) asserts that "the passive voice is particularly suited to scientific writing" (p.322); Rodman (1981) and Vande Kopple (1989) also describe the advantages of using passive voice for describing scientific procedures.

The passive[1] is also a very old sentence construction. Its use has been documented in Proto-Indo-European, which is at least 6000 years old (Crystal, 1992). Its documented use in English dates back at least as early as 1388, when the Bishop Wyclif, the first translator of the Bible from Latin into English, defined it much as we would today (Murray, 1933). Jesperson (1974) cites numerous examples of passives in passages dating back several centuries and discusses its evolution over the last several centuries. Coetzee (1980) discusses the evolution of passive voice in scientific writing over the last three centuries. If the passive is such a poor rhetorical choice, why did the passive originally develop and why does its use persist today? If it were without much merit as a linguistic construction, wouldn't it likely have passed into disuse by now?[2]

It was questions like these that spurred my interest in the passive voice,

---

[1] The term *passive* used as either a noun (*the passive)* or an adjective (*passive* constructions)in this paper always refers to the passive voice, which is not to be confused with the word *passive* as lacking in dynamism, liveliness, or motivation.

[2] I am not the first to ask this question; linguist Julia Stanley (1975) raises the same question, citing other linguists (e.g., Lakoff, 1971, and Palmer, 1968) who have done the same.

especially as it is used in scientific writing. In the last 30 years, textbooks and style manuals have increasingly discouraged the use of the passive voice, even scientific textbooks and style manuals (e.g, the widely-used *Publication Manual of the American Psychological Association*, 1994). Conway (1981) notes that the authors of twelve textbooks he surveyed disparaged the use of passive voice verbs. Of the 47 textbooks and style manuals that Warren (1981) included in his annotated bibliography of sources that discuss passive voice, 42 (89%) contained negative comments about the passive voice, characterizing it variously as vague, wordy, imprecise, impersonal, awkward, deadening, weak, confusing, or boring. More recently published texts often contain similar advice (see Study 1 for an example).

However, despite the plethora of advice to avoid writing in the passive, this advice rests upon rather shaky empirical foundations. Although short, isolated, narrative active voice sentences (e.g., *Jack threw the ball)* were shown three decades ago to require less effort to process than their passive counterparts (*The ball was thrown by Jack)*, it does not follow—nor has it been empirically demonstrated—that passive sentences that are longer, contextually-appropriate, and non-narrative are either processed more slowly or understood less well than active sentences.[3] In fact, a number of subsequent studies have shown that contextually-appropriate passive verbs[4] are processed as easily as active verbs.

However, as the literature review demonstrates, there are many empirical "holes" in the body of research on the passive voice. While some characteristics of the passive (e.g., transformational properties) have been thoroughly explored, others (e.g., the way that passives are used in context) have received less attention. In particular, little research has been done to determine the effects of voice on reading speed or comprehension in entire documents. Nor has decades-old research been re-examined in light of modern research methods or theoretical perspectives.

Despite these gaps in the research, the general public—and much of the academic community—has acquired the impression that passives cause problems for

---

[3] See the next chapter for a detailed discussion.

[4] In the interests of simplicity, I often refer to active voice verbs as *active verbs* and passive voice verbs simply as passive verbs. I also refer to passive voice verbs as *passives*.

readers and should therefore be avoided when possible. This anti-passive view is also evident in advice given by key arbiters of style (e.g., style manual editors and textbook authors) who influence writing conventions adopted in academic fields.

I became aware of this curious gap between theory and practice regarding the passive voice since completing my Master's degree in technical communication in 1986. I began teaching a writing course that attracted engineers and scientists. I discovered that many of these writers were puzzled about the reasons they were told by various sources—books, editors, or bosses— to avoid passives in scientific writing. I began to wonder myself after reading a number of crystal-clear scientific reports written in the passive voice.

My first goal in this thesis was to systematically document the history of empirical research into the passive voice. As far as I can determine, the literature review presented here constitutes the only recent review article on passive voice research.

My second goal was to empirically investigate the effect of passive verbs in the kind of documents where we would most expect to see them, i.e., scientific research reports. When I began this project, I wondered whether passive constructions in scientific prose were really problematic. My skepticism stemmed from the dearth of definitive research on this topic. As Rubens (1982) observed, "the grammatical structure of scientific prose has only been haphazardly studied" (p. 75). I found that not much had changed since Rubens made this observation sixteen years ago. Few empirical studies focused on sentence structure within scientific articles; fewer still specifically examined voice in scientific articles. Most of the studies I did locate (e.g., Riley, 1991) were relatively limited in scope or focused only peripherally on passive constructions (e.g., Paul & Rosner, 1983). Thus, it was clear that an in-depth investigation of the passive in scientific prose could fill a rather sizable gap in the research literature.

In the literature review (see next chapter), I summarize passive voice research conducted from the 1960's to the present. My main goal was to draw together a rather disparate collection of studies in an attempt to discern patterns that would permit the formulation of generalizations about the effects of voice in various contexts. In the

ensuing three chapters, I present the results of three separate studies designed to examine the use of passive voice in scientific reports. Study 1 focused on the incidence of passive constructions in published scientific articles in three fields (psychology, botany, and chemistry/physics) and in different article sections (Introduction, Method, Results, and Discussion). Study 2 assessed the effects of voice on (a) how quickly subjects read scientific articles and (b) how well they understood the articles. Study 3 focused on differences in weak vs. strong student lab reports written for psychology laboratory courses; voice was one of the variables assessed. In the final chapter, I discuss the implications of these studies and directions for further inquiry. I also speculate about the reasons why passives have become so criticized in the absence of compelling evidence that they inhibit either reading comprehension or reading speed.

Literature Review

When tracing a line of research on a topic, it is usually possible to find a relatively cohesive body of literature whose authors heavily cite one another and build on one another's research. However, the body of research on the passive voice is atypical in a number of ways. First, it is old—most of the studies reviewed here were performed during the 1960's and 70's. Second, much of it is fragmented—with the exception of early studies on transformational grammar (see below), authors performing closely-related studies do not generally cite one another, even when the studies were done at practically the same time. Third, there are no recent review articles summarizing the research. As a result, although some questions on voice have been thoroughly researched, there are also many areas that have barely been touched, e.g., the effects of passive constructions on reading comprehension. It is likely that some of these gaps in the research would not exist if the gaps were better documented.

The purpose of this literature review is to present a detailed account of the research on the voice of verbs. The subjects in existing studies were either adults or adults and children; the literature review does not include studies that used only children (e.g., Turner & Rommelveit, 1968) or special populations as their subjects. I discuss most of the studies I located, although I excluded studies that offered incomplete descriptions or that focused on voice only peripherally; I also excluded several studies that focused exclusively on the effects of voice in speech (e.g., Wannemacher, 1974, 1976).

Generally, studies are presented chronologically, in an attempt to trace the thread of the research from the early 1960's to the present. At the end of the literature review is a summary of major findings, limitations, and unanswered questions.

Overview

Studies on the passive voice during the 1960's were syntactically motivated—most focused on determining the psychological validity of Chomsky's transformational grammar. But as interest in transformational grammar waned, studies on the passive voice became more broadly focused, although most studies focused the role of voice as a semantic, rather than syntactic, variable. By the early 1980's, studies on voice

practically disappeared, perhaps in reaction to the strong backlash against Chomskean linguistics and those variables most associated with it. In any event, I was unable to locate any relevant empirical studies published after 1982.

Voice as a Syntactic Variable in Transformation Grammar Research

In 1957, Noam Chomsky published *Syntactic Structures*, which set forth a revolutionary new theory of language: transformational-generative grammar. Central to his initial theory was the notion that human language begins with deep structure mental representations (active, simple, declarative kernel sentences, p. 80) which are psychologically converted into the complex surface structures by means of grammatical transformations, e.g., declarative-to-interrogative transformations, positive-to-negative transformations or active-to-passive transformations.

Psychologists interested in the new cognitive perspective in psychology viewed transformational grammar as a theory which complemented the emerging paradigm that human behavior is cognitively motivated, i.e., that it is more than just the product of behavioral conditioning. The proposition that complex communication starts with simple mental representations was not only intuitively appealing, but testable. If simple, active, declarative sentences were more psychologically basic than complex, passive, non-declarative sentences, then the former should be processed with less effort than the latter.

George Miller (1962) was one of the first psychologists who attempted to test Chomsky's theory. He hypothesized that people would take longer to perform a complicated transformation than a simple one. Subjects matched two columns of sentences, one column of "kernel" sentences to another column of sentences that were transformed in some way (negative, passive, or negative-passive transformations). Miller tallied the mean number of sentences matched in one minute. Examining kernel sentences such as *Joe liked the small boy*, Miller reported the smallest difference ($\bar{x}$ = 1.1) for kernel-to-negative matches (*Joe didn't like the small boy)*, the next smallest difference ($\bar{x}$ = 1.5) for kernel-to-passive matches (*The small boy was liked by Joe)* and the greatest difference ($\bar{x}$ = 2.7) for kernel-to-negative passive matches (*The small boy wasn't liked by Joe*), although he did not give standards deviations for these means

or state that these differences were statistically significant.[5]

Miller's results inspired a large number of studies designed to determine whether the basic tenet of Chomsky's theory—that kernel sentences are psychologically most basic—was true. Researchers also wanted to ascertain whether the processing load imposed by transforming kernel sentences into non-kernel sentences was additive or multiplicative.

Sentence voice (active or passive) was a popular manipulation in these experiments because making an active construction passive was thought to add one additional grammatical transformation to a sentence while holding the meaning of the sentence constant. Subjects were typically presented with either active voice sentences (e.g., *John threw the ball*) or their passive counterparts (*The ball was thrown by John*), either alone or in combination with other transformations. Ease of processing was assessed using a variety of dependent measures: accuracy in the (verbatim) free recall of active versus passive sentences (Mehler, 1963); time to insert verb-object-subject sets into active versus passive sentence forms (Tannenbaum, Evans, and Williams (1964); time to verify the accuracy of pictures matched to active versus passive sentences (a sentence verification task) (Gough, 1965, 1966; Slobin, 1966); and animacy and uncertainty of words used to fill in actor-verb-object blanks in active versus passive sentence forms (Clark, 1965).

Initial studies using isolated sentences as stimuli showed that affirmative sentences were processed more quickly than negative sentences and that active sentences were processed more quickly than passive sentences (notably, Miller, 1962; Miller and McKean, 1964; Tannenbaum, Evans, and Williams, 1964; Gough, 1965), even when sentence length was controlled for by using truncated passives (Gough,

---

[5] It should be noted that this experiment was part of a speech and is described very briefly. Because no significance tests are given, the results are merely suggestive rather than conclusive. Miller himself appears to understand the preliminary nature of this study, remarking that "before we spend too much effort answering . . .[the questions raised by the results], we had better make sure that the data are correct" (p. 759). The study is reported here only because of its tremendous impact on subsequent developments in theoretical and experimental psycholinguistics.

1966).[6] These results were interpreted as offering at least tentative support for the transformational decoding hypothesis—the idea that people understand syntactically complex sentences by breaking them down into simple, declarative "kernels." However, as studies evaluating the transformational model continued, it became evident that these initial explanations were incomplete.

One line of studies compared the effects of voice and structural complexity on sentence recall. Martin and Roberts (1966) compared the role of three variables as predictors of verbatim sentence recall: voice (active, passive, and truncated passive), negation, and Yngve depth (a measure of structural complexity —see Yngve, 1960, for details). They found that Yngve depth (sentence complexity) was the best predictor of verbatim recall, in that simple sentences were recalled better than complex sentences, regardless of sentence voice. Thus, they concluded that "when sentence complexity and sentence length are controlled, the role of sentence kind [voice] becomes marginal" (p. 216). Comparing the effects of Yngve depth and voice on retention of individual sentence words over intervals of 10, 20, 30, and 40 seconds,[7] Martin, Roberts and Collins (1968) found that, although steady decrements in retention were observed as the retention intervals increased, (a) low-Yngve-depth passive sentences were recalled best at all increments and (b) recall for words in passive sentences was also less variable than for the other three types of sentences. When Perfetti (1969) varied Yngve depth and voice using a verbatim sentence recall task, he found no main effect for Yngve depth. But he did find an interaction between Yngve depth and voice, such that high Yngve-depth sentences in the active voice and low Yngve-depth sentences in the passive voice were better recalled than low Yngve-depth passives and high Yngve-depth actives. This finding led the author to conclude that "there is no simple and direct relationship between transformational complexity and psychological complexity"

---

[6] A full passive retains the actor in a *by*-phrase (e.g.,***The wall was painted by the boys***); a truncated passive does not (***The wall was painted***). Some researchers (e.g., Danks & Sorce, 1973) have defined truncated passives differently, designating passives with agentive *by*-phrases (***The wall was painted by the boys***) as full (non-truncated) and those with instrumental *by*-phrases (***The wall was painted by a brush***) as "truncated." Unless otherwise noted, however, I will use *truncated* to mean passives without *by*-phrases.

[7] Subjects counted backward by threes during the retention intervals to prevent rehearsal.

(p. 103).

Wearing (1970) assessed recognition for sentences varying in voice, Yngve depth, and a new variable: predictability.[8] Stimulus sentences that were identical or changed were presented either immediately or after 48 hours; subjects judged them as new or old. Wearing found that (a) recall for all sentences decreased over 48 hours, (b) unpredictable sentences were recalled better than predictable ones and (c) lowYngve-depth (simple) sentences were generally recalled better than high Yngve-depth (complex) sentences. Voice in isolation did not affect recall although it interacted with Yngve depth in the same way observed by Perfetti (1969). The authors of all four studies observed that Chomsky's theory did not account for their results because active sentences were not always recalled better than passive sentences. In fact, three of the four studies found that simple passive sentences were recalled better than simple active sentences. Sentence complexity, not transformational complexity, most influenced sentence recall.

Two of the preceding studies examined the recall of sentences over time. Martin, Roberts and Collins found that sentence recall decreased in as little as 10 seconds after presentation; Wearing (1970) found generally poor sentence retention after a 48-hour retention interval. Sachs (1967) found similar results, but her study is particularly notable because it compared the importance of syntactic versus semantic factors for sentences in connected discourse. Subjects who read 28 short passages[9] later heard a probe sentence that was either identical to the sentence in the passage or that varied either syntactically (i.e., changed word order or voice) or semantically (i.e., changed meaning). Subjects judged the probe sentence as either identical or changed, and the changed sentences as either changed in form or in meaning. The independent variable was the number of intervening syllables (0, 80, or 160) between the presentation of the target sentence and the probe. Subjects who heard the probe

---

[8] Predictability was operationalized as the likelihood of encountering the situation depicted in the sentence in the real world. Thus, *The weary traveller was soon annoyed by the long delay* is predictable; *The ugly boss was grandly entertained by the poor cleaner with Swiss liqueur* is unpredictable (the examples derived from Wearing, 1970).

[9] Passage length was characterized in terms of total syllables per passage; passages ranged between 27 and 180 syllables long.

immediately after the target sentence were able to accurately recognize the nature of all changes, but they did not retain this ability for syntactic changes after 80 and 160 syllables.[10] Sachs' study was the most definitive of several studies that had demonstrated, over time (i.e., after subjects had read some number of additional syllables), ideas are not retained as short, declarative kernels.[11]

Other researchers examined the effects of sentence reversibility. Slobin (1966) found that, for reversible sentences,[12] affirmative active voice sentences were verified as true or false (i.e., as matching or not matching pictures depicting the events in the sentence) more quickly than affirmative passive voice sentences. Affirmative passive voice sentences were in turn verified more quickly than negative active voice sentences. This finding was not consistent with the idea that active voice sentences are always verified more quickly than passive voice sentences. Furthermore, non-reversible passives were verified as quickly as non-reversible actives ($\bar{x}$ = 1.39 vs. 1.45 sec.), both of which were verified more quickly than non-reversible negative sentences that were either passive or active ($\bar{x}$ = 1.81 vs. 1.83 sec.). Thus, Slobin observed that "the difficulty in understanding passive sentences is partly attributable to the problem of

---

[10] Johnson-Laird and Stevenson (1970) replicated these results, finding that subjects alerted to the fact they would receive a recall test tended to recall sentence syntax; those not alerted performed more poorly. But both groups showed good retention for semantic content. However, in a later experiment, Sachs (1974) assessed both auditory and visual sentence memory at intervals of 0, 20, 40, and 80 seconds. She found that, while memory for formal or lexical changes was lost very quickly (by 20 syllables), changes in voice in the visual condition were recognized at better than chance levels, perhaps because voice switches involve a great deal more re-arrangement of visual stimuli than lexical or semantic alterations.

[11] Later studies found similar results. Anderson (1974) saw excellent verbatim recall of active versus passive sentences in a short passage only when probe sentences were presented directly after the stimulus sentences; Perfetti and Goldman (1975) found that verbatim recall after 48 hours was only 12% and meaning-preserving recall was only 25%. Only one study that varied voice (Wearing, 1972) found no significant difference in sentence retention after a significant (48 hour) delay, although Wearing's operationalization of "retention"—number of words recalled, regardless of function—may have demanded less effort than either verbatim sentence recall or retention of sentence meaning.

[12] Herriot (1969) defines a reversible sentence as "one in which no semantic rules would be broken if LO [the logical object] were to become LS [the logical subject] and vice-versa" (p. 166). Thus, *The boy chased the girl* is reversible because its inverse—*The girl chased the boy*—is a semantically acceptable sentence. However, a sentence like *The girl is watering the flowers* is not reversible because its inverse—*The flowers watered the girl*—is not semantically acceptable. (See also Glossary, p. 1.)

keeping track of which noun is the actor" (p. 226), a problem that does not exist for non-reversible sentences. Herriot (1969) found that sentences that violate pragmatic expectations (e.g., *The bather rescued the lifeguard*) were more difficult to process than those that do not (e.g., *The bather was rescued by the lifeguard*), regardless of voice. In a second experiment, where both unreversed and reversed sentences were pragmatically appropriate (*The tiger attacked the panther* versus *The panther was attacked by the tiger*), response times were faster for active voice versions. Such results suggest that passives elicit slower reaction times only for reversible sentences, an effect not predicted by transformational grammar.[13]

Researchers also examined the effects of passive truncation. Gough (1966) found that active sentences (e.g., *The girl hit the boy*) were verified faster than truncated passives (*The boy was hit*). But as Gough himself observed, "a speaker of English normally uses the active voice to describe events like those used in these experiments" (p. 495). Martin and Roberts (1966) found that subjects recalled the same percentage of simple passive and simple active sentences[14] (55%), but they recalled a higher percentage of passives, whether full or truncated, than actives (FP = 50%, TP = 41%, A = 36%).[15] Slobin (1968) examined the effects of truncation by asking subjects to re-tell stories originally presented using a mix of active, full passive, and truncated passive sentences.[16] Although only 36% of the full passive sentences remained passive during retelling, 80% of the truncated passives remained

---

[13] Later experiments elicited similar results for visually-presented sentences. Glucksberg, Trabasso, and Wald (1973) found that, although passive sentences were verified less quickly than active sentences, this effect was far stronger for reversible than non-reversible sentences. Moore and Biederman (1979) found no difference due to voice on detection times for anomalous (mostly non-reversible) sentences.

[14] It is unclear whether the authors used active and passive versions of the same sentences because they do not provide matching examples. The example they provide for a simple passive sentence is a negative truncated sentence (*They were not prepared for rainy weather*).

[15] However, because the authors were primarily interested in sentence complexity rather than sentence voice, they did not indicate whether these differences were statistically significant.

[16] One story example is as follows: *On the first day of school Bob was introduced to his new teacher (by the principal), and was given a reading book (by the teacher). He was shown around the classroom (by the teacher), and all of his friends were happy to see him. When he came home he was asked (by his father) to tell all about school, and he said it was fun.*

passive. Transformational theory would predict a strong tendency to convert all passives to actives, not just full passives. Thus, Slobin concluded that "there is no firm basis for asserting that change of a truncated passive to a corresponding active constitutes a 'simplification' in any psychological sense" (p. 880). Such results, noted the author, argue for the proposition that truncated passives are not simple transformations of active voice mental representations.[17,18,19]

Wright (1969) looked at the effects of matching the voice of sentences with questions about those sentences. She asked subjects questions about sentences five seconds after sentence presentation, varying the voice of both sentence and questions. She found that subjects made fewer errors when the sentence and question voice matched than when they did not match; furthermore, the fewest errors were made in the passive sentence-passive question condition. Transformational theory would predict the most errors in the passive sentence-passive question condition due to the greater

---

[17] When Wilson (1979) re-analyzed Slobin's data, he found that full passives were converted to actives when doing so would place an animate actor at the sentence beginning (e.g., *The jewels were cherished by the artist*). Passives in which this was not the case (e.g., *The artist was delighted by the jewels*) were not converted to actives, thus demonstrating that a preference for animate sentence topics may be a key motivator in voice switches; the relationship between voice and animacy is discussed later in the literature review.

[18] Linguistics such as Dagut (1985), Freidin (1975) and Coetzee (1980) have made similar suggestions. Freidin observes that truncated passives tend to function more like adjectives than verbs in sentences such as *The door was locked*. Coetzee goes further, suggesting that the *by*-phrase "should be thought of as added rather than deleted" (p. 208). He argues that (a) while there are languages that have only "truncated" passives, there are no languages that have only full passives; (b) full passives seem to have developed later in history than "truncated" passives, and (c) children learn full passives later than "truncated" passives.

[19] In a later study, Danks and Sorce (1973) looked specifically at the effects of passive truncation and the effects of sentence imagery on sentence recall. The authors varied truncation by including a *by*-phrase for all sentences, but varying the role played by the *by*-phrase, which was either agentive (a full passive) or instrumental (which the authors designated as a truncated passive). Please note that this definition is somewhat idiosyncratic in that most authors define a truncated passive as one in which the *by*-phrase is simply deleted. However, using this approach permits sentences to be matched on length. Example sentences from each condition include these: high-imagery, full-passive (*The grades were issues by the professor*); high-imagery, agent-replaced (*The grades were issued by letter*), low-imagery, full passive (*The game was played by substitutes*); and low-imagery, agent-replaced (*the game was played by permission*). High-imagery sentences were better recalled than low-imagery sentences but full passives were better recalled than truncated passives only for low imagery sentences, suggesting that the nature of the *by*-phrase was a less salient factor than image value.

number of transformations necessary to process two passive constructions. When Layton and Simpson (1975) replicated Wright's (1969) study, they obtained the same results when sentences were presented in groups of one or two sentences (light memory load). But for groups of four and eight sentences (heavier memory load), passive questions elicited more errors for both active and passive questions, leading Layton and Simpson to conclude that the match-mismatch effect holds only under light memory load conditions.[20]

Clark and Clark (1968) argued that changing the voice of a sentence changes not only its syntax but its meaning. Although they did not manipulate voice in this study, they did manipulate semantics independently of syntax by presenting subjects with sentences with information in or out of temporal order (e.g., *He tooted the horn before he swiped the cabbage* versus *He tooted the horn after he swiped the cabbage*). They found that subjects were better able to recall the meaning of sentences in temporal order than vice-versa; however, transformational complexity did not affect ease of recall. Concluding that Chomsky's theory cannot account for these results, Clark and Clark proposed an alternative semantic theory of memory in which semantic markedness—rather than syntactic "tags"—play a prominent role in helping people to retail information in memory. Prentice's (1966) finding that semantically priming the first (as opposed to the second) noun of a sentence resulted in better verbatim sentence recall for both actives and passives provided early support for this idea.[21]

The results of studies described in this section of the literature review call into question the idea that active sentences are more psychologically basic than passive sentences. Many showed no consistent advantage for actives over passives, and those that did tended to use sentences that were (a) reversible, (b) idiomatic only in the active voice and (c) transformed to full rather than truncated passives.

---

[20] The sentences were apparently unrelated, so the results are not applicable to the reading of connected ideas. Also, the examples provided were *The doctor helped the nurse* and its transformation *The nurse was helped by the doctor*. Thus, the results are generalizable to reversible sentences and full passives.

[21] Verbatim recall was better for actives than passives, a result primarily due to the tendency of subjects to convert passives to actives. This result is unsurprising, for all the passives were full rather than truncated (see discussion of full versus truncated passives later in this chapter).

It became apparent that the elegant transformational theory set forth in *Syntactic Structures* did not account for the results seen in study after study. As interest in transformational grammar waned,[22] interest in voice as a syntactic variable also declined.

Voice as a Semantic and Pragmatic Variable in Sentence Processing Research

Voice did remain of interest to psycholinguists as a *semantic* variable during the next decade, primarily for assessing how sentences are encoded, stored, and retrieved. Clark and Clark were not the only researchers interested in determining whether the role of voice was purely syntactic. In 1968, Philip Johnson-Laird asked the question "Does the passive mean the same thing as the active in English?" (1968a, p. 69). He maintained that it does not—that the passive is utilized in normal English usage to shift the emphasis from the actor to the action receiver.[23] Subjects asked to pictorially represent sentences such as *Red follows blue* or *Red is followed by blue* tended to emphasize subjects more (by making larger drawings of the color in the subject position), a tendency that was more pronounced when the sentence was passive. In a study using similar stimuli, Johnson-Laird (1968b) found that subjects asked to match sentences to red and blue strips used both voice and sentence position in deciding which sentences best described strips that varied in color emphasis. The fact that people perceived and produced passives differently than actives supported Johnson-Laird's hypothesis that differences in sentence voice reflect differences in sentence meaning.

A number of other studies in the next few years supported the idea that people use shifts in voice to indicate shifts in idea emphasis. These studies showed that people generally use active sentences to focus on actors and passive sentences to focus on action receivers (e.g., Hornby, 1972; Olson & Filby, 1972; Tannenbaum & Williams, 1968; Trabasso, 1972; Wearing, 1972). Tannenbaum and Williams (1968) showed junior high students "cue" passages that varied in voice (active or passive) and

---

[22] See Harris (1978) for a brief history of transformational grammar research.

[23] Johnson-Laird uses the terms *logical subject* for *actor* and *logical object* for *action receiver;* I use *actor* and *action receiver* in the interests of clearly distinguishing syntactic roles from semantic roles.

conceptual focus (neutral, actor, or action receiver).[24] These passages were followed by line drawings related to the "cue" passage (e.g., a picture of a train hitting a car) which the students were instructed to describe in a sentence as quickly as possible.[25] Although neutral and subject-focused passages elicited significantly longer response times from students in the passive condition, object-focused passages did not, thus demonstrating that context can affect the ease of encoding propositions as active or passive sentences. Olson and Filby (1972) used a similar approach, except that they showed subjects line drawings emphasizing either the actor (e.g., a train hitting a car) or the action receiver (e.g., a car being hit by a train) as cues for stimulus sentences which subjects verified as true or false. Reaction times were fastest when the picture focus and surface structure matched: active sentences were verified faster after actor-focused drawings and passive sentences were verified faster after action-receiver focused drawings.

Hornby (1972) obtained similar results in a task where subjects were shown a picture depicting the actor or the action-receiver and asked to select between an active and passive sentence as matching the picture: people viewing the picture depicting the actor were more likely to select the active sentence while those viewing the action-receiver were more likely to select the passive sentence. Hornby introduced a novel theoretical framework for interpreting his results. Pointing out the inadequacy of syntactically-oriented theories such as Chomsky's for interpreting the effects of semantic context on prose comprehension, Hornby proposed using the "topic-comment" approach used by Prague School linguists, a theory that focuses on the functional (psychological) roles of "actor" and "action-receiver" (also known as "agent" and "recipient") in the sentence instead of the traditional syntactic roles of "subject" and "object."[26]

---

[24] Conceptual focus was manipulated by changing the passage name and topic.

[25] Voice was a between-subjects variable, so the students responded using either active or passive sentences, but not both.

[26] See Sampson (1980) for a general discussion of the Prague School, Lyons (1968) for a general introduction to topic-comment analysis, and Grieve and Wales (1973) for a discussion of the relationship between topic-comment analysis and voice.

Trabasso (1972) also obtained results supporting the idea that subject (actor) focus is emphasized in active sentences and that object (action-receiver) focus is emphasized in passive sentences. He reported the results of an unpublished study done with Gough in 1963-1964 in which the researchers assessed how information was encoded in memory by asking subjects to learn concepts about a sentence[27] with the sex of either the subject or the object of the sentence as the salient feature. The sentences were presented in either the active or the passive voice. Errors in concept identification were fewest for active sentences when the subject was the salient feature and for passive sentences when the object was the salient feature. Trabasso observes that these results parallel those of Wright (1969), in which matches between sentence and question voice produced the fewest comprehension errors.

Wearing (1972) tested sentence memory for active and passive sentences using several types of cues, including subject and object cues. He found that, although more words from passive sentences were recalled than from active sentences, this effect was due entirely to the strong effect of the object cue, a result that is consistent with those described above.

These studies on the relationship between sentence voice and sentence emphasis demonstrate that the active voice does not mean the same thing as the passive voice. They illustrate the role voice can play to show readers or listeners what is most important about the sentence, the actor or the action receiver. When the actor was most important, active sentences served as a better cues; when the action receiver was most important, passive sentences served as better cues. The most frequent explanation for such a result is that people tend to assume that the noun at the beginning of a sentence is in some sense more primary—more important or more emphasized—than a noun later in the sentence. Grieve and Wales have commented that "[using] the passive indicates that 'importance' attaches to that entity which occupies initial, rather than subsequent, nominal position in the sentence, thereby serving a useful communicative function" (p. 174). Clark and Clark (1977) observed

---

[27] Probably a sentence based on one of these configurations: *A (boy or girl) is (kicking or hitting) a (boy or girl).* This configuration was used in a related experiment described more fully in the same article.

that "actives and passives fit different circumstances, and, when used appropriately, may be equally easy to understand" (p. 106).

Grieve and Wales (1973) also observed that, although voice is one tool communicators can use to focus attention on different sentence nouns, it is not the only tool for doing so. They examined the effects of word position and voice on the way that people use articles (*a* and *the*) in sentences. In their first experiment, they manipulated sentence voice and the definiteness of the articles preceding subjects and objects to produce stimulus sentences such as *The woman saw the man, The woman saw a man, The man was seen by the woman,* and *The man was seen by a woman.*[28] They asked subjects to formulate a question for which the stimulus sentence would be the answer to determine what readers perceived the sentence to be about. They found that subjects tended to use questions asking about the noun preceded by the definite article *the*, presumably because it implies a previous context in which the same noun had been discussed. In passive sentences where both nouns were preceded by the definite article (e.g., *The man was seen by the woman*), most questions focused on the man, rather than the woman or the event;[29] similar results were observed in a second experiment using truncated passives as stimuli. These results support the idea that article definiteness provides the strongest cue of noun importance, but that sentence position (as manipulated by voice) provides secondary cues for determining noun importance.

Hupet and Le Bouedec (1975) obtained similar results in an experiment where subjects were asked to complete French sentences presented in the following format: *I thought that the policemen had been injured by a gangster, but I was mistaken, in fact* . . . . Subjects tended to indicate that the mistake involved the indefinitely marked (i.e., the less certain) noun, regardless of sentence voice. When both nouns were preceded by the same type of article (i.e., *a-a* or *the-the*), questions focused on the

---

[28] Half the sentences were reversible (*The woman saw the man*) and half were not (*The porter hailed the taxi*), although the authors did not discuss the effects, if any, of reversibility.

[29] An example of a question focusing on the event is *What happened?*

event, not the subjects or objects.[30]

Wright and Glucksberg (1976) assessed article definiteness and sentence focus by varying both sentence voice and sentence reversibility. They allowed subjects to select either *a* or *the* before the second noun in the sentence (e.g., *The fox watched **a/the** hen*). They found that subjects preferred the definite article for active reversible sentences *(The fox watched **the** hen*) and the indefinite article for passive reversible sentences *(The hen was watched by **a** fox)*, a result consistent with the idea that passive is often used to emphasize sentence topics because using the indefinite article for the second noun de-emphasizes its importance. However, this pattern of results did not hold for non-reversible sentences, where subjects showed no consistent preference for either article in non-reversible sentences, regardless of voice *(The boy ate **a/the** cookie; The cookie was eaten by **a/the** boy)*. This result is consistent with the idea discussed earlier that non-reversible sentences are less ambiguous (less difficult to process) than reversible sentences, which leaves subjects freer to select either article for the second noun. However, as the authors point out, one limitation of this study was the use of isolated sentences as cues to indicate which ideas are to be highlighted: "The notion of topic, of theme, is. . . .stretched to its limits when the sentences being used do not have a prior context to generate a theme" (p. 569).

The relationship between noun animacy and voice in emphasizing concepts within a sentence has also been explored. In active sentences, the subject and actor are the same, and actors tend to be animate; in passive sentences, the subject is the action receiver, which may be either animate (*The girl was kicked*) or inanimate (*The papers were shredded*); the animate actor is present only in full passives (*The papers were shredded by the lawyers*).

Examining the role played by both animacy and certainty of nouns in determining how words are placed in sentences, Clark (1965) found that high school students asked to insert nouns into subject and objects slots in active and passive voice sentences inserted a higher percentage of animate actors in both active and passive sentences, but that this preference was stronger for active voice sentences (82% for

---

[30] Grieve and Wales (1973) also obtained this effect for *a-a* pairs, but not for *the-the* pairs.

actives vs. 68% for passives); in both active sentences, students preferred inanimate action receivers, but the percentage of animate action receivers was almost twice as great in passive sentences (27% vs. 46%). These results imply that although people generally like to make actors animate, they also like to place animate nouns at sentence beginnings, presumably because animate nouns are more likely candidates as sentence topics, and sentence topics are generally announced at sentence beginnings. This interpretation is consistent with the findings of Jarvella and Sinnott (1972), who also found that people were more likely to place animate nouns in pre-verb than in post-verb positions, regardless of sentence voice. Johnson (1967) obtained results similar to Clark's when asking people to rate the animacy of nonsense words in subject and object slots in active and passive sentences. For active sentences, nonsense words in subject slots were rated as more animate than those in object slots. Although the same pattern prevailed for passive sentences, the magnitude of the ratings differences were less, a result that also parallels that of Clark (1965).

Using a picture-sentence matching task, Harris (1978) found that children and adults asked to describe what was happening in a picture selected the active voice when the actor was more animate than the action receiver, and the passive voice when the action receiver was more animate than the actor, a result that Harris interpreted as reflecting the tendency to perceive animate nouns as more salient, and to place salient information at sentence beginnings: "It follows that in cases where the acted-upon [action receiver] is more animate than the actor, the passive voice may be used to realise this preference" (p. 496). Examining reasons why passives tended to be recalled as actives, Wilson (1979) varied the animacy of actors and action receivers in active and passive sentences and counted the number of voice transformations made during a verbatim sentence recall task. Experiencer-verb-object sentences tended to be transformed from passive to active voice (e.g., *The jewels were cherished by the artist* became *The artists cherished the jewels)*; object-verb-experiencer sentences tended to be transformed from active to passive voice (*The jewels excited the artist* became *The artist was excited by the jewels*).

Other investigators focused on a variable closely related to animacy: noun imagery. Clark and Card (1969) posited that, in isolated sentences, theme is equated

with the word in the subject slot. However, James (1972) suspected that semantic factors, such as noun imagery, played a prominent role in determining noun saliency and hence theme. He found that high-imagery subjects and objects were better recalled than their low-imagery counterparts, regardless of voice, a finding that appears to show that syntactic position is less salient than imagery as a determinant of sentence theme.[31] James, Thompson, and Baldwin (1973) found that sentences with low-imagery subjects or high-imagery objects (e.g., *The incident bothered the priest* or *The incident was reported by the students*) were the likeliest candidates for voice transformations, such that high-imagery nouns were moved to sentence beginnings. This result supports the notion that information made salient is perceived as thematic and thus placed in the subject slot if it is not already there.[32]

Studies of the processing of sentences in isolation such as those just reviewed support the idea that context affects voice preference, but the contextual cues provided consisted only of single words or sentences. The most extensive studies on the effects of context on sentence recollection were performed by Perfetti and Goldman (1974, 1975), who examined the effects of agent- or recipient-focus of a variety of contextual cues—prompt words, passage title, passage theme, and focus of preceding clause—on the way subjects recalled the final sentence in a short passage. Because of their complexity, these studies are be described in detail.

Perfetti and Goldman (1974) compared the effectiveness of cues presented for isolated sentences with that of cues presented for sentences occurring at the end of short (100+ word) passages. When they presented subjects with isolated sentences followed by single-word prompts consisting of a single word (the subject or object), they found that the probability that participants would be able to produce meaning-preserving paraphrases of the sentence was .36, regardless of prompt type. In a second

---

[31] However, as investigators such as Grieve and Wales (1973) have observed, nouns perceived as important are often placed in the subject slot because, in connected discourse, it is the subject that usually contains old information (information from the previous sentence). Thus, the subject often provides the link between two sentences, and thus helps maintain a coherent theme (see Haviland and Clark, 1974, for a discussion).

[32] But Wilson (1979) notes that the high-imagery nouns examined were animate, and that it is actually animacy rather than imagery that accounts for these results (see his discussion on pp. 283-284).

experiment, they introduced the theme variable by asking subjects to read a paragraph whose theme focused on either the subject or the object of the final sentence. When they presented participants with subject or object prompts from the final sentence, they found that probability of correct recall increased for both prompt types. However, the likelihood of correct recall for the subject-oriented passage after the object-prompt was lower (.48) than that after the subject-prompt (.65). But for the object-focused passage, the probability of correct recall was almost identical (.65 and .61). Thus, while providing a thematized passage appeared to assist the prompted recall of the target sentence (the final sentence of the passage), it also appeared to constrain the effectiveness of the object prompt, a result which the authors interpreted to reflect the normal practice (i.e., the practice in active voice sentences) of placing words relevant to the passage theme in the subject slot of the sentence).

In a 1975 follow-up study, Perfetti and Goldman, looked more closely at the effect of theme and a related variable, sentence topicalization, on the probability of generating a meaning-preserving paraphrase of the final sentence of a passage. While theme is determined by measuring the number of propositions contained in a passage and thus determines what that passage is about, the concept of "topic" refers to the focus of a particular sentence within a passage, and in English sentences—where word order is strongly constrained, it occurs at the beginning of the sentence, while the "comment"—a description of something about the topic—occurs at the end. For example, in the sentence *The admiral captured the bandit, The admiral* is the topic and *captured by bandit* is the comment.

Perfetti and Goldman manipulated the topic-comment focus of retrieval prompts by varying the voice of the final sentence of a short passage (*The admiral captured the bandit* versus *The bandit was captured by the admiral*). Subjects were asked to recall the final sentence of a passage after seeing a prompt that was either (a) thematized or not thematized; (b) an agent or recipient in the test sentence; or (c) topicalized or not topicalized in the test sentence (i.e., the test sentence was active or passive, respectively). Although agent prompts elicited better recall than recipient prompts and thematized prompts elicited better than non-thematized prompts, topicalized prompts had no effect on recall, although a nearly-significant interaction

($p < .06$) indicated that words (e.g., *bandit*) were better prompts for a passive sentence where they were the topic (e.g., *The bandit was captured by the admiral*) than for an active sentence where they were the comment (*The admiral captured the bandit*). This finding paralleled the results of the 1974 study in that placing the recipient at the beginning of the sentence (i.e., making the recipient the subject of a passive voice sentence) appeared to provide it with a special focus that enhances its value as a recall cue.

In a separate analysis comparing the voice of the passage with the voice used in sentences elicited by the prompts, Perfetti and Goldman noted that subjects exhibited a general tendency to recode passive sentences into active sentences, a tendency that was more pronounced when the prompt was the agent-comment (the second noun of the sentence) rather than the recipient-topic (the first noun of the sentence).

In a second experiment, passage thematization was made more salient by the addition of a title (e.g., *The Bandit* or *The Admiral*). With the addition of a thematized title, 50% of the recalled sentences were passive. In this experiment, recipient themes produced 47% passives compared with 32% passives for passages with an agent theme. The highest incidence of passive production (53%) occurred in response to a recipient prompt that was both titled and thematized, and where the first noun (the topic) of the preceding clause referred to the recipient. The lowest incidence of passive production (28%) occurred in response to an agent-focused prompt that was not titled nor thematized, and where the first noun of the preceding clause referred to the agent. Thus, these results show that, although passive sentences tend to be produced less often than active sentences, they can play a critical role in maintaining the thematic integrity of a passage. When the necessity of maintaining the thematic integrity of a passage required the use of the passive voice, subjects in this study preferred it over the active voice. Similar results were obtained in a French study by Parot and Kail (1980), where passive sentences were better recalled when the action receiver was thematized in a previous passage.

Kerr, Butler, Maykuth, and Delis (1982) also examined how passage theme affected the recollection of target sentences in short, narrative passages. The target sentence, which was pivotal to the meaning of the paragraph, had either the same theme

as the passage or a different theme.[33] Subjects listened to each passage and then re-wrote it based on their recollection.[34] Kerr et al. found that passive sentences were changed more often to active sentences than vice-versa, although passive targets still elicited more passive responses than active targets, regardless of whether the passage was subject- or object-focused.[35] But, as the authors note, the impact of voice may be less when passages are heard than when they are read, as previous research by Sachs (1974) suggests. When Kerr et al. assessed recognition memory for sentences presented either on tape or in writing, while changes in meaning were recognized at better than chance levels for all conditions, changes in voice were recognized at better than chance levels only for written—not oral—passages. These results replicates Sachs' findings (1967, 1974) that people retain semantic information much better than syntactic information, especially when asked to recall sentences they have heard rather than seen.

Voice as a Variable in Text Comprehension Research

As the preceding discussion shows, voice has been extensively investigated as a syntactic variable in transformational grammar research and as a semantic variable in sentence processing research. However, there is a surprising dearth of research examining the effects of voice on reading speed and comprehension, especially in connected discourse. I could locate only three studies primarily aimed at examining the effects of voice on prose comprehension: Charrow & Charrow, 1979; Coleman, 1964, 1965; and Siegel & Burkett,1974. But although these studies provide some insight on

---

[33] For example, in a story about two characters, one of the characters was thematized, and the target sentence had either the thematized or non-thematized person in the subject slot, and was in the active or the passive voice. However, it was not clear from the authors' description whether the position of the target sentence was held constant; in the sample passage, it was slightly more than half-way through a 126-word passage.

[34] Kerr et al. state their goal as the testing of "sentence memory," although it can be argued that the process used when recalling and writing down a passage is more reconstructive than memorial (see, e.g., James, Thompson, and Baldwin, 1973, for a discussion).

[35] If all the passages were like the example passage—i.e., about two persons—then the strong tendency to convert passives to actives may reflect the preference often seen for casting animate actors as subjects in active sentences.

the effects of voice on comprehension, they each have a number of shortcomings. Coleman's studies are not adequately described; Siegel and Burkett's study used a methodology ill-suited for assessing comprehension; and Charrow and Charrow's study looked at short passages, not entire documents. However, because these three studies constitute the corpus of research about voice and prose comprehension, they are described in detail below.

E.B. Coleman is the only investigator who has measured the effects of voice on the comprehension of lengthy written passages (although only one of his studies looks beyond the sentence level). Coleman was interested in the effects of transformations involving both voice and nominalizations and performed a series of experiments in the mid-1960s to see whether converting passive to active verbs and nominalized nouns to their verb counterparts could increase the comprehensibility of passages. Coleman and Blumenfeld (1963) used a cloze procedure to demonstrate that de-nominalized sentences were better recalled than nominalized sentences.[36] Coleman (1964) examined the effect of voice, nominalizations, and adjectivallizations on reading comprehension in connected discourse. He selected "two difficult passages" (each 2,969 words in length) from articles on group processes to simplify by converting passive verbs to active verbs, abstract nouns (nominalizations) to verbs and adjectivalized modifiers to verbs. Subjects were given twelve minutes to read each passage, one of which was in its original form and one of which was simplified as indicated above. Comprehension was assessed using 18 multiple choice questions and one matching question for each passage (subjects only answered questions on the part of passages they had read during the time allotted). Although subjects read almost exactly the same number of words for original and simplified versions (2,160 and 2,169), subjects who read the original versions received a mean score of 4.29 while those who read the simplified versions received a mean score of 5.38 (after the means were corrected for guessing), a finding which led Coleman to conclude that "anyone interested in improving readability would be heartened by the magnitude of the improvement: . . . 25.2%" (p. 187).

---

[36] Coleman (1964) gives the following examples of nominalized and de-nominalized prose, respectively: *His explanation of the design....* vs. *He explained the design.*

But this experiment has a number of weaknesses. First, because three different variables were simultaneously manipulated, it is impossible to isolate the effects of each variable in isolation or to pinpoint the nature of any possible interactions. Although Coleman used follow-up experiments (described later) to look more closely at each of the variables in isolation, he looked at their effects only in isolated sentences, not in passages or documents.

Second, the lack of verbatim examples of original and simplified passages (or even individual sentences) makes it impossible to independently evaluate how the sentences were altered.[37]

Third, the method for excerpting each of the articles (to create stimulus materials of equal length) was not given. Thus, it is not possible to determine which parts of each article were retained and which were discarded.

Fourth, almost no information about the nature of the multiple-choice questions used to assess comprehension was provided, e.g., the number of alternatives for each question and the form of the matching question. It is also unclear why Coleman used two different types of questions to assess comprehension.

Fifth, although means for the two conditions—original and simplified versions—are provided, no supporting statistics are given to indicate whether the differences are significant. This omission is remarkable, given that Coleman did provide statistical support when reporting a less important finding regarding the number of questions answered on the original versus the simplified versions. Sixth, other relevant analyses are omitted (e.g., an analysis of differences by article and interactions between article and condition). Such omissions in reporting give the experiment the appearance of a pilot study instead of a full-fledged research report. As a result of these weaknesses, the results cannot really be cited as

These problems are unfortunate, because this experiment is really the only one which had as its main goal the assessment how verb voice affects the comprehension of technical prose. But because of the above weaknesses, this study provides very limited empirical support for the proposition that active voice is easier to read than passive

---

[37] He does provide several examples of original and revised sentences, but these do not represent the text as a reader encounters it—as an entire text.

voice.

Due to the rarity of studies on voice and text comprehension, I thought it might be illuminating to perform my own analysis of the two articles originally analyzed by Coleman. One was an article on group interaction processes by Bales and the other article was on perceptions of social power by Lippitt, Polansky, Redl, and Rosen. Both were from a well-known collection of articles on group dynamics (Cartwright & Zander, 1953). Tallying the number of passive constructions in each of the articles, I found a surprisingly low number of passive verbs in each article—only 17% for *Lippitt et al.* and 18% for *Bales*, which is approximately half the number of passive verbs commonly seen in scientific writing.[38] However, I judged the passive usage in *Lippitt et al.*on the dynamics of power to be much more effective than in *Bales* on group dynamics. For example, consider the passive constructions in just the first paragraphs of each article:

Bales:
The ultimate stuff or empirical phenomena which the social scientist can observe, record, interpret, and arrange in many ways **may be thought of** under two heads: (a) action or interaction, i.e., the overt behavior of concrete human individuals, and (b) situation. Those things to which **action is addressed**, such as the self, other individuals, and physical objects, **may be said** to comprise the concrete situation of action for the acting individual. All of our relevant empirical generalizations must refer sooner or later to some aspect(s) of concrete action(s) or the situation(s) of action(s). This is true, whether the generalizations **are made** about personality, social system, or culture. Generalizations about any of these three types of systems or structures are at least one step removed (by abstraction) from the more complex and ultimate stuff we can all observe: activity addressed to persons and things. The observations of social interaction and its situation is the common starting ground for all of the social sciences.

Lippitt et al.:
This is one in a series of reports on a program of research into the process of social influence in groups of children (5, 8, 9, 10). Our initial curiosity focused on the phenomenon of behavioral contagion, described and clinically conceptualized by Redl (11) in an analysis of some of the operational problems of group therapy. We defined *behavioral contagion* as the spontaneous pickup or imitation by other children of a behavior initiated by one member of the group where the initiator did not display any intention of getting the others to do what he did. **This is distinguished** from *direct influence*, in which the actor initiates behavior which has the manifest objective of affecting the behavior of another member of the group. We decided to study the

---

[38] The results of several studies (e.g., Riley, 1991; Svartvik, 1966; Study 1 in this dissertation) indicate that approximately one-third of the verbs used in scientific articles are in the passive voice, although there is a great deal of variability in passive usage, both by individual authors and by field; see next chapter for a discussion.

hypothesis that the initiation of, and receptivity to, such social influences was related to the position of the actor in the social structure of the group.

Bales uses four passives, none of which are really necessary either to focus attention on a specific concept or to enhance sentence cohesion and all of which are somewhat awkward. In addition, *Bales* is abstract, pretentious, and propositionally dense. Thus, it is plausible that converting passives to active verbs (as well as nominalizations to verbs) in *Bales* might enhance text comprehensibility. But it is also possible that converting the awkward, unnecessary passives to rhetorically appropriate passives would have had a similarly beneficial effect (e.g., converting the first sentence as follows: *Human activity may be classified in two ways, as actions and situations).*

The first paragraph from *Lippitt et al.* contains only one passive construction, which is used to create a cohesive link to the previous sentence.[39] This paragraph appears to be much more cohesive, concrete, and direct than that from *Bales*; and its lower propositional density makes it less difficult to process. I was so curious about the comparability of these two paragraphs—paragraphs which closely mirror the style of the articles from which they are drawn—that I asked 14 undergraduates taking a psychology laboratory course in cognitive psychology to rate the general difficulty, concreteness, and concision of both paragraphs. The students rated *Lippitt et al.* more favorably on all three dimensions; 10 of the 14 also liked it better than *Bales*.

Coleman, however, characterized both articles as "difficult," an adjective which is at odds with the student's evaluations of the opening paragraphs and my own evaluation of both the opening paragraphs and remainder of the articles. It is not clear to me why converting passive to active verbs in *Lippitt et al.* would result in a 25% increase in reading comprehension scores, especially since most of the passives observed occur in the Methods and Results sections, and are used appropriately—to focus on the findings instead of the researchers, e.g.: *These sociometric interviews were conducted during the first and last weeks of the four week camp session* (p. 465); *The results of our two studies are summarized in four sections* (p. 467); or *A more refined analysis was made in the second study* (p. 471).

---

[39] The reason why *Lippitt et al.* has as many passive constructions as *Bales* is that the former article presents many empirical results, many of which are cast in the passive voice.

Based on my analysis of these two articles, the improvements in comprehension scores appear to be solely the result of revising *Bales*, not *Lippitt et al.* It is also possible that Coleman's stimulus materials may have been selected in such a way as to create experimental passages that are particularly benefitted by the revisions made. But the results were not reported in enough detail to ascertain whether either of these possibilities was true.

Coleman (1964) also asked subjects to read and write down eight short paragraphs (86-117 words in length; four simplified and four unsimplified[40]). Subjects displayed significantly better recall for simplified passages using three of four scoring systems, although the difference in recall was small (e.g., 43.4 versus 45.8 verbatim content words recalled). The same study contained two additional experiments focusing solely on the effects of nominalizations on comprehension. In his discussion, Coleman focuses very heavily on nominalizations, not voice. Thus, it is possible that the manipulations of text in both the two articles and the paragraphs primarily involved nominalizations; voice may have been only an incidental manipulation that accounted for little or no differences in comprehensibility.

Coleman (1965) again focused primarily on nominalizations, although one of the three experiments discussed examined the effects of voice on verbatim sentence recall of sentences. In this experiment, subjects viewed 96 sentences in sets of six (three active and three passive for each set) and subsequently wrote them down. Sentences were randomly selected from books in the college library; half were active sentences with transitive verbs and half were passives. For passive truncated sentences, actors were supplied to convert truncated passives to full passives. Recall was measured for verbatim sentences, content words in any order, and content words in correct order. Active sentences were better recalled than passive sentences using all three scoring systems, leading Coleman to conclude that "the implications of the data are obvious enough that little discussion is necessary" (p. 340).

I disagree with this conclusion. This study, like the 1964 study, had significant flaws, many of which stem from the way that sentences were selected for inclusion as

---

[40] Simplified passages were denominalized and converted from passive to active voice.

stimuli. First, no examples of the sentences used as stimuli were provided, so it is not possible to determine what these stimulus sentences looked like—whether the sentence topics were abstract or concrete; whether the actors inserted into truncated sentences were proper nouns, common nouns, or pronouns; or whether multi-clause sentences were passed over or whether only one of the clauses was used as a stimulus. It is impossible to know whether each sentence worked equally well as an active or a passive, especially truncated passives converted to full passives.

Second, converting truncated passive sentences to full passives may have resulted in more awkward passive sentences, which would create an unfair advantage for active voice sentences. This could constitute a serious problem, given the fact that 39 of the 48 (81%) of the passive sentences were originally truncated.

Third, the modifiers were deleted from all the sentences, thus resulting in creating the kind of "stripped down" sentences that are seldom seen in real prose. Such sentences may be useful for testing the speed of mental processing, but they are not suitable stimuli to use if the goal is to provide an author with advice about what kind of sentences will facilitate reading comprehension.

Fourth, although Coleman went to the trouble of randomly selecting sentences so that he could generalize his results to a larger population of sentences, making alterations on these sentences effectively negated the randomness of the sampling procedure.

Fifth, Coleman did not document the method by which he made these alterations, and this omission is not trivial. It is not easy to devise a sample of sentences using the approach Coleman describes. Consider the following sentence, which I randomly selected from Coleman's discussion section: *In 52 of the detransformations, one long clause was changed into two short coordinate clauses.* If I drop all of the modifiers, the sentence reads as follows: *The clause was changed into clauses.* However, the resulting sentence is so stripped down that it no longer makes any sense unless we, at a minimum, add back *two* so that the sentence reads *The clause was changed into two clauses.* Even so, this sentence is not very content rich when presented out of its natural context. One wonders whether Coleman would have retained such a sentence for his study. If he did, the next step would be to convert the

truncated passive sentence a full passive—and then into an active voice sentence. To do so, it is necessary to select which word should be used to indicate the actor. In this case, it was Coleman who was acting, so the most accurate rendering would be *The clause was changed into two clauses by me* (or in the active voice, *I changed the clause into two clauses*). But it might be tempting—so as to avoid too many sentences using first person pronouns—to convert the first person into the third person by using one of several possible actors: *The clause was changed into two clauses (by him / by the author / by Coleman)*. If we do this, do we select the pronoun, the common noun, or the proper name?

Coleman must have encountered such problems, but does not describe either the problems or how he solved them. Without such a description, it is impossible to meaningfully interpret his results. There are other problems, as well. Although Coleman provides statistical support for his general conclusion that actives were better retained than passives, he does not analyze mean differences by type of content word (actor, action, or action receiver) recalled. However, he does graph these results for the dependent measure showing the smallest, but still significant, difference between recall for actives and passives: content words retained in order (16.67% for actives and 14.67% for passives). The values of the graphed means for actives and passives can be approximated as follows: for actor recall, 17.6 % versus 15.3 %; for action recall, 15.3% versus 13.3%; for action receiver 16.7% versus 16.0%. The differences for actor and action recall approximate those between the (significant) grand mean difference of 2.0%. However, the difference between correct responses for the action receivers is only .7%—a difference which may not be statistically significant. If it is not, this result indicates that actor and action were more salient and thus better recalled in active sentences but that action receivers were equally salient in active and passive sentences. If Coleman had provided a more detailed description of this stimuli and a more complete statistical analysis of his data, he might not have characterized the implications of his findings as obvious.

Siegel and Burkett (1974) were also interested in the effect of voice on text comprehension. Their goal was to develop methods for evaluating the comprehensibility of Air Force training materials. They looked at a wide range of

variables, including Yngve sentence depth, morpheme depth, syllable length, and structural complexity (as measured by transformational complexity, the use of relative pronouns, self-embeddedness, and branchedness). Summarizing previous research results, Siegel and Burkett state that it appears that comprehensibility can be increased by decreasing word depth, decreasing morpheme depth, changing reversible passive sentences to active sentences, avoiding center embedding, and avoiding right-branching sentences. However, the authors wanted to independently assess the effects of each variable for themselves.

Because their object was to develop guidelines for Air Force training materials, it would have been useful to assess the effects of voice using materials similar to those they wanted to improve. Unfortunately, they employed a sentence verification task to do so, despite the fact that it is difficult to use the data from such a study to make conclusions about text comprehension. They provided subjects with 30 stimulus sentences: 10 active, 10 passive, and 10 negative-passive. Examples of stimulus sentences include these: *They found the money lying in the corner, The money was found lying the in the corner,* and *The money was not found lying in the corner.* Interestingly, the passive and passive-negative sentences were all truncated, although the authors did not discuss (and may not been not familiar with) the ramifications of using truncated as opposed to full passives. They also did not discuss the ramifications of using a pronoun, rather than a noun, as the subject of the sentence.[41] They gave subjects booklets containing sentences followed by pictures which either matched or did not match the sentence. The subjects, Air Force trainees, verified the sentence as true or false; their response times were recorded.

Contrary to the author's expectations, there was no difference in processing time for active, passive, or passive-negative sentences (for both true and false sentences) although weaker readers did experience more difficulty with true sentences that were both negative and passive, a finding in accord with previous studies showing that true negative sentences are generally harder to process than false positive sentences (e.g., Gough, 1965). They conclude that "when writing, to ensure . . .

---

[41] The example provided used a pronoun, but the authors did not indicate whether all the stimulus sentences did so.

comprehensibility, the use of the passive-negative voice should be avoided, . . . especially when writing for those with low reading grade level" (p. 105). However, if the authors had been more familiar with the literature on sentence voice, they might have designed their study differently or reached different conclusions about their findings. Using truncated passives likely decreased the difficulty of processing the passive sentences; using personal pronouns probably decreased the saliency of the subject in the active sentences. The combined effect of these two practices would be to create the impression that active and passive sentences are equally easy to process, whereas it is probably more accurate to state that, when sentences are presented in isolation, truncated passives are easier to process than actives employing anaphoric sentence subjects (i.e., pronouns).

Charrow and Charrow's (1979) study of jury instructions is arguably the best study on the passive voice intended to help authors revise technical documents. This study offers not only a thorough analysis of passives in naturally-occurring technical text, but an extended and thoughtful discussion comparing the effects of passives versus other psycholinguistic variables.

Traditional jury instructions are, like the law, rather arcane. They commonly present abstract ideas in passages containing multiply-embedded sentences, prepositional phrases, misplaced phrases, passive verbs, nominalizations, "whiz" deletions,[42] double negatives, and obscure words. The authors' goal was to determine the factors that make jury instructions unintelligible and to create and test an improved set of instructions. For Experiment 1, they randomly selected 35 members of a jury pool to listen to and then verbally paraphrase 14 California sets of civil jury instructions. They then determined the accuracy of critical ideas from the paraphrases to pinpoint problematic constructions.

Mean recall of key information in the instructions was 39%, which was compared to recall for key information associated with various discourse variables. Recall was lower for information presented in nominalizations, prepositional phrases

---

[42] I.e., deletions of the clarifying relative pronouns *which* or *that* in constructions such as *Any statement of fact [which was] made during the trial* . . . .

beginning with *as to*,[43] misplaced prepositional phrases,[44] "whiz" deletions, difficult lexical items (e.g., legal terms), double and triple negatives, passive constructions in subordinate clauses, redundant word strings (e.g., *give, devise, and bequeath*), poorly-organized instructions, and poorly-embedded sentences. Recall was higher for information presented using modal verbs (*must, should, may*) and for information presented using passives occurring in the main clause. Recall was unaffected by sentence length, single negatives, the use (as opposed to the location) of passive constructions, and the number of embedded constructions.

Discussing the passive voice results, Charrow and Charrow make several interesting observations. They note that, while the mean amount recalled for information presented in the passive voice was virtually identical to the overall mean of 39%, the type of passive and its location did affect recall. Location had the most dramatic effect: recall was 53% for main clause passives but only 27% for subordinate clause passives. Thus, main clause passives were associated with improved recall while subordinate clause passives were associated with diminished recall. In addition, the mean recall for truncated passives was 43% but only 27% for full passives.[45] These findings led the authors to conclude that "passives, when viewed as a class, are not an outstanding source of confusion" (p. 1325).

For Experiment 2, the authors revised the same jury instructions, based upon the results of Experiment 1. Specifically, they reorganized some instructions; they eliminated double and triple negatives, problematic embeds, and redundant word strings; they also converted abstruse words to simpler words, nominalizations to verbs, passives to actives (and vice-versa, in a few instances where using passives was thought to enhance passage cohesion).[46] Subjects read and then paraphrased 14 instructions,

---

[43] E.g., ***As to any question to which an objection was sustained, you must not speculate as to what the answer might have been or as to the reason for the objection.***

[44] E.g., ***If in these instructions any rule, direction or idea is repeated . . . .***

[45] The difference between full and truncated passives was not significant, perhaps because of the low number of full passives ($n = 9$).

[46] The authors include the original and revised version of each instruction, together with a discussion of the revisions made and an empirical justification for each revision.

half in their original form and half in their revised form.

Mean recall was higher in all cases for each of the improved instructions, although non-significantly so for five of the 14 instructions. Mean recall by type of revision was also higher in all cases, although the differences were significant only for denominalizations, revision of technical terms, and passive to active voice conversions in subordinate clauses. The authors concluded that comprehensibility of jury instructions can often be improved by revising passive constructions in subordinate clauses (especially by adding back "whiz" deletions), denominalizing nominalizations, moving misplaced phrases, removing double negatives, removing vague prepositional phrases, especially those starting with *as to*; eliminating unusual subordinate clause embeds and unusual discourse constructions (e.g., restating the same idea twice in different words).

It is instructive to compare passives that were easier to recall with those that were more difficult. The following passives from the original versions, which were not problematic, were retained in the revised instructions:

> p. 1341: *You **must not be influenced** by sympathy, prejudice or passion.*
>
> p. 1348: *The plaintiff . . . **is entitled** to recover compensation for such injury from that defendant.*

There are several reasons we can posit as to why these passives were not a problem. Both occurs in main clauses, which were generally not problematic for readers in this study. The first sentence likely works well as a passive because using the passive allows the focus to remain on the topic of the jury instructions—the juror. Also, it sounds natural cast in the passive voice. The second sentence is a construction that commonly occurs in everyday speech; because it is idiomatic, it is likely more easily processed. Ironically, in both examples, the passive voice is an effective tool for emphasizing actors—in the first example, the jurors (as opposed to the jurors' emotions); in the second example, the plaintiff (as opposed to an abstract entity, such as natural law).

Examples of problematic passives include the following, followed by paraphrases of the authors' comments:

> p. 1342: *If in these instructions any rule, direction, or idea **is repeated** or **stated** in varying ways, no emphasis thereon **is intended by me** and*

> *none **must be inferred by you**.* Jurors had the biggest problems paraphrasing the first passive; they thought that the instructions might be repeated, not the content within them. Their difficulty with the second two passives may have been due to the obscure term *thereon* preceding them, and with the fact that jurors generally had more difficulty describing the judge's duties rather than their own.
>
> p. 1344: *As to any question to which an objection **was sustained**, you must not speculate as to what the answer **might have been** or as to the reason for the objection. You must not consider for any purpose any offer of evidence that **was rejected**, or any evidence that **was stricken out by the court**; such matter is **to be treated** as though you had never known of it.* These four examples are part of a heavily-passivized instruction containing seven passive constructions. The first two both occur in "as to" phrases; all four passives are in subordinate clauses.
>
> p. 1346: *In determining the weight **to be given** such opinion you should consider the qualifications and credibiliity of the expert and the reasons given for his opinion.* Only 12% of the jurors correctly paraphrased this sentence, presumably because the passive in the subordinate clause makes it unclear who should be weighing the opinion.

Eight of the ten problematic passives occur in subordinate clauses; the two that are in main clauses (see the first instruction) are particularly awkward, full passives that are natural candidates for conversion to the active voice (*is intended by me* becomes *I intend*; *must be inferred by you* becomes *you must infer*). Similarly, the two infinitive passives *to be treated* and *to be given* are also much more direct as *You must treat* and *You must give*, respectively. All these "bad" examples have abstract subjects, and two are in "as to" phrases that the authors characterize as overly vague.

The following revised instructions were converted—with good results—from the active to the passive voice because the original active voice constructions were poorly understood:

> p. 1344: *A question **can only be used** to give meaning to a witness's answer.*
>
> p. 1351: *That situation is where . . . John Smith was the plaintiff's agent, and was performing duties he **was hired by the plaintiff** to do.*
>
> p. 1347: *Even though I allow expert witnesses to give their opinions, you **are not required** to accept those opinions.*

All three passives are commonly-used passive voice constructions; two of the three are in main clauses. The second example, however, is in a subordinate clause. This is an

interesting revision, in light of the authors' conclusions that passives in subordinate clauses are generally problematic. But in this example, the need to maintain parallel structure (to present content words in the same order they were presented in the main clause) took precedence over the need to avoid passives in the subordinate clause.

Charrow and Charrow's study is fruitful for many reasons: (a) it examined passives in naturally-occurring text; (b) it examined passives in the context of a variety of other psycholinguistic variables in order to compare the contribution of each variety to document clarity; and (c) it offered a very detailed discussion of the results. This approach enabled them to perform an exceptionally fine-grained descriptive analysis of passive constructions as used in actual documents. Although their results enable us to make some new observations about the passive voice—e.g., that it may be more problematic in subordinate clauses—it also demonstrates that good revising is not merely a matter of mechnically applying writing rules, even those based on solid empirical evidence: the authors successfully converted an unclear active voice construction to a clear passive voice construction, despite the fact that the successful passive was in a subordinate clause.

Summary of the Findings on Voice Manipulations

Tables 1 and 2 at the end of this chapter present the findings of almost two decades of research on the passive voice. The most relevant findings (organized by dependent variable) are summarized below.

Speed of sentence processing. In some early studies, active sentences that are short, concrete, affirmative and presented in isolation were verified more quickly (Miller and McKean, 1964; Gough 1965) than passive sentences, even truncated passives (Gough, 1966). However, it was soon apparent that this was not always the case. Tannenbaum, Evans, and Williams (1964) found that affirmative active sentences and negative passive voice sentences were verified faster than affirmative passive sentences and affirmative passive sentences were verified more quickly than negative active sentences; Siegel and Burkett (1974) found no difference in verification times for active, passive, and passive-negative sentences.

Slobin (1966) found that non-reversible passive sentences were verified as quickly as active sentences; Glucksberg, Trabasso, and Wald (1973) found that, although passives sentences were verified less quickly than active sentences, this effect was seen primarily with reversible sentences. When saliency of subjects and objects was varied, subjects took less time to verify actor-focused pictures matched with active sentences and action receiver-focused pictures matched with passive sentences (Olson & Filby, 1972; Tannenbaum & Williams, 1968). When sentence expectation was varied, subjects took less time to verify expected sentences, regardless of voice. Voice also had no effect on the speed of identifying actors and action receivers in sentences where reader expectation was manipulated (Herriot, 1969); neither did it affect the speed of detecting anomalous sentences (Moore & Biederman, 1979).

Accuracy of sentence recall.[47] Although Mehler (1963) found that active sentences were better recalled than passives, later studies failed to find a uniform advantage for the verbatim recall of active voice constructions. Some studies found no voice effects on sentence recall (e.g., James, 1972); others found voice effects only in interaction with other variables, such that active sentences were better recalled in some situations and passive sentences were better recalled in others. Variables interacting with voice include delay in presentation (Sachs, 1967; Anderson, 1974), Yngve depth (Perfetti, 1969; Parot & Kail, 1980; Martin & Roberts, 1966), passage thematization (Perfetti & Goldman, 1975; Parot & Kail, 1980), mode of presentation (oral versus written—Sachs, 1974; Kerr et al., 1982), object versus subject prompt (Wearing, 1972) and subject versus object saliency (Trabasso, 1972).

Preference for voice and sentence focus. In tasks in which pictures were matched to

---

[47] Sentence recall is distinguished from sentence comprehension (see next section) in that it focuses on verbatim recall of sentences, with attention paid to syntactic structures (e.g., whether the sentence is recalled in the active or passive voice) and the retention of syntactic versus semantic information as an indicator of how sentences are represented in and retrieved (or reconstructed) from memory. Conversely, sentence and passage comprehension measures use a variety of techniques, including the ability to accurately reproduce discourse units, as indicators of prose comprehension. The measures are sometimes similar, but the goals of the study and the ways that authors discuss the results are quite different.

sentences describing those pictures, people generally preferred using active sentences to describe actor-focused pictures and passive sentences to describe action receiver-focused pictures (Harris, 1978; Hornby, 1972; Johnson-Laird, 1968a, 1968b; Wilson, 1979). Johnson (1967) found that voice had no effect on perceptions of animacy: nonsense words at sentence beginnings were perceived as more animate than words that occur later in the sentences, such that subjects (a) use whichever voice (active or passive) allows them to place animate words at sentence beginnings (see also Clark, 1965) and (b) to recode sentences with animate actors and inanimate action receivers as active but to recode sentences with inanimate actors and animate action receivers as passives. Voice did not affect judgments about the importance of articles (*a* or *the*) (Grieve & Wales, 1973), and selection of articles (Hupet & Le Bouedec, 1975; Wright & Glucksberg, 1976).

Preference for voice and passage focus. Perfetti and Goldman (1975) found that passive sentences at the end of a passage were preferred over active sentences when the passives served to maintain the thematic integrity of the passage. Kerr et al. (1982) apparently failed to find such an effect, but their results were less well described than those of Perfetti and Goldman's.

Accuracy of sentence and passage comprehension. For two technical articles and eight paragraphs, Coleman (1964) found that active-voice, denominalized versions were better recalled than passive-voice, nominalized versions. However, in a more extended study comparing the effects of a variety of psycholinguistic variables on the comprehensibility of jury instructions, Charrow and Charrow (1979) found that passives were comprehended less well only when located in subordinate clauses; when located in main clauses, they were comprehended as well as actives.

In studies where the relationship between the voice of a question was compared to that of information previously presented, sentences presented in groups of 1 or 2 were better comprehended when the voice of the question and the sentence matched (Wright, 1969), although this effect did not persist when sentences were presented in larger groups (Layton & Simpson, 1975).

Evaluation of Past Research on the Passive Voice

Researchers have manipulated voice for a number of different purposes, both theoretical and practical, and have used a variety of dependent measures to assess the effects of voice on speed of sentence processing, accuracy of recall and comprehension, and preferences for using active or passive voice in a variety of contexts. Given the number of studies and the variety of research questions addressed, it is surprising how few of these findings are relevant for understanding how voice affects either reading speed or reading comprehension.

There are several reasons for this. First, as mentioned earlier, very few researchers who manipulated voice were interested in assessing its effects on reading speed or comprehension. Although Coleman (1964, 1965) specifically looked at the effect of voice on comprehension, his methodology and reporting procedures were substandard. For two of his experiments, he did not manipulate voice in isolation, and for all of his experiments, he did not describe his results in enough detail to make them interpretable. Thus, whatever support his results offer for the idea that passives inhibit comprehension is very weak. Siegel & Burkett (1974) were also interested in determining how voice affects passage difficulty, but they inexplicably chose a sentence verification task to try to answer this question. Their choice of task, their choice of stimuli, and the incompleteness of their literature review calls into question their finding that casting negative sentences in the passive causes problems. Charrow & Charrow's 1979 analysis of jury instructions was both more thoughtful and more statistically rigorous, but it also did not manipulate voice in isolation, but in combination with several other variables.

Second, the great majority of the studies used isolated sentences as stimuli, not sentences occurring in connected prose. And the handful that did examine sentences in connected prose either manipulated several variables simultaneously (e.g., Coleman, 1964; Charrow & Charrow, 1979) or manipulated only the voice of the target sentence rather than the voice of sentences in the passage surrounding the target (Perfetti &

Goldman, 1975; Kerr et al., 1982).[48]

Third, the most widely-cited studies used isolated sentences of a particular sort—sentences that are short, concrete, and narrative, e.g.: *John liked the small boy* (Miller, 1962), *She liked it* (Miller & McKean, 1964), *The girl has worn the jewel* (Mehler, 1963), *The boy hit the girl* (Gough, 1965). In these examples, the actor is more salient than the action receiver: in *Miller*, the actor is named and is thus thematized; in *Miller and McKean* and *Mehler*, only the actor is animate; in *Gough*, the violence of the act makes the actor more salient. Because people prefer to place salient information at sentence beginnings (Clark, 1965), sentences such as these are biased in favor of the active voice. This bias is even more evident when we convert these sentences to full passives: *The small boy was liked by John*, *It was liked by her*, *The jewel was worn by the girl*, and *The girl was hit by the boy*. As Gough pointed out, such sentences all make awkward, unidiomatic passives that are unlikely to occur in English speech or writing. It is not surprising that people tend to convert such passives to actives.

Fourth, the studies are old. This means that computers were generally not used for presenting stimuli, collecting data, or performing statistical analyses. Thus, it is hard to know whether the same results would be seen for some experiments using modern equipment. Also, some studies did not present statistical analyses that were sufficiently thorough: quite a few researchers frequently focused only on effects they were interested in, not all the possible effects and interactions. Whether this was due to less rigorous reporting standards, the difficulty of procuring time on the university mainframe, or the tediousness of computing statistics by hand, it often made the results more difficult to interpret.

As a result of these methodological problems and limitations, the generalizability of this body of research is limited. Although the preceding discussion provides ample justification making this claim, I would like to discuss two additional characteristics of stimuli sentences that are of special relevance when generalizing to

---

[48] Anderson (1974) varied actives and passives in a short passage, but the passage was atypical of those seen in actual prose. (I.e., it featured many actors, all acting upon one another in a fairly confusing way; the content development was minimal. Also, the all passives used were both full and reversible.)

scientific prose: sentence reversibility and passive truncation. In addition to using stimuli sentences that are short, concrete, narrative, isolated, and idiomatic only in the active voice, researchers have usually used (a) sentences that are reversible for both active and passive versions and (b) passive sentences that are full passives—passives containing an agentive *by*-phrase.

In a reversible sentence, when the nouns occupying the subject and object slots are switched, the result is still a semantically acceptable sentence (e.g., *The girl hit the boy* is reversible; *The girl hit the wall* is not). It is easy to see why a reversible sentence might be at least marginally harder to memorize and recall than a non-reversible sentence: the respondent must expend extra effort distinguishing the actor from the action receiver. The difficulty is increased when the sentence is presented in isolation, due to the absence of other recall cues that might otherwise be gleaned from the surrounding context. But if the sentence is in the active voice, the respondent still has another cue he can use to decrease the difficulty of his task: the ability to map the actor onto the subject slot and the action receiver onto the object slot. However, if the sentence is presented as a passive (*The boy was hit by the girl*), that cue is no longer available to aid in disambiguation. As a consequence, more recall errors or slower reaction times might result. But if this in fact happens (i.e., if voice and reversibility interact), should we conclude that it is the passive voice that caused poorer recall or slower reaction times? Probably not.[49] This is not what Slobin (1966) concluded when he found voice differences in sentence verification times only for reversible (but not non-reversible) sentences. He argues instead that "the difficulty in understanding passive sentences is partly attributable to the problem of keeping track of which noun is the actor. In the case of non-reversible sentences, this problem is solved" (p. 226).[50]

Similarly, passives appear to be problematic mainly when they are full (i.e.,

---

[49] Interestingly, these results argue equally well for the case that reversible sentences decrease comprehension, etc., but I have yet to see anyone argue that writers should avoid reversible sentences when they write.

[50] Slobin's results also support the idea that reversibility alone increases the difficulty of sentence processing, in that mean reaction times were the lowest for non-reversible sentences, whether active or passive (.69 vs. .70 secs), but were higher for active reversible sentences (.93) and higher yet for passive reversibles (1.21).

when they retain the *by*-phrase) rather than truncated. Yet most of the studies in the literature review used full—not truncated—passives as stimuli. There are several probable reasons for this. For the studies testing Chomsky's transformation theory, a proposition presented in the passive voice must of necessity include the *by*-phrase if it is to semantically match the same proposition in its active voice form. Thus, all the earliest experiments testing this hypothesis (with the exception of Gough, 1966) compared full passives with actives.[51] Also, the terminology most frequently used to describe passives—*full passive* and *truncated passive*— implies that the former is the "true" passive and that the latter is a cut-down version. However, linguists such as Coetzee (1980) have observed that, in most languages, truncated passive forms are both more common and more etymologically ancient than full passives. He also cites the fact that, while many languages possess both the short and long forms of the passive, no language possesses only the long form. For these reasons, Coetzee seriously questions the idea that the "truncated" passive is simply an abbreviated version of the full passive.

Whatever the reasons for comparing full passives to actives in experimental studies, the result has been to limit the generalizability of results obtained in any study that operationalizes passives as full passives. This includes the vast majority of voice studies, even many studies which did not test Chomsky's transformational hypothesis (e.g., Coleman, 1965; Anderson, 1974; Perfetti & Goldman, 1975).

Of the studies that did use truncated passives as stimuli, most found that the negative effects associated with full passives were absent with truncated passives.[52] For example, Slobin (1968) found that people re-telling stories were much less likely to transform truncated passives (vs. full passives) into actives. Charrow and Charrow (1979) found that comprehension of jury instructions, as measured by mean recall of

---

[51] In addition, Chomsky's (1957) formal rule for transforming actives to passives includes the *by*-phrase. Chomsky must have been aware that this phrase is optional but does not clearly indicate this either in the way he formulates the rule nor in the way he discusses active-passive transformations in the text. Thus, it is not surprising that researchers testing the transformational hypothesis would manipulate full passives, not truncated passives.

[52] Gough (1966) found otherwise: that truncated passives such as *The boy was hit* were verified less quickly than their active counterparts. But he used sentences that were reversible and which make awkward passives, as he himself noted.

ideas presented in active or passive constructions, was 43% for ideas presented as truncated passives (slightly higher than overall mean recall of 39%), as opposed to 27% for ideas presented as full passives.

These studies on reversibility and truncation suggest that passives may be problematic for a very restricted range of sentences—reversible sentences that employ full passive constructions (e.g., *John was picked by Bill*).[53] If this is true, then the already limited generalizability of the results from studies using reversible sentences and full passives is even further limited. But in order to know how limited, we would need to know the incidence of such constructions in the domain to which we want to generalize.

However, I was able to locate only general estimates on the incidence of truncated passives in scientific writing and no estimate of the incidence of reversible sentences in scientific writing. Thus, I expanded my study of the incidence of passive verbs in scientific articles (see Study 1) to include the categorization of a sample of passive sentences with regard to both reversibility and truncation. I found that over 97% of the 751 passives I examined were both non-reversible and truncated. Thus, the claim I have already made that most studies on voice have limited generalizability is particularly applicable in the case of scientific prose. Most sentences in scientific prose do not resemble the sentences used in studies on voice. They are not predominantly short, concrete, or narrative. They do not occur in isolation. And they are very rarely reversible[54] or truncated.

Thus, I must conclude that the research literature on voice does not presently support the claim that passive verbs make scientific documents difficult to read. It does, however, point out one critical fact: that the effects of voice are highly sensitive to context, as Spyridakis and Wenger (1992) have also noted. Past research on passives has amply demonstrated that voice frequently interacts with other variables,

---

[53] Making such sentences non-reversible (*The flowers were picked by Bill*) or truncated (*John was picked* or *The flowers were picked*), appears to eliminate passive problems.

[54] Although I did not count them, it is doubtful that there are many active sentences that are reversible, either, because most sentences describe either abstractions or the manipulation of objects. The occasional exception occurs in psychology, where the "objects" of the study are often people. So there are slightly more opportunities for reversible sentences as a result.

producing either no difference across voice or an advantage for active voice in some contexts, passive voice in others.

Therefore, if we want to determine the effects of passive voice in the context of scientific writing, it is best to study the effects of voice by using actual scientific reports or documents closely simulating such reports. The studies described in the three ensuing chapters represent a preliminary effort to do three things: to examine the how scientists in three fields use passives when they write (Study 1); how passives in scientific reports affect reading speed and reading comprehension (Study 2); and how passives affect instructors' evaluations of student lab reports (Study 3).

Table 1
Passive voice studies by dependent variable measured

Dependent variable	Dependent measure	Studies
Ease of sentence processing as measured by RT measures other than sentence verification tasks	RT for sentence transformations RT for naming actors and action receivers RT for judging a sentence as anomalous RT for inserting v-o-s sets into sentence forms	Miller (1962), Miller & McKean (1964) Herriott (1969) Moore & Biederman (1979) Tannenbaum, Evans, & Williams (1964)
Ease of processing as measured by RTs for sentence verification tasks, i.e., matching subject- or object-focused stimuli to active or passive stimuli	Stimulus sentences elicit matching sentences Stimulus pictures elicit matching sentences Stimulus sentences elicit matching pictures Stimulus passages elicit matching pictures	Anderson (1974), Siegel & Burkett (1974) Olson & Filby (1972),Glucksberg, Trabasso, & Wald (1973), Johnson-Laird (1968b) Slobin (1966), Gough (1965, 1966), Johnson-Laird (1968a), Glucksberg, Trabasso, & Wald (1973) Tannenbaum & Williams (1968)
Ease of processing as measured by accuracy of recall and recognition of active vs. passive sentences or passages	Verbatim sentences reproduced Meaning-preserving sentence paraphrases produced Recognition for sentences changes No. of words recalled as a function of prompt type No. of words recalled as a function of imagery Passage recall	Mehler (1963), Martin & Roberts (1966), Sachs (1967), Martin, Roberts & Collins (1968), Perfetti (1969), Coleman (1965) Perfetti & Goldman (1974, 1975) Sachs (1967,1974), Wearing (1970), Kerr et al. (1982) Wearing (1972) James (1972) Coleman (1964), Charrow & Charrow (1979)
Comprehension as measured by accuracy of answering questions	Accuracy of answers to comprehension questions Accuracy of identifying subject or object concepts	Wright (1969), Layton & Simpson (1975), Coleman (1964) Trabasso (1972)
Preference for active or passive sentences as moderated by subject or object focus	Selection of one of two sentences descriptions for a picture Selection of sentences in a passage recall task	Hornby (1972) Kerr et al. (1982), Perfetti & Goldman (1965)
Preference for active or passive sentences as moderated by noun animacy	Selection of nouns to insert into sentences frames Placement of nouns into pre- or -post-verb positions Animacy ratings of nonsense words Descriptions of pictures Voice transformations of sentences	Clark (1965) Jarvella & Sinnott (1972) Johnson (1967) Harris (1978) Wilson (1979)
Preference for voice switches as moderated by passive truncation	Voice switching when re-telling a story	Slobin (1968)
Preference for sentence wording as a function of voice and article definiteness	Preferred questions for sentences Preferred endings for sentences Preferred article for 1st and 2nd noun	Grieve & Wales (1973) Hupet & Le Bouedec (1975) Wright & Glucksberg (1976)

Table 2
Key passive voice studies by topic and findings

General effects of voice on speed of sentence processing	
Miller (1962)	**Simple** transformations were faster than complex transformations.
Miller & McKean (1964)	**Negative** transformations were faster than passive transformations.
Tannenbaum, Evans, & Williams (1964)	**Verb**-object-subject sets were inserted into sentence forms as follows: active < negative & passive-negative < passive.
Gough (1965, 1966)	**Active** sentences were verified faster than passive sentences.
Slobin (1966)	**Passive** sentences were verified faster than negative sentences.
Anderson (1974)	**After** no delay, (animate, reversible) target sentences matching probe sentences were verified faster, regardless of voice; after a delay, active sentences were verified faster than passives .
Moore & Biederman (1979)	**Speed** of detecting anomalous (mostly non-reversible) sentences was not affected by voice.
Effects of voice on accuracy of verbatim sentence recall	
Mehler (1963)	**Active** sentences were better recalled (verbatim) than passive sentences
Coleman (1964)	**Simplified** (active, denominalized) paragraphs were better recalled than non-simplified paragraphs.
Sachs (1967)	**Subjects** retained verbatim recall for target sentences presented in connected discourse only after no delay between time of presentation and recall; after a delay, only meaning, but not syntactic form, was recalled.
Martin & Roberts (1966)	**Differences** in recall were accounted for by differences in Yngve depth, not voice (active, passive, passive-truncated) or negation.
Prentice (1966)	**Semantically** priming initial sentence nouns facilitates verbatim recall for both actives and passives.
Martin, Roberts & Collins (1968)	**Simple**-passive sentences were recalled best at a variety of retention intervals; recall for words in passive sentences was also less variable than for simple-active, complex-active, and complex-passive sentences.
Perfetti (1969)	**High**-Yngve depth actives and low-Yngve-depth passives were recalled best.
James (1972)	**High**-imagery subjects & objects were better recalled than low-image subjects & objects, regardless of voice.
Anderson (1974)	Target sentences which matched probes in voice were better recalled than those that did not but after a delay this interaction was much weaker.
Effects of voice on sentence/passage comprehension	
Coleman (1964)	**Articles** with active verbs and denomizalized articles were better comprehended; similar paragraphs were better reconstructed.
Wright (1969)	**Sentences** were better comprehended when the voice of both sentence and question matched.
Trabasso (1972)	**Errors** in concept identification were fewest for active sentences were the subject was the salient feature and for passive sentences where the object was the salient feature.
Layton & Simpson (1975)	**Replicated** Wright (1969) for sentences presented in groups of 1 or 2, but not 4 or 8 (where passive questions elicited more errors, regardless of voice).
Charrow & Charrow (1979)	**Only** passives located in subordinate clauses were less well comprehended.
Effects of voice when subject and object are reversible	
Slobin (1966)	**Non-reversible** passives were verified as quickly as non-reversible actives.
Herriot (1969)	**Actors** and action receivers were identified faster for reversible-but-expected sentences than for reversible-but- unexpected sentences, regardless of voice.
Wright &Glucksberg (1976)	**Sentence** voice affected article selection (*a, the*) for reversible, but not non-reversible, sentences
Glucksberg, Trabasso, & Wald (1973)	**Sentence** verification tasks using sentences > pictures and pictures > sentences reflected different processing strategies; passive sentences and reversible sentences took longer to verify only when true.

Effects of voice when passives are truncated	
Gough (1966)	**Active** sentences were verified faster than truncated passives.
Slobin (1968)	**In** the retelling of a story, most full passives were converted to actives; most truncated passives were not.
Charrow & Charrow (1979)	**Truncated** passives in jury instructions were as well recalled as actives.
Effects of voice when article definiteness is varied	
Grieve & Wales (1973)	**Nouns** with definite articles were judged as more important than those with indefinite articles, regardless of voice or regardless of whether the passives were full or truncated.
Hupet & Le Bouedec (1975)	**Nouns** with definite articles were less often changed than those with indefinite articles, regardless of voice.
Wright & Glucksberg (1976)	**Sentence** voice affected article selection (*a, the*) for reversible, but not non-reversible, sentences.
Effects of voice when sentence/passage focus is varied (focus on actor vs. action receiver)	
Johnson-Laird (1968a, 1968b)	**Active** sentences elicited subject-focused drawings while passive sentences elicited object-focused drawings and vice-versa.
Hornby (1972)	**Pictures** depicting psychological subjects were described using the active voice; pictures depicting psychological objects were described using the passive voice.
Olson & Filby (1972)	**Actor**-focused line drawings were verified faster when paired with active-voice descriptions; the reverse was true for line drawings focused on action receivers.
Wearing (1972)	**More** words were correctly recalled from passive than active sentences , mainly due to a strong effect of object (vs. subject) prompts; the pattern of results was similar for a 0 and 48 hour delay.
Glucksberg, Trabasso, & Wald, 1973	**Sentence** verification tasks using sentences > pictures and pictures > sentences reflected different processing strategies; passive sentences and reversible sentences took longer to verify only when true.
Tannenbaum & Williams (1968)	**Neutral** and subject-focused passages using passive sentences elicited longer response times to describe stimulus pictures, but object-focused passages did not.
Effects of voice when actors and action receivers vary in animacy or imagery	
Clark (1965)	**Animate** words were inserted at sentence beginnings (i.e.,animate actors elicited active sentences and animate objects elicited passive sentences).
Johnson (1967)	**Subjects** rating nonsense syllables in subject or object slots rated those in the subject slot as more animate in both active and passive sentences, although active sentences received higher animacy ratings.
Jarvella & Sinnott (1972)	**Sentences** with animate actor and inanimate action receivers are recoded as active; sentences with inanimate actors and animate action receivers are recoded as passives.
James, Thompson, & Baldwin (1973)	**Voice** preference in sentence reconstruction depended upon which voice permitted movement of high-imagery nouns to sentence beginnings.
Harris (1978)	**Picture** descriptions were in the active voice when picture focus was on the actor, in the passive when focus was on the acted-upon; when animacy was equally stressed, more actives were produced, although the tendency was less for inanimate-inanimate sentences.
Wilson (1979)	**Experiencer**-verb-object sentences were transformed from passive to active voice while object-verb-experiencer sentences were transformed from active to passive voice.

Study 1:
Incidence of Passive Constructions in Empirical Articles from Three Scientific Fields

Many current scientific style manuals disparage the use of the passive voice:[55]

> Use the active rather than the passive voice, and select mood and tense carefully. **Prefer the active voice** [bolding theirs] (p. 32), *Publication Manual of the American Psychological Association* (1994)

> In any type of writing, the active voice is more precise and less wordy than the passive voice. It is the natural voice in which most people speak and write. The active voice also adds energy to your writing (p. 98). Matthews, Bowen, & Matthews (1996) *Successful Scientific Writing*

> In any type of writing, the active voice is usually more precise and less wordy than is the passive voice (p. 160). Day (1994), *How to Write and Publish a Scientific Paper*

> It is not only "more scientific" and "objective" to use the passive voice; it is only more imprecise—and cowardly (p. 41). Schwager (1991), *Medical English Usage and Abusage*

> The passive voice retards and even reverses the movement of a sentence . . . . the passive voice may muddle . . . . Most sentences should be written in the active voice (p. 145-146). Lynch & Chapman (1980), *Writing for Communication in Science and Medicine*

> Active verbs give vitality to writing, whereas the passive voice tends to dull prose (p. 170). Fondiller (1992), *The Writer's Workbook*

> [Taking] the point of the view of the experimenter ("we"), is undeniably more lively because it usually requires the use of the active voice (p. 128). Zeiger (1991), *Essentials of Writing Biomedical Research Papers*

---

[55] Most of these authors also include examples of instances in which the passive voice is acceptable. However, these examples are always cited as exceptions to the general rule that passive should be avoided (Riley, 1991, also noted that most textbooks authors recommend passives only as a "last resort".) In no case was the passive presented in a truly positive light—as a useful, constructive tool for scientific authors.

These comments present a very negative view of the passive voice as (a) imprecise, (b) wordy, (c) unnatural, (d) confusing and even (d) cowardly. The same authors characterize the active voice much more favorably—as vital, lively, precise, and concise. While not all authorities are so sharply critical as those cited above, the general consensus from those who give scientific writing advice about passives is that they are inferior constructions and should be avoided whenever possible.

Despite this "pro-active voice" bias on the part of most style guide authors, a casual glance at most scientific articles reveals an abundance of passive constructions, although the exact incidence of passives in published scientific prose has not been systematically investigated. This discrepancy between what scientific authors are advised to do (avoid passives) and what they actually do (to use passive constructions rather liberally) is curious. Some authorities have tried to explain this discrepancy to tradition (Lynch & Chapman, 1980; Matthews et al., 1996) or to the desire of scientists to avoid using personal pronouns (Matthews et al., 1996). Day (1994) characterizes the widespread use of passives as a "bad habit" that arises out of the desire to avoid first-person pronouns.

There is probably some truth in the assertion that scientists persist in using the passive voice because of linguistic traditions or discomfort with using personal pronouns (especially *I*). But it is difficult to believe that this is the only reason that scientists continue to use passives in the face of ongoing criticism. To determine possible reasons *why* scientists use the passive voice, we need to determine *when* and *how* they use it.

Most of the studies on the use of passives in scientific writing have simply counted the number of passives in the entire article, focusing primarily on comparing passive usage in different writing genres. Svartvik (1966) examined the incidence of passive verbs in different prose genres. A preliminary analysis of 5000-word excerpts from two novels and one scholarly scientific monograph[56] revealed that although the novels contained 7% and 5% passive verbs, the scientific monograph contained 32%

---

[56] *The Echoing Grove*, by R. Lehmann; *Eating People is Wrong*, by M. Bradbury; *The Life of Vertebrates* by J. Z. Young.

passive verbs; 77% of the passives across the three texts were truncated.

Svartvik performed a more extensive analysis of 28 texts in eight different genres (science, news, arts, speech, sports, novels, plays, and advertising). However, he calculated the number of passive verbs per 1000 words, rather than the number of active versus passive verbs. Thus, while it is possible to report the frequency of passives within the sample, it is not possible to compare the incidence of passive versus active verbs for each genre. The incidence of passive verbs/1000 words (from highest to lowest) is the same order as the list above:

science 19.3
news 13.6
arts 10.0
sports 7.0
speech 6.5
novels 4.5
advertising 1.8
plays 1.2

Although no figures are given for the total percentage of passives truncated, a figure depicting truncated passives by genre shows a relatively stable percentage of truncations across genres, approximately 80%, a percentage very similar to that obtained in Svartvik's preliminary study.

Thus, the results of this cross-genre study demonstrate that, not surprisingly, scientific texts have a higher percentage of passive verbs than other writing genres. However, his analysis offers little insight into the pattern of usage in different scientific fields, because Svartvik drew from a variety of scientific texts, including a science book characterized as "popular science,"[57] scholarly articles on neutron diffraction, mathematical-physical state changes, homolytic aromatic substitution, and a selection of other articles and letters on biochemistry, biology, genetics, metallurgy, meteorology, psychics, and physiology.

Other writers have described the use of passive in scientific texts, although

---

[57] *The Psychology of Thinking* by R. Thomson.

these descriptions usually been relatively brief. Examining passive usage in six scientific articles, Rodman (1981) found that most passives were truncated: in four articles, more than 90% of the passives were truncated; in the other two articles, just 80% and 73% were truncated. Paul and Rosner (1983) tallied the incidence of various syntactic devices, including the passive voice, in three fields—engineering, agronomy, and home economics—in popular, semi-technical, and technical articles. They found that the popular literature contained the fewest number of passives (engineering = 18%, agronomy = 10%, home economics = 6%, $\bar{x}$ = 11%), the semi-technical the next highest percentages (engineering = 24%, agronomy = 26%, home economics = 14%, $\bar{x}$ = 21%), and the technical the highest percentages (engineering = 39%, agronomy = 42%, home economics = 37%, $\bar{x}$ = 39%). Six publications from each field (two from each level of technicality) were sampled, but the authors did not reveal either the sample size or the specific articles sampled.

Gross (1984) compared the writing advice given to scientific writers and editors with the actual content of numerous articles that had been submitted, edited, and published in a "well-respected" scientific journal. He notes that although scientific writers are told to avoid the passive voice, it was widely used in the articles he examined. He noted that, although editors often changed constructions such as "*it is suggested* to *we suggest*, "they seemed to have no preconceived preference [for the active voice]" (p. 247). He also found that passives were more widely used in Method and Discussion sections, which he said "are generally more comfortably phrased in the passive" (p. 247).

Riley (1991) conducted a more detailed analysis of the incidence of passive verbs in the field of speech pathology. She was interested in how passive usage varied as a function of the rhetorical role assumed by the scientific writer in different sections of a scientific article (Introduction, Method, Results, and Discussion). She examined the percentage of passives in each section of 12 published scientific articles in the field of speech pathology. Her analysis revealed four ordinal patterns of passive

usage, shown with the number of articles exhibiting each pattern in parentheses:

	lowest incidence <————> *highest incidence*	
Pattern 1	Discussion < Introduction < Results < Method	(7 articles)
Pattern 3	Introduction < Discussion < Results < Method	(2 articles)
Pattern 2	Results < Discussion < Introduction < Method	(2 articles)
Pattern 4	Discussion < Results < Introduction < Method	(1 article)

For all articles, passive usage was highest in the Method section; for 75% of articles (9/12), passive usage was higher in the more objective sections (Method and Results) than the more interpretive sections (Introduction and Discussion). Pattern 1 (Discussion < Introduction < Results < Method), seen in 58.3 % (7/12) of the articles, was clearly the most common pattern of passive usage.[58]

The fact that authors were more likely to use active voice when interpreting information and passive voice when describing it led Riley to conclude that these writers used passives in appropriate ways. Noting that writing advice about using stylistic features such as the passive voice may easily be overgeneralized or incomplete, Riley recommended that writers exercise caution when interpreting the advice offered by many writing texts.

The purpose of the current study was to replicate and extend Riley's study by documenting the incidence of passive verbs by section (Introduction, Method, Results, and Discussion) to see whether the patterns of voice usage by section in three scientific fields would be similar to those observed by Riley in speech pathology articles. In addition, I wanted to systematically examine the incidence of passives both within and across scientific fields. Thus, I chose fields representing three branches of scientific inquiry: the social sciences (psychology), life sciences (botany), and physical sciences (chemistry/physics) in order to determine whether there would be a relationship between field focus and passive usage.

Examining the use of passives in different article sections, I expected to find a pattern similar to that observed by Riley—more passives in the objective sections of scientific reports (Method and Results). In addition, I expected that scientists in a

---

[58] Using Riley's table of percentages, I calculated the mean percentage of passive main clauses as follows: grand mean = 32.71, Discussion = 18.0, Introduction = 24.5, Results = 32.83, Method = 55.5

highly "object-oriented" field such as chemistry/physics[59] would tend to use more object-oriented prose (i.e., more passives) than a more "subject-oriented" field such as psychology. By "object-oriented" field, I refer to a field in which scientists literally study objects (e.g., chemicals, electrons) but also to the idea that, in object-oriented fields (chemistry, physics, geology, astronomy), it is easy for scientists to psychologically separate themselves from what they study. In subject-oriented fields (psychology, sociology, economics, anthropology), it is difficult to psychologically separate oneself from one's field because the field focuses on some facet of human behavior (or on animal behavior that can be generalized to humans) and the investigators are of course human. In addition, there are scientific fields that lie midway between these two extremes (e.g., botany, zoology, animal behavior), in which investigators are not studying human behavior, but are studying the behavior of living organisms that share similarities—to one degree or another—with humans. I expected that scientists studying plants would used more passives than psychologists but fewer than chemist/physicists.

I had three hypotheses about the pattern of passive usage in the articles examined: (a) there would be a high incidence of passive constructions in all the scientific articles examined (higher than 25%); (b) passive usage would relate to focus of inquiry, such that the incidence will be highest in chemistry/physics, second-highest in botany, and the lowest in psychology; and (c) there would be a higher incidence of passives in the relatively objective sections of articles (Method and Results) than in the relatively interpretative sections (Introduction and Discussion).

To gain additional insight into the patterns of passive usage in scientific writing, I also examined the incidence of different types of passives by type—truncated, reversible, and cleft passives (e.g., *It was found that*)— in a subset of the articles previously sampled. I thought that a closer examination of truncated and reversible passives might shed some light on the reasons that some early studies that demonstrated faster processing or better comprehension for active (vs. passive) sentences. Most of these studies compared reversible sentences (e.g., *The boy chased*

---

[59] The journals used were *Analytical Chemistry*, *Journal of Chromatography*, and *Chemical Physics Letters*.

the girl vs. *The girl was chased by the boy*). Furthermore, in these studies, the passive sentence versions were non-truncated (full passives); i.e., the agent of the action was retained using a *by*-phrase, rather than deleted (*The girl was chased*). Thus, in order to know whether even the handful of studies that showed a definite advantage for active voice are generalizable to the kinds of sentences we see in scientific writing, it is necessary to know the extent to which the passive sentences used in scientific reports are either (a) reversible or (b) full (non-truncated).

I wanted to look at the incidence of cleft passives (*It was shown that*) because such constructions are widely cited by passive voice critics as examples of awkward passives in scientific prose. I suspected that such constructions were atypical of the passives most often used in scientific writing.

Thus, I hypothesized that the great majority of passives examined would be truncated and non-reversible; I also hypothesized that the incidence of cleft passives would be very low.

Method

Thirty-six articles, twelve from each of three fields (psychology, botany, chemistry/physics), were selected for analysis (see Appendix A for article citations). For each field, four articles from three different journals were used, in order to ensure that the pattern of voice usage was not due solely to editing practices of any particular journal. Each article was randomly selected from a mainstream scientific journal, was organized using conventional scientific format (Introduction-Method-Results-Discussion) and was published since 1993; each reported the results of only one experiment[60] and was relatively short (5 pages or less in length). No two articles were on the same topic or were authored by the same investigators.

For the initial analysis, the total number of verbs was tallied for each article, and then verbs were categorized as either active or passive voice, and tallies and percentages were calculated for each category. This approach was somewhat different

---

[60] Analyzing a single-study report made it possible to compare article sections (Introduction-Method-Results-Discussion). This would not be possible with multi-study reports, which often have a different organization (i.e., more abbreviated Method, Results, and Discussion sections and an expanded General Discussion section dealing with several studies).

than that used by Riley, who just tallied verbs in main clauses. However, I tallied verbs in both main and subordinate clauses because I believe that both clauses typically make major contributions to the understanding of a sentence. For example, in the sentence preceding this one, the main clause (*However, I tallied verbs . . . .*) explains what I did, but the subordinate clause (*because I believe that . . . .*) explains the reason I did it. I would argue that both clauses are equally important—that there is no good reason to count only the verbs observed in main clauses.

To be categorized as a verb, a word was required to function as a verb, rather than as an adjective, adverb, or noun. Thus, only verbs paired with subjects (e.g., *The results **were reported***) were rated as verbs; the following examples would not be considered verbs: *The theory **proposed** by Jones was not supported, Smith tried to **solve** these methodological problems with a new approach, **Reporting** potential confounds is essential in research papers.* Verbs were identified as active voice if the subject of the clause corresponded to the actor (e.g., *We **interpreted** the results*), as passive voice is the subject corresponded to the action receiver: *The results **were interpreted**.* If the participle of a potentially passive construction acted primarily adjectival in function (e.g., *These variables are **correlated***), the sentence was categorized as active voice (the verb in this example is *are*).

To examine the incidence of truncated, reversible, and cleft passives, I examined all the passive verbs in a subset of the same articles (four articles from each field, $N = 12$). Each passive was classified with regard to reversibility (reversible vs. non-reversible) and truncation (present vs. absent). Cleft constructions (e.g., *it was found that, it was reported that*) were independently tallied and were also evaluated as part of the analysis of truncation and reversibility. However, because of the rarity of cleft constructions, results for the analysis of truncation and reversibility are not broken down by cleft versus non-cleft constructions.

Results

The data were analyzed using SPSS-PC+, v. 4.0. Figure 1, which shows the mean percentage of passive verbs by field, demonstrates a pattern of results in the predicted direction. An omnibus ANOVA revealed significant differences between the means:

psychology (23.97) < botany (34.69) < chemistry/ physics (32.91); the means for each individual article section—Introduction, Method, Results, and Discussion—were also compared (see Table 3). For three of the four sections (i.e., Introduction, Results, and Discussion), this pattern of means (psychology < botany < chemistry/physics) was observed. Only for the Method section was a different pattern observed (psychology < chemistry/physics < botany).

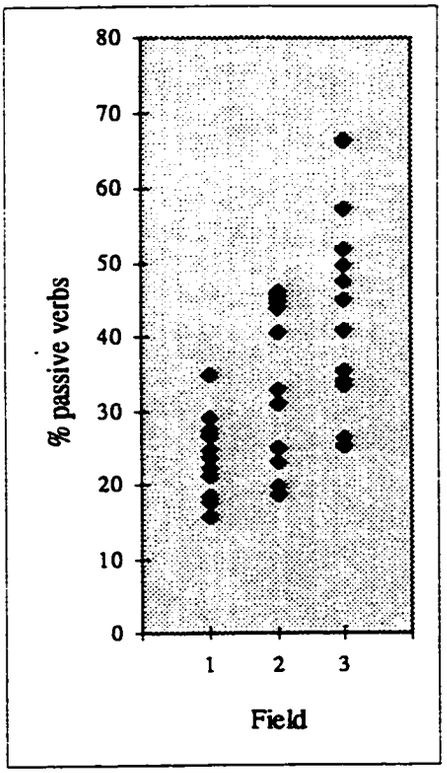

Figure 1. *Distribution of passive verbs by article within each field,* n = *36 (1 = psychology, 2 = botany, 3 = chemistry/physics).*

Analytical comparisons were performed to determine the source of the differences between the percentage of passives by field. Thus, three analyses were performed: psychology vs. botany, psychology vs. chemistry/physics, and botany vs. chemistry/physics. Because psychology is the most subjective field and chemistry/physics is the least, we would expect that differences there to be significant, which they were, both (a) for all sections individually: Introduction (17.62 vs. 32.25), Method (43.11 vs. 67.16), Results (26.95 vs. 41.82), and Discussion (20.59 vs. 27.87) and (b) for all sections combined (23.97 vs. 42.71) (see Table 4.)

In addition, *t*-tests comparing the percentage of passive verbs across sections, regardless of field (see Table 5), demonstrated the following pattern of results: Discussion (21.10) < Introduction (24.92) < Results (32.33) < Method (63.57). All differences except those between Introduction and Discussion were significant. In addition, differences between the means of both objective sections (Method/Results) and interpretative sections (Introduction/Discussion) were also significant (22.13 vs. 46.10, respectively).

Table 3. ***Percentage of passive verbs used in three scientific fields (psychology, botany, and chemistry/physics)***

Section	Psychology ($n = 12$)		Botany ($n = 12$)		Chem/Physics ($n = 12$)		All articles ($n = 36$)		p[1]
	$\bar{x}$	*SD*	$\bar{x}$	*SD*	$\bar{x}$	*SD*	$\bar{x}$	*SD*	
Introduction	17.63	11.38	24.88	12.65	32.25	11.69	24.92	13.06	.019
Method	43.11	16.02	80.43	17.21	67.16	17.63	63.57	22.73	<.001
Results	26.95	16.08	28.22	18.34	41.82	11.71	32.33	16.63	.048
Discussion	14.86	4.35	20.59	8.37	27.87	11.17	21.10	9.82	.003
Entire article	23.97	5.46	34.69	10.91	42.71	12.53	32.91	11.34	<.001

[1] Probabilities for differences in means in the first three columns (Psychology, Botany, Chemistry/Physics)

Table 4

Percentage of passive verbs across fields and article sections

Comparison	Psychology vs. Botany			Psychology vs. Chemistry/Physics			Botany vs. Chemistry/Physics		
	$\bar{x}_1$	$\bar{x}_2$	p	$\bar{x}_1$	$\bar{x}_2$	p	$\bar{x}_1$	$\bar{x}_2$	p
Introduction	17.62	24.87	.146	17.62	32.25	.005	24.88	32.25	.139
Method	43.11	80.43	<.001	43.11	67.16	.001	80.43	67.16	.064
Results	26.95	28.22	.844	26.95	41.82	.026	28.22	41.82	.041
Discussion	14.86	20.59	.106	20.59	27.87	.042	20.59	27.87	.042
All sections	23.97	34.69	.014	23.97	42.71	< .001	34.69	42.71	.060

Table 5
Percentage of passive verbs used across article sections regardless of field

Comparison of sections: $\bar{x}_1$ *versus* $\bar{x}_2$	Means $\bar{x}_1$	$\bar{x}_2$	*t*	*p*
Introduction vs. Method	24.92	63.57	10.20	< .001
Introduction vs. Results	24.92	32.33	2.71	.010
Introduction vs. Discussion	24.92	21.10	1.67	.105
Method vs. Results	63.57	32.33	7.99	< .001
Method vs. Discussion	63.57	21.10	12.52	< .001
Results vs. Discussion	32.33	21.10	4.13	< .001
Interpretative vs. Objective sections[1]	22.13	46.10	9.90	< .001

[1] Combined Introductions/ Discussions vs. combined Methods/Results

Table 6 shows the categorization of all of the passive verbs in four of the articles from each category (n = 12) as full vs. truncated and reversible versus non-reversible.[61] These results illustrate that almost 100% of the passive verbs in the articles examined were both truncated and non-reversible, e.g., *Seeds of cheat were **germinated** on wet ashless filter paper wetted with double-deionized water in petri dishes* (Gali & Smith, 1992).[62]

In addition, only eleven of the passive constructions (1.5%) were cleft passives (e.g., *It was hypothesized that*). The rarity of such cleft constructions is interesting in light of the fact that they are so often cited as examples of "bad" passives (see, e.g., Day, 1994, p. 166; Matthews, Bowen, & Matthews, 1996, p. 98).

---

[61] See previous chapter for a discussion of truncation and reversibility.

[62] Examples of the other categories are as follows: (1—truncated reversible) *The subjects were **asked** to sort the cards into two piles of three cards each* (Beatty, 1993); (2—non-truncated non-reversible) *The magnitude of viscous damping, which **has been studied carefully by other workers**, depends on the viscosity and density of the solution* (Redepenning, Schlesinger, Mechalke, Puleo & Bizios, 1993).; and (3—*We **are encouraged by the work of Nevins et al.** in which they succeeded in minimizing problems associated with long-term drift of the oscillator during their studies involving bacterial films* (Redepenning et al.). Please note that truncated passives can be analyzed for reversibility if the implied actor can easily be determined, as it usually can in scientific prose, when the actors are usually the authors or other researchers.

Day even characterizes such constructions as "typical" (p. 166) in scientific writing. But the results of this study suggest that such uses are the exception, not the rule.[63]

Table 6
Incidence of truncation and reversibility in 751 passive verbs from 12 scientific articles

		Truncation: Absent	Truncation: Present	
Reversibility	Not reversible	2	740	742
	Reversible	1	8	9
		3	748	751

Discussion

These results demonstrate that passive constructions are common in scientific writing—approximately one-third of all verbs in the articles examined were in the passive voice. The incidence of passive usage appeared to be related to what I have termed field objectivity (psychology < botany < chemistry/physics). For all three fields, the section order of lowest to highest incidence of passive verbs was the same as that observed by Riley (1991) (Discussion < Introduction < Results < Method). Thus, all three hypotheses regarding the incidence and pattern of passive usage in scientific articles were confirmed.

The higher number of passives in objective fields and objective sections of scientific articles is hardly surprising. As Gross (1985) observed, "Methods . . . and Results emphasiz[e] . . . the objective over the subjective, the events of the laboratory

---

[63] This very low incidence of cleft passives contrasts with the higher incidence of cleft passives ($\bar{x} = 7.5\%$) found in eight scientific articles drawn from geology, chemical engineering, and physics by Rodman (1991). However, as she points out, one of the articles had an unusually high number of cleft passives (31); when that article is eliminated from the analysis, the mean number of cleft passives is 4.14%, which is still somewhat higher than the 1.45% that I counted.

(p. 21). In these two sections, the passive maintains the focus of the discourse on the thing being described. But the use of passives to achieve this goal has evolved over time. Describing the role of passives in the evolution of scientific language, Coetzee (1980) observes that the sentence *Lime deters insects* would be more common during the time of the Royal Society (1661) while *Insects are deterred by lime* would probably be more common today. Coetzee points out that the active version focuses too much attention on lime: "This 'activity' of lime is a red herring, hinting distractingly at animistic powers in the substance" (p. 215).[64]

Another result of this study was that, as predicted, scientists rely almost exclusively on passives that are both truncated and non-reversible—not the full, reversible passives often cited as bad examples in textbooks and used as stimuli in active-passive experiments. Most writers also used potentially awkward cleft passives very sparingly—at the mean rate of about one per article. Thus, it seems that, although the authors of these scientific articles used passive verbs when they wrote, most used them intelligently: to focus attention on important factual information such as procedures and experimental results. Passive verbs were used less frequently in literature reviews or in the discussion of experimental results. Also, the absence of full passives and cleft passives indicates that few scientific writers compose sentences such as *Lactate was produced by S. aureus* or *It was found that.*[65]

The results of this study seem to indicate that most scientific writers and editors are already using passives in the way that style guide writers recommend. Riley (1991) arrived at the same conclusion, noting that the authors of the speech pathology articles she assessed

> neither avoided the passive voice (as many textbooks advise) nor used it indiscriminately (as the stereotypical view of scientific writing would have us believe). Instead, the writers varied their use of passive and active voice in rhetorically significant ways (p. 253).

---

[64] This illustration also points out the fact that we cannot assume—as many textbooks authors seem to—that casting sentences in the active voice guarantees a "livelier" sentence, either because the verb is "livelier" or because the subject is the actor. *Lime deters insects* is a good example of an active voice construction which lacks an actor, if we think of an actor as an entity that possesses volition (or, at a minimum, animacy) Another example is the first clause in this footnote.

[65] These examples are drawn from Day (1994).

The current study extended Riley's research in establishing that scientific writers make liberal use of passive verbs, and do so to a greater extent in objective article sections and in object-oriented fields. It also showed that the incidence of potentially problematic passive constructions such as cleft passives and passives that are both full and reversible is very low, a finding that (a) supports the proposition that scientific writers use passives in rhetorically-approprate ways and (b) limits the generalizability of voice studies that used full reversible passive sentences as stimuli.

Study 2:
Comprehensibility of Empirical Articles Written in the Active versus Passive Voice

Although passives have been examined in isolated sentences and in short, narrative passages, no study to date has examined reading speed and comprehension for conventionally-written scientific articles that vary only in voice (high-active vs. high-passive).

I performed two studies designed to assess the effect of voice on reading speed and comprehension of scientific articles. If the information presented in "high-passive" articles takes longer to encode than that presented in "high-active" articles, then we would expect to see longer reading times for passive articles. If the information presented in "high-passive" articles is less thoroughly understood, then we would expect lower reading comprehension scores by readers of passive accounts. If, however, voice exerts little influence on encoding, processing, or retrieval, then we would expect to see no difference in either reading times or comprehension scores.

STUDY 2A

Method

Subjects. Subjects were 267 subjects participating in the University of Washington Psychology Subject Pool. Participants received course credit in introductory psychology courses for participating in the experiment.

Materials. To assess the comprehensibility of voice (active vs. passive) in scientific research articles closely resembling published accounts, two different articles on psychology topics were adapted from published studies. Article X (*Counterfactual Thinking*), adapted from Rhodes (1988), focused on the effect of counterfactual information on judgments of liability and damage awards in a civil suit scenario; Article Y (*Mood and Memory*), adapted from Parrott & Sabini (1990), examined the effect of non-induced mood on memory retrieval in a classroom setting. Although neither article required any special content knowledge to understand, the specific

research they discussed would not likely be covered in an introductory psychology class such as the one from which subjects were recruited. The article adaptations were matched as follows:

Counterfactual Thinking, active voice	1593 words, Flesch-Kincaid Grade Level:12
Counterfactual Thinking , passive voice	1625 words, Flesch-Kincaid Grade Level:12
Mood and Memory, active voice	1543 words, Flesch-Kincaid Grade Level:12
Mood and Memory, passive voice	1573 words, Flesch-Kincaid Grade Level:12

The average length by voice was 1568 for the active versions and 1596 for the passive versions (the passive versions were slightly longer because of the necessity to add auxiliary verbs (e.g, *gave* vs. *were given*) and the necessity of ensuring that content for both articles was as close to identical as possible. Thus, in some cases, *by*-phrases were added in the passive versions (e.g., *Task outcome . . . was controlled by the experimenters)* to match the active version (*The experimenters controlled task outcome)*. Many of these short additions would not have been necessary to ensure comprehensibility. Indeed, if anything, they probably diminished comprehensibility slightly. But they were necessary in this experiment to ensure the closest possible content match for the active and passive versions of the article. The slightly longer length of the passive versions should not be taken as evidence that passive passages are wordier than active passages.

In the "high-active" version of each article, 100% of all verbs were in the active voice; in the "high-passive" version, 58% of verbs were in the active voice, 42% in the passive voice.[66] Each subject read one "high-active" article and a different "high-passive" article; order was counterbalanced, which resulted in four conditions. Subjects answered 16 five-alternative multiple-choice questions designed to test each subject's understanding for each article just read.[67] Questions for both active and

---

[66] The "high-passive" version was operationalized as 42% passive based on the results from Study 1, which revealed a mean incidence of passive verbs ranging from 24% (psychology) to 43% (chemistry/physics). I wanted to ensure that the high passive version was as high as one would expect to see in most scientific fields.

[67] Multiple choice questions are the most common method to assess reading comprehension, although their use has been criticized as old-fashioned because they focus on literal rather than inferential comprehension (see Farr & Carey, 1986, for a discussion). But as Faist (1984) points out, what constitutes an appropriate measure of reading comprehension depends largely upon the goals of the research. My purposes here were to assess fairly literal comprehension that was suitable for use with

passive articles were always presented in the active voice to ensure that any possible bias resulting from similarities in voice between question voice and article voice operated in favor of the active voice. Subjects also completed a questionnaire in which they were asked to rate the comprehensibility of the each article and the interestingness of the article topic on a 9-point scale where 0 = very low and 8 = very high.

Procedure. Subjects were randomly given one of four packets, each of which contained materials presented in the following order: consent form, instructions, first article, first article questions, second article, second article questions, demographic questionnaire. Subjects additionally received verbal instructions explaining the general purpose of the experiment (to test the comprehensibility of scientific articles, not their individual reading ability) and how to record reading time using the experimenter's clock (which gave the time in hours, minutes, and seconds). They were instructed to read the article carefully once through. They were told that they would be tested on its contents, and that they would have as long as they needed to complete the task, and that their performance on the test would in no way affect their course grade. After subjects read one article, they answered 16 five-alternative multiple choice questions on it; this procedure was then repeated for the other article. Then they filled out a demographic questionnaire.

Design. The effects of three manipulated variables on reading comprehension and reading time were assessed. The three variables were voice (active and passive), article (X and Y), and order (first and second). This results in eight conditions (Article X active first, Article X active second, Article X passive first, Article X passive second, Article Y active first, Article Y active second, Article Y passive first, and Article Y passive second) although each subject participated in only one of the

---

a large-*n* sample; multiple choice questions provided the best available measure.

following four combinations of conditions:

1 Article X active first followed by Article Y passive
2 Article X passive followed by Article Y active
3 Article Y active followed by Article X passive
4 Article Y passive followed by Article X active

Because each subject participated in all the levels of each independent variable, it was possible to analyze the separate effects of each variable using a within-subjects approach (i.e., a *t*-test). However, it was not possible to look at interactions between variables using a conventional analysis of variance because combining these variables results in a 2 x 2 x 2 incomplete design (incomplete for all three of the variables). Thus, although each subject participated in exactly two cells of the eight-cell design, the within-subject characteristic cannot be assigned to any of the three independent variables if they are analyzed simultaneously.

Another way of explaining the same problem is to start by conceptualizing the design as a 2 x 2 instead of a 2 x 2 x 2. If both variables are between-subjects, there is one observation per subject (one observation per subject). If both variables are within-subjects, there is one observation per cell per subject (4 observations/subject). If the design is mixed (e.g., the row variable is within-subjects and the column variables is between-subjects), there are two observations in one row and none in the other (two observations/subject). The present design, however, has only two observations per subject, and these observations are on the diagonal, an arrangement which does not conform to the arrangement of data in a conventional between-subjects, within-subjects, or mixed design. In addition, if we add back the third variable to create a 2 x 2 x 2 design, then the two observations are on a three dimensional diagonal. Thus, although having more than one observation per subject is a characteristic of a within-subjects design, having those observations on the diagonal is not.

As a result of this data configuration, each main effect can be analyzed using a within-subjects *t*-test but interactions require between-subjects analysis of variance. However, the interactions are of minor interest in this experiment because of the three variables in the experiment—voice, order, and article—only voice is of real conceptual interest. Order and article are nuisance variables that were controlled by counterbalancing. However, the presence of a voice by article interaction might

indicate that the effects of voice depend upon article content. Thus, the independent effects of voice, article, and order were assessed using a within-subjects *t*-test; the effects of voice in combination with article were also assessed using a between-subjects analysis of variance.

Results

Within-subjects *t*-tests were performed in SPSS-PC+, v. 4, on three variables: order, article, and voice. Order of presentation did not affect reading comprehension scores ($\bar{x}_{1st} = 8.65$ vs. $\bar{x}_{2nd} = 8.78$, $t(266) = .75$, $p = .46$) but did affect reading times ($\bar{x}_{1st} = 8.14$ vs. $\bar{x}_{2nd} = 8.71$, $t(266) = 5.89$, $p < .001$), such that the second article was read more slowly than the first. Article content affected reading comprehension scores ($\bar{x}_X = 8.15$ vs. $\bar{x}_Y = 9.28$, $t(266) = 7.21$, $p < .001$), such that Article X scores were lower than Article Y scores; but it did not affect reading times ($\bar{x}_X = 8.35$ vs. $\bar{x}_Y = 8.50$, $t(224) = 1.50$, $p = 14$).[68] The variable of primary interest, article voice, affected neither reading comprehension scores ($\bar{x}_{act} = 8.82$ vs. $\bar{x}_{pas} = 8.61$, $t(266) = 1.28$, $p = .20$) nor reading times ($\bar{x}_{act} = 8.35$ vs. $\bar{x}_{pas} = 8.51$, $t(224) = 1.50$, $p = .14$).

A between-subjects analysis of variance was also performed to eliminate the possibility of an interaction between voice and article.[69] To perform this analysis, it was necessary to randomly eliminate the minimum number of subjects possible to create cells with equal numbers of subjects. Thus, seven of the original 267 subjects used in the reading comprehension analysis were randomly eliminated in order to create cells with 65 subjects each ($N = 260$); 19 of the original 225 subjects used in the reading times analysis were eliminated in order to create cells with 51 subjects each ($N = 204$). See Table 7 for reading comprehension means and Table 8 for reading time means.

For the between-subjects analysis, reading comprehension scores were

---

[68] Although $n = 267$ for this comprehension analysis, $n = 225$ for the reading time analysis due to the failure of 42 subjects to correctly record both their start and finish reading times.

[69] A between-subjects comparison was necessary because of the incompleteness of the within-subject design when both variables are conjointly analyzed.

significantly lower for Article X than the Article Y, regardless of the order in which the articles were read: for order = 1, $F(1, 256) = 7.31, p = .007$; for order = 2, $F(1, 256) = 16.99, p < .001$. However, scores were unaffected by voice: for order = 1, $F(1, 256) = .52, p = .47$; for order = 2, $F(1, 256) = .06, p = .80$. There was also no voice by article interaction: for order = 1, $F(1, 256) = .009, p = .92$; for order = 2, $F(1, 256) < .001, p = .98$.

Table 7
Effects of voice and article content on response accuracy for 16 five-alternative reading comprehension questions (N = 260)

Order = 1

Article	Voice: Active	Voice: Passive	Mean
X	8.40	8.14	8.27
Y	9.23	9.03	9.14
Mean	8.82	8.59	

Order = 2

Article	Voice: Active	Voice: Passive	Mean
X	8.15	8.08	8.12
Y	9.54	9.45	9.49
Mean	8.85	8.76	

Reading times were not affected by voice, regardless of order (for order = 1, $F(1, 200) = .71, p = .40$; for order = 2, $F(1, 200) = 3.10, p = .079$). Reading times were not affected by article content for the first article read (for order = 1, $F(1, 200) = 1.91, p = .16$) but were lower for Article X when it was read second (for order = 2, $F(1, 200) = 4.47, p = .036$). There was no interaction between voice and article regardless of order (for order = 1, $F(1, 200) = 1.91, p = .169$; for order = 2, $F(1, 200) = .746, p = .389$).

To ensure that subject performance was not the result of prior content knowledge, I performed a passage dependency test in which subjects who had not read the two articles were asked to answer the same questions. If the comprehension questions used in this study measured article comprehension, rather than prior knowledge, then subjects who have not read the articles should do no better than chance when asked to answer the questions without having read the articles (for 16 five-alternative questions, chance performance = 3.2 correct answers). This prediction

was confirmed. For both sets of 16 questions, the mean number of correct answers for 77 subjects was 3.05 for Article X and 3.17 for Article Y—not significantly different than the 3.2 expected by chance.

Table 8
Effects of voice and article content on mean article reading time in minutes (N = 204)

Order = 1

Article	Active	Passive	Mean
X	7.98	8.12	8.05
Y	8.59	8.00	8.29
Mean	8.28	8.06	

Order = 2

Article	Active	Passive	Mean
X	8.63	9.32	8.97
Y	8.29	8.53	8.41
Mean	8.46	8.93	

In addition to assessing reading speed and comprehension, I assessed the participants' subjective reactions to the articles. Subjects were asked (a) how interesting they found each article and (b) how comprehensible they thought each article was, where 0 = very low and 8 = very high. Although subjects found Article X more interesting than Article Y ($\bar{x}$ = 4.51 vs. 4.16, $t(264) = 2.32$, $p = .02$), they rated Article X and Y as equally comprehensible ($\bar{x}$ = 4.90 vs. 4.88, $t(264) = .28$, $p = .78$). Ratings were unaffected by article voice; subjects indicated that they liked active and passive versions equally well ($\bar{x}$ = 4.38 vs. 4.29, $t(264) = .57$, $p = .57$) and found them equally comprehensible ($\bar{x}$ = 4.90 vs. 4.90, $t(264) = .00$, $p = 1.00$).

Discussion

An initial analysis revealed that one of the articles (*Mood and Memory*) was better comprehended than the other (*Counterfactual Reasoning*), despite the fact that subjects liked the latter better and that they rated the articles as equally comprehensible. In addition, order of presentation affected reading times: the articles presented first were read more quickly than those presented second, perhaps because

subjects were slightly fatigued after reading and answering questions on the first article.

The results of the voice analysis did not support the proposition that passive voice passages (a) take longer to read and (b) are harder to comprehend, despite the fact that this experiment was designed to create a stringent test of this proposition, in that the "high passive" condition had an exceptionally high percentage of passive verbs (42%), the design was very powerful (i.e., large *n* and within-subjects analysis of voice effects), and the subjects were mostly freshman college students with limited experiences as consumers of scientific prose.

It should be noted, however, that the pattern of means for both variables favors active voice very slightly over passive voice. Using the t-test data (which uses the largest samples), active voice article comprehension scores were approximately 1% better than passive voice article comprehension scores; reading times for active voice articles were about 10 seconds faster for active voice versions than for passive voice versions. However, for reasons discussed earlier, the passive articles were just slightly longer than the active articles. If we compute the words per minute as a function of voice, the means are virtually identical: 187.78 wpm (active) vs. 188.11 wpm (passive).

None of the results obtained in this experiment support the proposition that using the passive voice in a scientific article decreases the extent to which readers like the article or their ratings of article comprehensibility. Neither did voice affect actual reading comprehension or reading speed.

STUDY 2B

This study was a close replication of Study 2a.. I performed this study in order to ascertain whether the results obtained in Study 2a were replicable using two entirely different scientific articles. For Study 2b, I used two articles focusing on animal rather than human behavior. The methodology was identical to that for Study 2a, with one exception: in order to examine the effects of voice on ease of encoding and subsequent recall, each comprehension question assessed recall for a fact specifically presented in either an active or passive clause. Thus, to answer each question (e.g.,

Where did the study take place?), subjects would have to remember information presented in either an active clause (*We performed the study in Norway)* or in a passive clause (*The study was performed in Norway*).

This approach enabled me to look more specifically at the link between the form in which information was presented (active or passive clause) and the accuracy of answering a question requiring correct recall of that information. In Study 2a, because no special effort was made to ensure that subjects were queried only on information presented in active vs. passive clauses, it would be possible to argue that this study shows only that a generally high incidence of passive verbs does not adversely affect overall reading comprehension. It fails to eliminate the possibility that passive voice might diminish reading comprehension only when specifically used to present information critical to answering the comprehension questions. Indeed, some of the questions (e.g., Question 2 for both articles) were clearly designed to tap the respondents' understanding of theories presented over the course of several paragraphs. The Study 2b methodology eliminates that possibility by asking subjects to recall specific facts presented in active versus passive clauses. The disadvantage of this approach is that it mandates the use of narrowly-focused, factual comprehension questions (those requiring subjects to access a specific piece of information that is presented only once). However, it can be argued that, unlike many types of writing, scientific articles tend to contain much narrowly-focused, factual information that is presented only once. Thus, questions that elicit one's recall of fairly detailed, non-redundant information are an appropriate measure of article comprehension.

If the information conveyed in passive voice clauses is encoded less well than that in active voice clauses, we would expect that subjects answering questions from passive passages would have more difficulty retrieving that information than subjects reading active passages, especially when the questions are in the active voice.

Method

Subjects. Subjects were 268 subjects participating in the University of Washington Psychology Subject Pool. Participants received course credit in introductory psychology courses for participating in the experiment.

Materials. Two articles on new topics were used for this study. The first article, adapted from Saetre, Fossnes & Slagsvold (1995), examined the relationship between physical characteristics of male pied flycatchers and their ability to care for their brood in the absence of the female. The second article, adapted from Forthman, Elder, Bakeman, Kurkowski, Noble & Winslow (1992), compared the success of traditional vs. innovative feeding methods in reducing stereotypic behavior among zoo bears. In all other respects, articles from Study 2b were closely matched with both each other and the articles from Study 2a (e.g., all articles had the same Flesch-Kincaid Grade Level), although the articles from Study 2b were slightly longer than those used for Study 2a:

Pied Flycatchers, active voice	1732 words, Flesch-Kincaid Grade Level: 12
Pied Flycatchers, passive voice	1742 words, Flesch-Kincaid Grade Level: 12
Feeding Bears, active voice	1725 words, Flesch-Kincaid Grade Level: 12
Feeding Bears, passive voice	1750 words, Flesch-Kincaid Grade Level: 12

The mean length by voice was 1728 for the active versions and 1746 for the passive versions. The percentage of active and passive verbs was the same as that used for Study 2A (100% active verbs for the "high active" versions and 42% passive verbs for the "high passive" versions).

In addition to rating article comprehensibility and how well they liked the article topic (9-point scale, 0 = very low, 8 = very high), participants were also asked to indicate which article sounded more scientific and to explain the reason for their response. My purpose was to ascertain whether they might think the high-passive articles sounded more scientific and, if so, the degree to which they were aware of voice as influencing their response in any way.

Procedure. The procedure was identical to that used for Study 2a, except that subjects were more carefully instructed to record their time in hours, minutes, and seconds, in order to decrease the number of subjects failing to record reading times for the articles.

Results

The analysis procedure was the same for Study 2b as for Study 2a. Within-subjects *t*-tests were performed to assess the effects of order, article, and voice on reading comprehension and reading times. There were 268 subjects for the within-subjects reading comprehension analysis but 251 subjects for the reading times within-subjects reading times analysis due to the failure of 17 subjects to provide complete information on reading times. Between-subjects analyses of variance were performed to assess possible interactions between voice and article; for the between-subjects analyses, subjects were randomly eliminated to create groups of equal sizes, resulting in a sample of 260 for the reading comprehension analysis and 244 for the reading times analysis.

The pattern of results for the *t*-tests was the same as that for Study 2a. Order of presentation did not affect reading comprehension scores ($\bar{x}_{1st}$ = 8.17 vs. $\bar{x}_{2nd}$ = 8.48, $t(267) = 1.55$, $p = .12$) but did affect reading times ($\bar{x}_{1st}$ = 9.05 vs. $\bar{x}_{2nd}$ = 9.48 , $t(267) =$, $p < .001$), such that the second article was read more slowly than the first. One of the articles—Article A (*Birds*)—elicited higher reading comprehension scores than Article B (*Bears*)($\bar{x}_A = 8.85$ vs. $\bar{x}_B = 7.80$, $t(267) = 5.58$, $p < .001$). But there was no difference in reading times attributable to article content ($\bar{x}_A = 9.18$ vs. $\bar{x}_B = 9.35$, $t(250) = 1.61$, $p = .11$).[70] Similarly, article voice elicited no differences in either reading comprehension scores ($\bar{x}_{act}$ = 8.44 vs. $\bar{x}_{pas}$ = 8.21 $t(267) = 1.21$, $p = .23$) or reading times ($\bar{x}_{act}$ = 9.21 vs. $\bar{x}_{pas}$ = 9.33, $t(250) = 1.09$, $p = .28$).

The analysis of variance also revealed results similar to those of Study 2a. Reading comprehension scores were unaffected by voice: for order = 1, $F(1, 256) = .19$, $p = .66$; for order = 2, $F(1, 256) = 2.73$, $p = .10$. Reading comprehension scores were lower for one article than the other, regardless of order: for order = 1, $F(1, 256) = 16.02$, $p < .001$; for order = 2, $F(1, 256) = 3.99$, $p = .047$. There was also no voice by article interaction: for order = 1, $F(1, 256) = .92$, $p = .34$; for order = 2, $F(1, 256) = .001$, $p = .97$. See Table 9.

---

[70] All but 17 of the 268 participants recorded all reading time information, resulting in a larger *N* for the reading time analyses ($n = 251$).

Table 9
Effects of voice and article content on response accuracy for 16 five-alternative reading comprehension questions (N = 260)

Order = 1

Article	Active	Passive	Mean
X	8.66	9.15	8.91
Y	7.58	7.40	7.49
Mean	8.12	8.28	

Order = 2

Article	Active	Passive	Mean
X	9.14	8.57	8.85
Y	8.45	7.85	8.15
Mean	8.79	8.21	

Reading times were not affected by voice, regardless of order: for order = 1, $F(1, 240) = 3.14$, $p = .08$; for order = 2, $F(1, 240) = .31$, $p = .58$). Neither were reading times affected by article content: for order = 1, $F(1, 240) = .33$, $p = .57$; for order = 2, $F(1, 240) = .58$, $p = .45$. There was no interaction between voice and article: for order = 1, $F(1, 240) = .09$, $p = .76$; for order = 2, $F(1, 240) = .003$, $p = .95$. See Table 10.

To ensure the passage dependency of the comprehension questions, 33 subjects answered the comprehension questions without reading the articles. For *Birds*, subjects answered 3.12 questions correctly; for *Bears*, subjects answered 3.30 questions correctly. Neither of these results were significantly different than the 3.2 expected by chance, demonstrating that the questions were indeed passage dependent.

The pattern of results for participant ratings was identical to that observed in Study 2a. The 265 participants who completed the ratings found one of the articles more interesting than the other—in this case, the *Bears* article ($\bar{x} = 4.18$ vs. 3.36 for *Birds*, $t(264) = 5.78, p < .001$). But they found both articles equally comprehensible ($\bar{x}_A = 4.83$ vs. $\bar{x}_B = 4.77$, $t(264) = .52, p = .61$). As in Study 2a, voice did not affect ratings of either interestingness ($\bar{x}_{act} = 3.73$ vs. $\bar{x}_{pas} = 3.82$, $t(264) = .55, p = .58$) or comprehensibility ($\bar{x}_{act} = 4.75$ vs. $\bar{x}_{pas} = 4.85$, $t(264) = .86$, $p = .39$).

Table 10
Effects of voice and article content on mean article reading time in minutes (N = 244)

Order = 1

Article	Active	Passive	Mean
X	8.83	9.16	8.99
Y	8.89	9.36	9.12
Mean	8.86	9.26	

Order = 2

Article	Active	Passive	Mean
X	9.48	9.36	9.40
Y	9.66	9.54	9.60
Mean	9.57	9.43	

In answer to a question added for Study 2b (i.e., *Which of these articles sounded more scientific?*), 40% chose Article A (*Birds*) and 20% chose Article B (*Bears*); 40% thought there was no difference. Although subjects were given an opportunity to give reasons for their responses, none of the 160 subjects who thought one article sounded more scientific than the other cited voice as a factor (few subjects cited any reason). Forty-six percent thought the active version sounded more scientific; 54% thought the passive version sounded more scientific; this difference was not significant.

Discussion

The pattern of results in Study 2B was very similar to that observed for Study 2A: a high incidence of passive voice verbs neither increased reading times nor decreased comprehension scores. However, as in Study 2A, one of the articles elicited lower comprehension scores than the other. For Study 2B, directly tying the article comprehension questions to information presented once in either active or passive voice—a manipulation designed to make voice more salient—did not affect the pattern of results in any detectable way. The only difference between these results and those of Study 2A was that there were no effects for reading time, regardless of article or order of presentation.

General Discussion

In these two studies, the use of large samples ($n = 267$ for Study 2a; $n = 268$ for Study 2b) and a within-subjects approach was intended to create an experimental design with sufficient power to detect any meaningful differences in reading speed and reading comprehension resulting from the manipulation of article voice, should any exist. However, the only significant differences observed was for reading comprehension, and this was due to article differences,[71] not voice. There were no effects of voice even in Study 2b, where comprehension questions were specifically tied to information presented only once in either the active or passive voice. In both studies, both the reading time and comprehension measures were potentially biased in favor of the active voice in that the active articles were slightly shorter than the passive articles and the comprehension questions were phrased in the active rather than the passive voice.

These results are different with those obtained by Coleman (1964). Manipulating voice, nominalizations, and adjectivalizations, Coleman found that revising two passages by converting passive to active verbs, nominalizations to verbs, and adjectivalizations to verbs enhanced reading comprehension as measured by higher scores on multiple choice comprehension questions. However, he did not manipulate voice in isolation in his study. When I did so in this study, I found no effects due to voice.

My results call into question the common assumption that active verbs require less cognitive effort than passive verbs. Although the limits of hypothesis testing as a methodology make it impossible to unequivocally state that no difference exists, the power of the experimental design does permit me to argue that any actual, but undetected, differences in reading speed or comprehension stemming from voice manipulations must be slight.

One potential weakness of these two studies concerns the type of questions used to assess reading comprehension. The questions were quite specific—some could even be considered picky in that they demanded a fairly detailed recollection of

---

[71] Alternatively, differences in reading comprehension scores could have been due to to differences in the difficulty of the comprehension questions used for each article.

specific points. There were a number of reasons for using detailed questions. First, to avoid ceiling effects, the questions needed to be of at least medium difficulty. Second, to ensure that each question had one and only one correct answer, I needed to ask questions for which the answers were extremely specific. In defense of these questions, I think that such questions are more appropriate to assess the comprehension for scientific articles than for some other materials (e.g., Shakespearean plays or newspaper editorials) because remembering specific details about the procedure, design, etc., is often critical in scientific articles.

One limitation of these studies is that they examined the comprehensibility of scientific prose on educated non-scientists (undergraduates in entry-level psychology courses). They did not examine the effect of voice on scientific readers, i.e., scientists (a) in general, (b) within a specific field (e.g., psychology), or (c) who are working on the same line of research. However, this weakness is in some sense a strength, in that it is likely that college freshman would find passives harder to understand than scientists who have greater familiarity both with scientific articles and with the content of those articles. Thus, if even the freshman did not have measurable problems with passives, it is unlikely that the scientists would have problems.

Another limitation is that the studies used article adaptations rather than the actual articles. Although the adaptations closely resembled actual articles—and were indeed based upon four articles—they were somewhat shorter and less complex than the originals. This was due to several factors. First, real scientific reports tend to contain complications such as multiple citation strings, extremely complex data analyses, and references to theories or methods that are not fully explained—and which are not readily understandable—to non-specialists. Second, I needed to create two articles that were matched in length, difficulty, and number of passive verbs. To do so, I needed to adapt a number of sentences so that they could be expressed both in the active and passive voice without either construction sounding unduly awkward. Last, despite the best efforts of both writers and editors, not all scientific articles are well-written—topic sentences do not always encapsulate the author's paragraphs and sentences do not always overlap as theorists such as Haviland and Clark (1974) recommend. For this initial investigation, I wanted to examine the effects of voice in

articles that were maximally cohesive and coherent, so I was careful to create tight-knit adaptions. In the future, it would be desirable to assess how well naturally-occurring articles are comprehended by the scientific audiences for whom they are intended.

A third limitation was the use of multiple choice questions to measure reading comprehension. Although this method is a traditional means of testing reading comprehension, it provides only an indirect means of assessment. A more direct, process-oriented method (e.g., eye tracking, see Just and Carpenter, 1979) or a secondary reaction time task (see Britton, Glynn, Meyer, and Penland, 1982) might provide a potentially more sensitive and direct measure of cognitive effort. But such methods are extremely labor intensive and thus are not well-suited for the kind of large-*n* study I wanted to perform.

A related limitation mentioned earlier is the use of questions aimed at assessing subjects' recall for specific facts as opposed to subjects' understanding of the broader theories or related ideas implied by these facts. But the advantage of the approach used is that the correct answer to the question was unambiguously correct. This was particularly important in Study 2b, which targeted recall for information presented in specific clauses that were either active or passive. In defense of this approach, however, I would argue that understanding a scientific article requires more than simply understanding the author's gist. It requires the understanding of details, because it is these details that determine the validity and generalizability of the work. Thus, multiple choice questions testing literal recall are particularly well-suited for testing a reader's understanding of scientific articles. Also, their use permitted me to compare my results with those of Coleman (1964), who also used multiple choice questions to assess reading comprehension.

Study 3:
Evaluation of Text Variables in
Strong versus Weak Student Lab Reports

Study 1 demonstrated that passives are widely-used in published scientific writing and Study 2 demonstrated that scientific articles with many passive verbs can be read as quickly and understood as well as their active voice counterparts. However, the articles examined in these studies were written by experienced scientific authors; the published articles were also professionally edited.

But how does this research apply to scientific articles written by less experienced authors? While experienced writers can successively use passives to communicate, perhaps novice writers are less successful. This idea has intuitive appeal; any composition teacher can think of numerous anecdotal examples of awkward passives in student writing. And of course many of the stylists that counsel writers to rely on the active voice are primarily directing their comments to inexperienced writers. If novice writers take the advice proffered, their writing should improve, assuming the advice is accurate. Thus, students who write the best lab reports might be expected to use fewer passives than students who write the worst reports.

To look more closely at this issue, I compared the incidence of passives in strong versus weak laboratory reports written by psychology majors in a scientifically-oriented psychology program. I also examined the incidence of other quantitative prose characteristics that might affect prose quality (e.g., number of studies cited) and used subjective quality ratings to compare factors such as organization, cohesion, and content.

Method

Subjects and Procedure. During Spring Quarter 1996, students enrolled in two psychology laboratory courses (Psych 231—Laboratory in Human Performance and 232—Laboratory in Animal Learning) were informed that researchers from the Psychology Writing Center would be conducting a study on writing in psychology. They were told that the study would focus on a number of aspects of writing,

including strengths and weaknesses of student papers. Students were asked to indicate whether they would be willing to allow their papers to be examined as part of the study by indicating their decision on a written consent form.

A list of students willing to participate in the study was given to course instructors (four instructors, two from each of the two lab courses) each of whom selected several eligible papers which they judged to be weak, typical, or strong. All lab papers meeting the criteria for inclusion in the study (i.e., those which were individually written and which used conventional scientific format—Introduction, Method, Results, and Discussion)—in the weak and strong categories were selected for the current analysis, resulting in a sample of 19 papers (7 weak and 5 strong papers from Psych 231; 4 weak and 4 strong papers from Psych 232).[72]

Selected papers were analyzed both quantitatively and qualitatively. For the quantitative analysis, all the text in each paper was examined to determine the incidence of the following text characteristics: (a) percentage of passive verbs (the overall percentage and percentage by article section); (b) paper length,[73] (c) number of studies cited in the reference section, (d) number of citations in the text of the paper, and (e) number of American Psychological Association (APA) citation violations. For the qualitative analysis, quality ratings were performed on just the Introduction and Discussion sections to ascertain the quality of (a) scientific content development (i.e., quality of eight specific content "chunks," four in the Introduction and four in the Discussion[74]); (b) scientific organization of the Introduction and Discussion (degree to which content chunks occurs in the right order); and (c) sentence cohesion of the Introduction and Discussion (degree to which adjacent sentences provide adequate

---

[72] Only strong and weak papers were analyzed, both because of the small sample size and my suspicion that there would be more variance in instructor ratings of "typical" papers than in strong and weak papers, a suspicion that was confirmed in a primliminary investigation.

[73] Operationalized as number of total verbs in paper.

[74] I assessed these two sections because the rhetorical demands they impose are much greater than those imposed by Method and Results, which are more formulaic in nature. For the Introduction, I assessed the quality of the research question, previous research, interpretation of previous research, and hypotheses; for the Discussion section, I assessed the quality of the relationship of results to hypotheses, implications of results, potential study weaknesses, and directions for future research.

content overlap).[75] For the quality analysis, papers were rated on a 0 - 4 scale, where 0 = very weak, 1 = weak, 2 = fair, 3 = good, 4 = very good.

Initial ratings and tallies done by me indicated major differences between weak and strong papers on all key variables except voice. However, because I was familiar with the conditions of the study, I trained an independent rater blind to the conditions of the study to rate the 20 studies. I subsequently computed correlation coefficients for our ratings, which ranged between .57 and .70 for the nine content variables and two organizational variables; all of these correlations were statistically significant. Correlations were lower (and non-significant) for only two of the 13 variables: sentence cohesion for Introduction sentences (r =.24) and for Discussion sentences (r = .35) (see Appendix F for a summary).

Results

T-tests comparing strong and weak lab reports revealed significant differences in all eight areas of content development (four in the Introduction, four in the Discussion); organization of both Introductions and Discussions; sentence cohesion of both Introductions and Discussions, paper length (as measured by number of total verbs); and incidence of APA citation violations. Differences between number of studies cited as references and total citations per paper achieved near-significance ($p < .10$) . All differences were in favor of stronger papers (see Table 11). For only one set of variables were there neither significant nor near-significant differences between weak and strong papers: percentage of passive verbs (either overall or by article section).

The overall percentage (percent of passive verbs in the entire paper) was remarkably similar for weak and strong papers: 28.03 (weak) vs. 29.62 (strong). By section, the range of differences was small (1% - 7%). In all cases, stronger paper used a slightly higher percentage of passive verbs than weak papers, but these differences did not approach significance.

---

[75] I examined just the Introduction and Discussion sections for several reasons. One is that the type of analysis performed is extremely labor intensive. But the main reason is that these two sections are clearly the most difficult to compose; the Method and Results sections are much more formulaic. Thus, an examination of the Introductions and Discussion yields a superior measure of writing ability.

Interestingly, the pattern of passive verb usage was the same as that of the published scientific articles in Study 1: (a) the rank order of passive verbs by section was the same (Discussion < Introduction < Results < Method); and (b) the mean percentage of passive verbs used was 28% (similar to that of published articles in psychology, $\bar{x} = 24\%$).

Discussion

In this descriptive study, lab reports rated as strong by four instructors in two courses were significantly different from weak papers on numerous dimensions. Strong papers had better content development, organization, and cohesion; they were also longer, listed more studies in the Reference section, and contained fewer APA violations. .They did not, however, differ in the number of passive constructions used either globally or in individual paper sections. Thus, although a great number of factors accounted for differences in quality ratings of "weak" and "strong" by instructors, voice did not appear to be one of them.

The percentage of passives in both strong and weak papers was similar to the percentage of passives in papers written by experienced psychology researchers. This similarity in passive usage is probably due to the fact that most of these report writers were seniors—students who had read quite a few scientific papers by the time they produced their lab reports. It seems likely that, whatever their other strengths or weaknesses as writers, most of them had learned that writing like a scientist entails using the passive voice, not indiscriminantly, but in the appropriate context. Even the weak writers at least tacitly understood that passives are "supposed" to appear in Results and Method and active verbs in the Introduction and Discussion.[76]

---

[76] Indeed, if students were actually aware that they were using so many passive verbs, they would probably try to get rid of them (because they have been told that passives are bad). It is only because most of them can't recognize the difference that they are able to freely model what they have seen in published reports.

Table 11
Ratings of lab reports in two psychology lab courses (Psych 231, 232) where ratings range from 0 (very poor) to 4 (very good)

Text Characteristic Ratings (Qualitative Analysis)	Weak (n = 11)[1]	Strong (n = 9)[1]	t	P
Content variable—paper title	1.98 (.84)	2.72 (.48)	2.49	.024*
Content variables—Introduction:				
Rating of research question	2.02 (.99)	2.89 (.87)	2.08	.052
Rating of past research	1.84 (.79)	3.31 (.67)	4.50	<.001*
Rating of implications	1.45 (.67)	2.89 (.66)	4.79	<.001*
Rating of hypothesis	1.77 (.98)	3.03 (.68)	3.36	.004*
Content variables—Discussion				
Rating of tie to hypothesis	1.50 (.60)	3.28 (.51)	7.17	<.001*
Rating of data interpretation	1.47 (.53)	3.03 (.46)	7.01	<.001*
Ratings of study limitations	1.25 (.88)	2.89 (.98)	3.90	.001*
Ratings of future study	1.48 (.80)	2.81 (.89)	3.47	.003*
Organization—Introduction	2.30 (.95)	3.58 (.35)	4.17	.001*
Organization—Discussion	1.95 (.77)	3.30 (.50)	4.92	<.001*
Sentence cohesion—Introduction	2.43 (.41)	2.92 (.31)	2.97	.008*
Sentence cohesion—Discussion	2.38 (.44)	2.87 (.25)	3.10	.007*
Text Characteristic Tallies (Quantitative Analysis)				
No. of studies in Reference list	3.55 (2.54)	8.11 (6.72)	1.93	.083
No. of citations in paper, including duplicates	4.73 (3.38)	13.00 (13.11)	1.84	.099
No. of APA citation violations	2.55 (1.13)	1.00 (.87)	3.46	.003*
Paper length (total no. of verbs)	82.00 (37.14)	130.56 (37.52)	2.89	.010*
Percentage of passive voice verbs (entire paper)	28.02 (8.60)	29.62 (12.80)	.32	.752
Introduction	20.24 (7.60)	24.38 (14.50)	.77	.455
Method	44.70 (24.90)	50.89 (23.30)	.57	.573
Results	28.56 (20.90)	32.56 (19.70)	.44	.663
Discussion	20.12 (12.60)	21.07 (14.10)	.16	.877

[1] Standard deviations are in parentheses.

However, weaker writers may have misused passives in less obvious ways—in convoluted constructions (e.g., *Subjects were to be given instructions*), pretentious comments (*It was found that the results were significant*), or confusing voice switches (e.g., *We passed out the materials and then subjects were given instructions).* Stronger writers may have avoided such pitfalls, instead using passives appropriately—to emphasize key points, to de-emphasize themselves as actors, or to enhance sentence cohesion. If so, weak writers may have produced weaker papers as a result of using the passive voice while strong writers may have produced stronger papers. Thus, it is possible that weak or inexperienced writers might produce better papers by using the active voice, just as speakers using a foreign language for the first time often communicate best when speaking in short, simple sentences.

For these reasons, I would hesitate to say that these results demonstrate that voice plays no role in determining the quality of student papers. However, the fact that strong and weak papers differed in so many respects, but not in the incidence of passives, certainly calls into question the idea that encouraging student writers to using the active voice will generally result in better student papers. Future studies could use a larger sample of papers and sample lab reports from a variety of scientific disciplines.

General Discussion

"All things being equal, readers more easily understand the active voice than the passive" (Rowan, 1989, p. 173). Early studies on the passive voice seemed to bear this statement out. But the assumption of "all things being equal" is a big one. Seldom are all things equal in written prose—contexts, topics, emphases, and readers vary. Later studies on voice showed that their effects are extremely context-sensitive, demonstrating the need to examine passives in the context in which they are actually read. However, none of the studies done to assess the effects of passives on processing difficulty examined the use of passives in scientific writing and, in particular, in "stand alone" scientific research reports.

The investigation documented here examined the use of passives in a scientific context: (a) their incidence in published scientific papers, (b) their effect on reading speed and comprehension in experimentally-manipulated reports, and (c) their incidence in weak vs. strong student lab reports. I found that (a) published scientific articles have a relatively high incidence of passive constructions ($\bar{x} = 32\%$), (b) such constructions produced no negative effects on either reading comprehension nor reading speed in two experiments (even when comprehension scores depended specifically on subjects' understanding of information presented in active vs. passive sentences), and (c) voice did not affect instructor ratings of student lab papers; weak and strong psychology student lab reports contained an almost identical mean percentage of passive verbs—29% vs. 28%—which closely matches their in published psychology papers (24%).

The results for these three investigations may help resolve some of the ongoing disputes about the advisability of using passives in scientific and technical prose. Writers such as Lynch and Chapman (1980) assert that "scientists often argue that methods and results can only be reported in the passive, but the only support for this viewpoint is tradition" (p. 146). But others have taken a different view. Analyzing a Method section from a scholarly botany journal, Martin (1993) defends the use of passive voice in the Method section:

> Had these sentences been written in the active voice, then the scientists conducting the experiment would have been [the] topical Theme in

> every clause. But the text is not about scientists; it is about leafy shoots in a rubber tube (p. 194).

Others have defended the use of well-crafted passives in scientific and technical writing (e.g., Scriven, 1989; Walpole, 1979; Wilkinson, 1992), but have had little empirical evidence with which to buttress their arguments. The results obtained in this investigation provide that evidence.

The results of the current investigation suggest that there is no cognitive justification for generally discouraging the use of passive constructions in any section of a scientific report. The results argue for a re-evaluation of the notion that passive voice is an inherently problematic construction that writers should simply be told to avoid. Despite its currently negative reputation in guides to scientific writing, passive voice enjoys wide use in scientific articles and does not appear to cause detectable problems with article comprehension. Neither is it seen more frequently in weak vs. strong student lab reports.

This study highlights a discrepancy between what style manuals advise scientists and students to do when writing scientific reports—eschew passives—and what writers actually do—to use passives in rather predictable ways in rather predictable locations within scientific articles. As pointed out in Study 1, this discrepancy between writing advice and actual usage is by no means unique. Although composition instructors routinely encourage students to write simply and directly, Hake and Williams (1981) performed a series of experiments demonstrating that high school and college English teachers asked to evaluate nominalized versus simplified passages strongly preferred the nominalized versions, despite the fact that the simplified versions were easier to understand, as evidenced by the fact that they were typed faster and more accurately than the nominalized versions. The authors wryly observe that prolix writing may be the norm as long as writing teachers discourage what they claim to admire and encourage what they claim to deplore.

Source of Misconceptions about the Passive Voice

If the findings of the current investigation reflect how scientific papers are actually read and written—i.e., if the scientific passive is widely used and understood as well

as the active voice—how has it come to acquire such a poor reputation among both scholars and the general public? Below I discuss seven factors that have contributed to this poor reputation, both of the passive in general and the passive as used in scientific documents.

(1) **The term *passive* is ambiguous.** "Passive" can refer to either verbs in the passive voice or to non-dynamic (stative) verbs in the active voice. Although most people who have attended junior high are familiar with the term "passive voice," only a small percentage understand what it means.[77] Many people confuse the generic term "passive" (meaning non-dynamic, static or unmotivated) with the passive voice, which has a much more specific meaning. As a result, they tend to equate stative verbs (e.g., *is, are, seem, believe)* with passive voice verbs and action verbs with the active voice. (This is ironic in light of the fact that stative verbs cannot by definition be cast into the passive voice.) That this kind of confusion should prevail among non-linguists is not surprising, since the term "passive" is inherently ambiguous. What is surprising, however, is that some textbook writers and psycholinguistic researchers blur this distinction between properties of verbs and properties of sentence construction.

For example, Flesch & Lass (1949) claim that writers who use too many passive constructions risk making their writing "dull and uninteresting, instead of making it interesting by using live, kicking verbs" (p. 172), thus implying that passive constructions use verbs that are less action-oriented than their active counterparts. Gunning (1968, p. 108) presents a problematic paragraph containing the clause *The decision has been rescinded*; the revised version, which "restores action" to the paragraph, replaces this clause with a sentence that begins thusly: *Members of the Boat Producers Association decided . . . . .* The revision is certainly an improvement, but not only the verb, but the voice, has been changed, which will lead some readers to confuse action verbs with active voice constructions.

The above examples are not unique. Coleman (1965) describes the effects of denominalization (converting nouns into their corresponding verbs counterparts, e.g.,

---

[77] At least half of the students who take the style course I teach (Style in Technical Writing) routinely confuse stative verbs with passive verbs *after* receiving instruction on the passive voice.

investigation > *investigate*) as "transforming them to simpler forms using active verbs" (p. 332), but his denominalized sentences include both active voice and passive voice verbs. Klare and Buck (1954) advise readers "to write actively rather than passively. Active writing is characterized primarily by a large proportion of verbs compared to adjectives or nouns." (p. 125). Twenty three years later, Klare (1977) compares the "active verb versus [its] nominalized form" (p. 150), although he discusses active and passive *voice* in the same article, in a section that begins with the statement that "the active form of a statement leads to easier recall and verification than the passive form" (p. 151). Booher & Hill (1989) also juxtapose a discussion of "passive style" with a discussion of "passive voice," referring to both verbs that are stative and those in the passive voice as "passive": "a passive style, though perhaps difficult to define, is . . . easy to recognize. Those who write in a passive style bury the meat of their ideas in passive verbs" (p. 113). In this quote, the authors are referring simply to weak verbs, not those in the passive voice. Several pages later, however, they use the same terminology—"passive verb"—to describe a verb cast in the passive voice.

A related problem is the lack of uniformity in definitions of the passive voice, which leads to more confusion. Although passive voice sentences are most conventionally defined as those in which the action receiver, rather than the actor, is the subject of the sentence (e.g., Cook, 1985; Williams, 1989), different stylists use different definitions. For example, Svartvik (1966) and Kies (1985a) conceptualized passive voice much more broadly; Svartvik includes any construction with TO BE + part participle[78], including constructions such as *The door remained locked* as *The village seemed deserted.* Kies includes *John is tired* and *Students are often more interested in style* on the same basis; he also includes past participles in isolation as passive constructions (e.g., *Release the merchandise marked for identification*, 1985b).

Given the potential confusion about the term *passive*, one might expect that textbook authors would explicitly discuss the inherent ambiguity of the term and

---

[78] Svartvik includes other stative verbs as acceptable *to be* substitutes, e.g., *appears, seems, looks.*

caution writers not to confuse stative with passive verbs. But few authors include such discussions.

Of course, just because a verb is not stative does not mean it is dynamic—e.g., *generate*, *analyze*, or *distribute* are technically action verbs, but such verbs typically receive low marks for dynamism in either the active or passive voice.[79] Stative verbs that are only possible as active voice constructions (e.g., *is* or *appears*) also lack dynamism. Thus, the argument that writers should switch to the active voice to make the verbs more lively rests upon weak foundations.

(2) **Passive verbs are widely associated with other (undesirable) attributes that collectively contribute to prose characterized as "bureaucratese."** These characteristics include a high incidence of obscure jargon, vague or abstract verbs, nominalized verbs, prepositional phrase stacks, and a generally pretentious and impersonal tone. This idea has intuitive appeal, but the role played by passives in bureaucratic prose has not been well-documented. It is widely assumed that readability studies have included voice as one of the myriad of variables analyzed, but this is not the case; readability studies have focused primarily on variables such as word length, sentence length, number of personal pronouns, number of prepositional phrases, and number of different hard words (Felker, 1980). Thus, until the nature and extent of a relationship of passive voice with undesirable text characteristics is clearly established, we should exercise caution in assuming a relationship where none may exist. Even if such a relationship should exist, it would be of limited applicability to scientific writing, which must be relatively terse to be publishable.[80]

(3) **Passive verbs as used in actual documents are assumed to be wordy.** As Rodman (1981) has pointed out, the argument that passives are wordy can apply only to full passives. And as shown in Study 1, full passives are rare in scientific articles. Even the addition of the auxiliary verb (e.g., *is* or *was*) should not increase the

---

[79] There are stylists (e.g., Kies, 1985), that characterize such non-dynamic "action" verbs as stative.

[80] This extreme compression of information in scientific articles may in fact be the largest single factor responsible for their perceived difficulty, especially for novice readers. So far, I know of no studies done to investigate this question.

length of a sentence if the passive is truncated: *We analyzed the samples* becomes *The samples were analyzed.* Thus, the argument that passives are wordy has little merit. Nevertheless, so many people now perceive the passive as wordy that it is difficult to change this impression. In Warren's (1981) annotated bibliography of texts that discuss passives usage, 19 of the 64 (30%) of the annotations specifically characterize the passive as wordy. When I surveyed more recently published style guides, I saw little evidence that this outlook has changed.

(4) **Poorly-crafted passive constructions are extremely noticeable to readers.** Like any linguistic construction, passives can be misused. And when they are, the result is often awkward (e.g., *It will be determined that*). Such constructions are more noticeable than other problematic writing constructions (e.g., vagueness or imprecision) and may thus tend to elicit far more attention, especially from people who are sensitive to linguistic nuances, such as writers, editors, and writing teachers—all of whom are in a position to influence current usage standards.

(5) **Passive voice is an easy target.** Unlike many characteristics of prose, voice is a concrete, easily identifiable linguistic form. It thus provides an easy target for writing instructors and style manual editors who are seeking ways to provide writing guidelines to fledgling writers. An instructor can define the passive voice, show students bad examples, and tell them to avoid it when possible. Ironically, the concept of passives is easily taught, but not easily learned—students routinely confuse passives with stative verbs. Thus, instructing people to avoid passives is often an exercise in futility.

(6) **The effects of voice are generally assumed to have been more thoroughly researched than they actually have been.** Many authors of style guides are aware that some research (especially early research) has shown an advantage for active over passive voice. Thus, they assume that the proposition that "passive is bad" has been empirically verified. But most are unaware of the details of this research, especially the fact that many of the best-known studies on voice were designed to evaluate the psychological validity of Chomsky's theory transformational grammar, not to assess the effects of voice on reading comprehension or reading speed. The fact that the investigators in such studies often used passive sentences of a type seldom found in

real documents did not matter, because their goal was not to determine writing guidelines for the use of passives. But the sheer volume of such studies has created the impression that passive voice has been extensively researched as a discourse variable when, in fact, it has not.

(7) **Passive constructions are not currently fashionable.** Language, like ideas, goes in and out of fashion. Although the reasons for shifts in linguistic fashion are complex, I can offer one possible explanation for the demise of the passive as a popular form of written discourse. Until the last few decades, when a positivist philosophy of science was influential, passives were favored as exemplars of the "neutral observation language" touted as desirable for scientific research. However, as the influence of logical positivism waned, the idea that a neutral observation language was desirable, or even possible, was challenged by critics such as N. R. Hanson and Thomas Kuhn (Suppe, 1977). Constructions which formerly flourished—among them, the passive voice—came to be associated with a discarded scientific paradigm. Such old-fashioned constructions were discouraged by the current arbiters of acceptable scientific style.

These seven factors have worked together to produce the widespread impression that the passive voice is a dysfunctional language tool whose use must be generally proscribed. But the results of the current investigation suggest that many of our assumptions about the limitations of passive constructions in scientific writing may be based on information that is incomplete, erroneous, or overgeneralized. The results of this investigation point out the need for additional inquiries into the way that linguistic constructions such as the passive are used and understood not only in isolated sentences and passages, but in actual documents.

Study Applications

I became interested in the topic of passive voice in scientific writing as the result of reading a number of well-written scientific reports containing appropriately-used passive constructions whose authors had been told by editors and other higher-ups to remove the passive verbs "because they are harder to understand." I did not find most

of these passives hard to understand, but what I did find hard to understand was why conscientious writers so consistently received advice which seemed at odds with reality. This investigation was my effort to determine whether the scientific passive voice is really more difficult to read.

It is important to note that the articles that first sparked my interest in passives were well-written. As noted above, the passives were used "appropriately"—i.e., they were generally used to focus on what was done (rather than the doer) and to maintain a consistent focus throughout a passage.

The passive voice is one of a number of language tools that writers can either use or misuse, depending upon their skill as wordsmiths. A study such as the present one should not be construed to endorse the use of awkward, wordy, and otherwise inappropriate passive constructions. The articles used for Studies 2a and 2b (the controlled experiments) were as coherent as I could make them. Thus, strictly speaking, the findings that "voice makes no difference" should be applied only to articles that are otherwise well-written. Just as it would be a mistake to say that passives should be avoided in scientific writing, it would be poor advice to say that passives should be used in scientific writing. Such advice misses the point: it is not the specific technique that is good or poor, but the *context* in which it is used. As English professor Richard Lanham notes, "Each prose stye is itself not only an object seen but a way of seeing . . . . Obviously, there can be no single verbal pattern that can be called 'clear'. All depends on context" (p. 33).

Having said that, I think that the passive is often a useful tool in scientific writing, and that its use should not be discouraged, especially in the Methods and Results sections of scientific reports. A better approach would be to provide (a) a clear definition of voice (one that clearly distinguishes "passive" as a characteristic of voice from "passive" as an adjective applied to weak or lifeless verbs) and (b) good and bad examples of passive constructions in a variety of scientific passages.

Directions for Future Inquiry

A sizable number of studies have manipulated voice as a variable. But only a handful have focused on the role of passives as a means of communicating information,

especially written information. As a result, there is still much we do not know about the role of voice as a rhetorical tool.

The present study examined (a) the incidence of passive verbs in published research reports in different fields, (b) the effects of passives on the ease of reading and understanding scientific articles, and (c) passive usage in weak versus strong student papers. The main focus was on the role of voice in connected scientific prose—in particular, the comprehensibility of scientific prose in which passive verbs play a prominent role.

However, there are still unanswered questions regarding passives that have yet to be addressed. Although the present study and previous studies (e.g., Charrow & Charrow, 1980) have focused briefly on the role of sentence voice in different sentence types (i.e., reversible sentences and truncated versus full passives), this area has not been yet been well-researched.

Also, as noted earlier, the present study used adaptations of published articles rather than the articles themselves to assess article comprehension. Although the adaptations look and read much like their published counterparts, it would still be desirable to assess the comprehension of the published articles.

It would also be desirable to assess article comprehension using subjects that are the intended scientific audience for these articles—scientists specializing in the field. In light of the results I obtained using an audience of non-specialists (i.e., college freshmen), it seems likely that passives would pose no problem. But it would be interesting to see whether active constructions in unexpected places, such as the Method section, would adversely affect comprehension, reading speed, or preference ratings for readers accustomed to scientific style .

Assessing how the passive voice is used by writers of different ability or experience would also be desirable. As noted earlier in Study 3, it is possible that writers of well-written papers may make more effective use of passives than writers of less well-written papers. It is thus possible that voice and writing ability interact, such that proficient writers create stronger papers using rhetorical tools such as the passive voice while non-proficient writers create weaker papers using the identical tools. The problems in developing an objective method for judging the quality of writing,

however, might make it difficult to devise a method of evaluation which would yield an acceptable level of inter-rater reliability.

Examining other ways that voice interacts with a variety of variables would be fruitful, but challenging. I performed a pilot study using the articles from Study 2a in which I attempted to vary two factors, voice and tone (personal vs. impersonal). But I did not pursue the study further, in part because tone impersonality seemed to have little effect on comprehension, but also because I had to change the wording too much in order to vary the tone, which unfortunately introduced a new variable (content) that co-varied closely with tone. I was not able to find a practical way to separate the effects of tone and content.[81]

Finally, it would be desirable to better understand the social role that passives play in scientific discourse. Most psycholinguistic research focuses on the effects of text variables such as voice on ease of cognitive processing; writing instructors use the results of such research to advise writers to use writing techniques that facilitate processing and avoid those that impede it. However, ease of processing is not the only yardstick we can use for evaluating linguistic constructions. We can also examine the contextual role that linguistic constructions play in differentiating discourse genres, in reinforcing the social roles that practitioners assume when they write, and in accomplishing other goals (e.g., persuasion, entertainment, or obfuscation).

Scientists try to preserve objectivity when conducting research and to convey that objectivity when reporting their results. Passive verbs help them to do this by enabling them to place the action receiver in the subject slot. Passives are also used to define the genre of scientific writing, to differentiate it from other writing genres. Passives can also be used to block communication—to obfuscate. Passive critics tend

---

[81] It is possible to use multiple regression to analyze text variables, but this approach yields only correlational data. However, using this approach might still provide insights, because the studies that use this approach—readability studies—have not included voice as a text variable. I have no explanation for this omission, given the fact that many of these studies (e.g., Gray & Leary, 1935) examined literally hundreds of other text variables and that readability theorists (e.g., Flesch & Lass, 1947; Gunning, 1968; Klare, 1976) tend to be among the most vocal critics of the passive voice.

to assume that such uses are misguided, i.e., that the problem lies with the construction itself, rather than with the intent of the writer who uses it. But before we give people advice to avoid obfuscatory passives, we should find out whether maximizing clarity is always the writer's goal. If it is not, then our advice may be misguided.

In a related vein, we need to examine not only what makes writing comprehensible, but what makes it successful (i.e., publishable). Not only are these two factors different, but they often tend to conflict with one another, as Lipson (1984)[82] and Hake and Williams (1981) have noted. Understanding such contextual demands may facilitate the development of writing guidelines based on realistic assumptions about real world writing situations.

---

[82] She discusses the dilemma of technical writing instructors who must decide whether to teach students how to write clearly or how to write in a way that is acceptable in the field.

References

Anderson, J. (1974). Verbatim and propositional representation of sentences in immediate and long-term memory. *Journal of Verbal Learning and Verbal Behavior, 13*, 149-162.

Booher, D., & Hill, T. (1989). *Writing for technical professionals*. Wiley: New York.

Britton, G., Glynn, S., Meyer, B., & Penland, M. (1982). Effects of text structure on use of cognitive capacity during reading. *Journal of Educational Psychology, 74*, 51-61.

Cartwright, D., & Zander, A. (1953). *Group dynamics*. Row Peterson: Evanston, IL.

Charrow, R., & Charrow, V. (1979). Making legal language understandable: a psycholinguistic study of jury instructions. *Columbia Law Review, 79*, 1306-1374.

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

Clark, H. (1965). Some structural properties of simple active and passive sentences. *Journal of Verbal Learning and Verbal Behavior, 4*, 365-370.

Clark, H., & Clark, E. (1968). Semantic distinctions and memory for complex sentences. *Quarterly Journal of Experimental Psychology, 20*, 129-138.

Coetzee, J. (1980). The rhetoric of the passive in English. *Linguistics, 18*, 199-221.

Coleman, E.B., & Blumenfeld, J.P. (1963). Cloze scores of nominalizations and their grammatical transformations using active verbs. *Psychological Reports, 14*, 651-654.

Coleman, E.B. (1964). The comprehensibility of several grammatical transformations. *Journal of Applied Psychology, 48*, 186-190.

Coleman, E.B. (1965). Learning of prose written in four grammatical transformations. *Journal of Applied Psychology, 49*, 332-341.

Conway, W. (1981). The passive voice: Friend or foe of the technical communicator? *Technical Writing Teacher, 8*, 86-90.

Cook, C.K. (1985). *The MLA's line by line: How to edit your own writing*. Houghton Mifflin: Boston.

Crystal, D. (1992). *The Cambridge Encyclopedia of Language*. Cambridge, England: Cambridge University Press.

Dagut, M.B. (1985). A 'teaching grammar' of the passive voice in English. *IRAL, 23,* 1-10.

Danks, J., & Sorce, P. (1973). Imagery and deep structure in the prompted recall of passive sentences. *Journal of Verbal Learning and Verbal Behavior, 12,* 114-117.

Day, R. (1994). *How to write and publish a scientific paper.* ISI: Philadelphia.

Faist, R. (1984). *Defining reading comprehension.* Unpublished doctoral dissertation, University of Washington, Seattle, WA.

Farr, R., & Carey, R. (1986). *Reading: What can be measured?* International Reading Association: Newark, Delaware. (ERIC Document Reproduction Service No.266 438).

Felker, D. (1980). *Document design: A review of the relevant research.* Washington, D.C.: American Institutes for Research.

Flesch, R. & Lass, A.H. (1949). *The way to write.* Harper: New York.

Freidin, R. (1975). The analysis of passives. *Language, 51,* 384-405.

Glucksberg, S., Trabasso, T., & Wald, J. (1973). Linguistic structures and mental operations. *Cognitive Psychology, 5,* 338-370.

Gough, P. (1965). Grammatical transformations and speed of understanding. *Journal of Verbal Learning and Verbal Behavior, 4,* 107-111.

Gough, P. (1966). The verification of sentences: the effects of delay of evidence and sentence length. *Journal of Verbal Learning and Verbal Behavior,* 5, 492-496.

Gray, W., & Leary, B. (1935). *What makes a book readable.* Chicago: University of Chicago.

Grieve, R., & Wales, R. (1973). Passives and topicalization. *British Journal of Psychology, 64,*173-182.

Gross, A. (1984). Style and arrangement in scientific prose: the rules behind the rules. *Journal of Technical Writing and Communication, 14,* 241-253.

Gunning, R. (1968). *The technique of clear writing.* New York: McGraw-Hill.

Gross, A. (1985). The form of the experimental paper: A realization of the myth of induction. *Journal of Technical Writing and Communication, 15,* 15-26.

Hake, R., & Williams, J.M. (1981). Style and its consequences: do as I say, not as I do. *College English, 43,* 433-451.

Harris, R.A. Linguistics, technical writing, and generalized phrase structure grammar. *Journal of Technical Writing and Communication, 18*,227-240.

Harris, M. (1978). Noun animacy and the passive voice: a developmental approach. *Quarterly Journal of Experimental Psychology, 30*, 495-504.

Haviland, S., & Clark, H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning & Verbal Behavior, 13*, 512-521.

Hebb, D., & Bindra, D. (1952). Scientific writing and the general problem of communication. *American Psychologist, 7*, 569-573.

Herriot, P. (1969). The comprehension of active and passive sentences as a function of pragmatic expectations. *Journal of Verbal Learning and Verbal Behavior, 8*, 166-169.

Hornby, P. (1972). The psychological subject and predicate. *Cognitive Psychology, 3*, 632-642.

Hupet, M, & Le Bouedec, B. (1975). Definiteness and voice in the interpretation of active and passive sentences. *Quarterly Journal of Experimental Psychology, 27*, 323-330.

Huth, E. (1990). *How to write and publish papers in the medical sciences.* Williams & Wilkins: Baltimore.

James, C., Thompson, J., & Baldwin, J. (1973). The reconstructive process in sentence memory. *Journal of Verbal Learning and Verbal Behavior, 12*, 51-63.

James, C. (1972) Theme and imagery in the recall of active and passive sentences. *Journal of Verbal Learning and Verbal Behavior, 11*, 205-211.

Jarvella, R., & Sinnott, J. (1972). Contextual constraints on noun distributions to some English verbs by children and Adults. *Journal of Verbal Learning and Verbal Behavior, 11*, 47-53.

Jesperson, O. (1974). *A modern English grammar.* Oxford: Alden.

Johnson, M. (1967). Syntactic position and rated meaning. *Journal of Verbal Learning and Verbal Behavior, 6*, 240-246.

Johnson-Laird, P.N., & Stevenson, R. (1970). Memory for syntax. *Nature, 227*, 412.

Johnson-Laird, P.N. (1968b). The choice of the passive voice in a communicative task. *British Journal of Psychology, 59*, 7-15.

Johnson-Laird, P.N. (1968a). The interpretation of the passive voice. *Quarterly Journal of Experimental Psychology, 20* , 69-73.

Just, M., & Carpenter, P. (1980). A theory of reading : From eye fixations to comprehension. *Psychological Review, 87*, 329-354.

Kerr, N., Butler, S., Maykuth, P., & Delis, D. (1982). The effects of thematic context and presentation mode on memory for sentence voice. *Journal of Psycholinguistic research, 11*, 247-264.

Kies, D. (1985b). Some stylistic features of business and technical writing: The funcations of passive voice, nominalization, and agency. *Journal of Technical Writing and Communication, 15*, 299-308.

Kies, D. (1985a). *A functional linguistic view of language development: Common links between writing centers, composition courses, and the academy.* Paper presented at the Midwest Writing Centers Association Conference, St. Louis, MO. (ERIC Document Reproduction Service No. 266 462).

Klare, G. (1976). A second look at the validity of readability formulas. *Journal of Reading Behavior, 8*, 129-152.

Klare, G. (1977). Readable technical writing: Some observations. *Technical Communication, 24*, 1-3.

Klare, G., & Buck, B. (1954). *Know your reader: The scientific approach to readability.* Hermitage House: New York.

Kolln, M. (1991). *Rhetorical grammar: Grammatical choice, rhetorical effects.* New York: Macmillan.

Layton, P., & Simpson, A. (1975). Surface and deep structure in sentence comprehension. *Journal of Verbal Learning and Verbal Behavior, 14*, 658-664.

Lipson, C. (1984). *Style in technical writing: Foundations for pedagogy.* Unpublished paper. Syracuse, NY: Syracuse University.

Lynch, B.S., & Chapman, C. (1980). *Writing for communication in science and medicine.* New York: Van Nostrand.

Martin, E., & Roberts, K. (1966). Grammatical factors in sentence retention. *Journal of Verbal Learning and Verbal Behavior, 5*, 221-218.

Martin, J.R. (1993). Literacy in science: learning to handle text as technology. In M.A.K. Halliday, & J.R. Martin (Eds.). *Writing Science: Literacy and Discursive Power* (pp. 166-202). University of Pittsburg: Pittsburg.

Martin, E., Roberts, K., & Collins, A. (1968). Short-term memory for sentences. *Journal of Verbal Learning and Verbal Behavior, 7,* 560-566.

Matthews, J., Bowen, J., & Matthews, R. (1996). *Successful scientific writing.* Cambridge University Press: New York.

Mehler, J. (1963). Some effects of grammatical transformations on the recall of English sentences. *Journal of Verbal Learning and Verbal Behavior, 2,* 346-351.

Miller, G., & McKean, K. (1964). A chronometric study of some relations between sentences. *Quarterly Journal of Experimental Psychology, 16,* 297-308.

Miller, G. (1962). Some psychological studies of grammar. *American Psychologist, 17,* 748-762.

Moore, T., & Biederman, I. (1979). Speeded recognition of ungrammaticality: double violations. *Cognition, 7,* 285-299.

Murray, G. (1933). *The Oxford English Dictionary.* Oxford: Clarenden Press.

Olson, D., & Filby, N. (1972). On the comprehension of active and passive sentences. *Cognitive Psychology, 3,* 361-381.

Parot, F., & Kail, M. (1980). Topicalisation et thématisation dans le rappel de phrases. *L'Annee Psychologique, 80,* 99-120.

Paul, T., & Rosner, M. (1983). Discovering the teaching syntactic structures three technical disciplines. *Journal of Technical Writing and Communication,* 13, 109-122.

Perfetti, C., & Goldman, S. (1974). Thematization and sentence retrieval. *Journal of Verbal Learning and Verbal Behavior, 13,* 70-79.

Perfetti, C., & Goldman, S. (1975). Discourse funcations of thematization and topicalization. *Journal of Psycholinguistic Research, 4,* 257-279.

Perfetti, C. (1969). Sentence retention and the depth hypothesis. *Journal of Verbal Learning and Verbal Behavior, 8,* 101-104.

Prentice, J. (1966). Response strength of single words as an influence in sentence behavior. *Journal of Verbal Learning and Verbal Behavior, 5,* 429-433.

Publication Manual of the American Psychological Association, 4th ed. American Psychological Association. Washington, D.C., 1994.

Riley, K. (1991). Passive voice and rhetorical role in scientific writing. *Journal of Technical Writing and Communication, 21,* 239-257.

Rodman, L. (1991). Anticipatory *IT* in scientific discourse. *Journal of Technical Writing and Communication, 21,* 17-27.

Rodman, L. (1981). *The passive in technical and scientific writing.* Paper presented at the 32nd Conference on College Composition and Communication, Dallas, TX.

Rowan, K. (1989). Moving beyond the *what* to the *why*: Differences in professional and popular science writing. *Journal of Technical Writing and Communication, 19,* 161-179.

Rubens, P.M. (1982). Rule making in technical communication: A prolegoma to a study. *Teaching Writing Teacher, 9,* 75-81.

Sachs, J. (1974). Memory in reading and listening to discourse. *Memory and Cognition, 2,* 95-100.

Sachs, J. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics, 2,* 437-442.

Schwager, E. (1991). *Medical English Usage and Abusage.* Oryx: Phoenix.

Scriven, K. (1989). Actively teaching the passive voice. *Teaching English in the Two-year College, 16,* 889-93.

Siegel, A., & Burkett, J. (1974). *Application of structure-of-intellect and psycholinguistic concepts to reading comprehensibility measurement.* (Report No. APHRL-TR-74-49). Lowry AFB, CO: Air Force Human Resources Lab. (ERIC Document Repreoduction Service No. ED 099 817).

Slobin, D. (1966). Grammatical transformations and sentence comprehension in childhood and adulthood. *Journal of Verbal Learning and Verbal Behavior, 5,* 219-227.

Slobin, D. (1968). Recall of full and trancated passive sentences in connected discourse. *Journal of Verbal Learning and Verbal Behavior, 7,* 876-881.

Spyridakis, J. & Wenger, M. (1992). Writing for human performance: Relating reading research to document design. *Technical Communication, 2,* 202-215.

Stanley, J. (1975). Passive motivation. *Foundations of Language, 13,* 25-39.

Stratton, C. (1984). *Technical Writing: Process and Product.* Harcourt Brace: New York.

Strunk, W., & White, E. (1979). *The Elements of Style.* Macmillan: New York.

Suppe, F. (1977). *The structure of scientific theories.* Urbana: University of Illinois, 1977.

Svartvik, J. (1966). *On voice in the English verb.* The Hague: Mouton.

Tannenbaum, P., & Williams, F. (1968). Generation of active and passive sentences as a function of subject or object fonts. *Journal of Verbal Learning and Verbal Behavior, 7,* 246-250.

Tannenbaum, P., Evans, R., & Williams, F. (1964). An experiment in the generation of simple sentence structures. *Journal of Communcation, 14,* 113-117.

Trabasso, T. (1972). Mental operations in language comprehension. In J.B. Carroll & Roy O. Freedle (Eds). *Language Comprehension and the Acquisition of Knowledge* (pp. 113-137). Washington, D.C.: Winston.

Turner, E., & Rommetveit, R. (1968). Focus of attention in recall of active and passive sentences. *Journal of Verbal Learning and Verbal Behavior, 7,* 543-548.

Vande Kopple, W. (1989). *Clear and coherent prose: A functional approach.* Scott Foresman: Glenview, IL.

Walpole, J. (1979). Why must the passive be damned? *College Composition and Communication, 30,* 251-254.

Wannemacher, J. (1976). Processing strategies in sentence comprehension. *Memory & Cognition, 4,* 48-52.

Wannemacher, J. (1974). Processing strategies in picture-sentence verification tasks. *Memory & Cognition, 2,* 554-560.

Warren, T. (1981). The passive voice verb: A selected, annotated bibliography—parts II-V. *Journal of Technical Writing and Communication, 11,* 373-389.

Wearing, A.J. (1970). The storage of complex sentences. *Journal of Verbal Learning and Verbal Behavior, 9,* 21-29.

Wearing, A. (1972). Remembering complex sentences. *Quarterly Journal of Experimental Psychology, 24,* 77-86.

Wilkinson, A.M. (1992). Jargon and the passive voice: prescriptions and proscriptions for scientific writing. *Journal of Technical Writing and Communication, 22,* 319-325.

Williams, J.M. (1989). *Style: Ten Lessons in Clarity and Grace.* Scott Foresman: New York.

Wilson, S. (1979). Factors influences the frequency of switches of voice during the recall of active and passive sentences. *Journal of General Psychology, 100,* 277-285.

Wright, P. (1969). Transformations and the understanding of sentences. *Language and Speech, 12,* 156-166.

Wright, P., Glucksberg, S. (1976). Choice of definite versus indefinite article as a function of sentence voice and reversibility. *Quarterly Journal of Experimental Psychology, 28,* 561-570.

Yngve, V.H. (1960). A model and an hypothesis for language structure. *Proceedings of the American Philosophical Society, 104,* 444-466.

Zeiger, M. (1991). *Essentials of writing biomedical research papers.* McGraw Hill: New York.

Appendix A:

Citations for Articles Analyzed in Study 1

Psychology:

Bulletin of the Psychonomic Society:

Beatty, W.W. (1993). Age differences on the California Card Sorting Test: Implications for the assessment of problem solving by the elderly. *Bulletin of the Psychonomic Society. 31*, 511-514.

Davis, S. F., Armstrong, L. W., & Huss, M. T. (1993). Shock-elicited aggression is influenced by lead and/or alcohol exposure. *Bulletin of the Psychonomic Society, 31*, 451-453.

Dewsbury, D. A. (1993). Sperm competition and effects of mating order on copulatory behavior in meadow voles (*Microtus pennsylvanicus*). *Bulletin of the Psychonomic Society. 31*, 437-439

Higginbotham, P., & Bartling, C. (1993). The effects of sensory distreactions on short-term recall of children with attention deficit-hyperactivity disorder versus normally achieving children. *Bulletin of the Psychonomic Society, 31*, 507-510.

Journal of Clinical Psychology:

Wong, J. L., & Whitaker, D. J. (1993). Depressive mood states and their cognitive and personality correlates in college students: they improve over time. *Journal of Clinical Psychology, 49*, 615-621.

Beck, A. T., Brown, G. K., & Beck, J. S. (1993). *Journal of Clinical Psychology, 49*, 603-614.

Janzen, B. L., Saklofske, D. H., & Kelly, I. W. (1993). Personality and bulimic symptomatology. *Journal of Clinical Psychology, 49*, 649-653.

Wierzbicki, M. (1993). Use of MCMI subtle and obvious subscales to detect faking. *Journal of Clinical Psychology, 49*, 809-814.

Journal of General Psychology:

Mandal, M. K., Pandey, R., & Madan, S. K. (1992). Exposed eye area (EEA) in the expression of various emotions. *Journal of General Psychology, 119*, 385-389.

Stone, N. J. (1993). Performance, mood, satisfaction, and task type in various work environments: a preliminary study. *Journal of General Psychology, 120, 489-497.*

Weiler, E. M., Smith Gold, L., Sandman, D. E., & Warm, J. S. (1992). Four triggering factors in loudness adaptation. *Journal of General Psychology, 119*, 325-334.

Zimmerman, D. W. (1993). Mimicking properties on non-parametric rank tests using scores that are not ranks. *Journal of General Psychology, 120*, 509-516.

Botany:

American Journal of Botany

Andersen, M. C. (1993). Diaspore Morphology and seed dispersal in several wind-dispersed Asteraceae. *American Journal of Botany, 80*, 487-492.

Gali, Hala U. (1992). Effect of silicon supply on growth, fertility, and mineral composition of an annual brome, bromus secalinus l. (Graminaeae). *American Journal of Botany, 79*, 1259-1263.

Lammers, T. G., & Hensold, N. (1992). Chromosome numbers of campanulaceae. II. the lobelia tuba complex of Chile. *American Journal of Botany, 79*, 585-587.

Wakamiya, I., Newton, R. J., Johnston, J. S., & Price, H. J. (1993). Genome size and Environmental factors in the genus *Pinus*. *American Journal of Botany, 80*, 1235-1241.

Phytopathology

Cappaert, M. R., Powelson, M. L., Christensen, N. W., Crowe, F. J. (1992). Influence of irrigation on severity of potato early dying and tuber yield. *Phytopathology, 82*, 1448-1453.

Kolmer, J. A. (1993). Selection in a heterogeneous population of *Puccinia recondita* f. sp. *tritici*. *Phytopathology, 83*, 909-914.

Kousik, C. S., Snow, J. P., Valverde, R. A. (1994). Comparison of double-stranded RNA components and virulence among isolates of *Rhizoctonia solani* AG-1IA and Ag-1 IB. *Phytopathology, 84*, 44-48.

Vega, F. E., Davis, R. E., Barbosa, P., Dally, E. L., Purcell, A. H., & Lee, Ing-Ming. (1993). Detection of a plant pathogen in a nonvector insect species by the polymerase chain reaction. *Phytopathology, 83*, 621-624.

Zoomorphology

Cifrian, B., Martinez-Alos, S., & Gremigni, V. (1993). Ultrastructural and cytochemical studies on the germarium of *Dicrocoelium dendriticum* (Plathelminthes, Digenea). *Zoomorphology, 1993*, 165-171.

Fish, F. E., Stein, B. R. (1991). Functional correlates of differences in bone density. *Zoomorphology, 1991*, 339-345.

Hou, S. T., & Maxwell, W. L. (1992). Evidence for an intermediate type of spermatozoon: ultrastructural studies of spermiogenesis in the cuttlefish *Rossia macrosoma* (Cephalodpoda, Decabrachia). *Zoomorphology, 112*, 209-7-215.

Huvard, A. L. (1990). Ultrastructural study of the naupliar eye of the ostracode *Vargula graminicola* (Crustacea, Ostracoda). *Zoomorphology, 110*, 47-51.

Chemistry/Physics:

Analytical Chemistry

Carroll, J. A., Ngoka, L., Beggs, C. G., Lebrilla, C. B. (1993). Liquid secondary ion mass spectrometry/Fourier transform mass spectrometry of oligosccharide anions. *Analytical Chemistry, 65*, 1582-1587.

Lemmo, A. V., Jorgenson, J. W. (1993). Transverse flow gating interface for the coupling of microcolumn LC with CZE in a comprehensive two-dimensional system. *Analytical Chemistry, 65*, 1576-1581.

Redepenning, J., Schlesinger, T. K., Mechalke, E. J., Puleo, D. A., & Bizios, R. (1993). Osteoblast attachment monitored with a quartz crystal microbalance. *Analytical Chemistry, 65*, 3378-3381.

Whelan, J.P., Kusterbeck, A.W., Wemhoff, G.A., Bredehorst, R., & Ligler, F.S. Continuous-flow immunosensor for detection of explosives. (1993). *Analytical Chemistry, 120*, 489-497.

Chemical Physics Letters

Davies, P. B., Liu, Y., & Liu, Z. (1993). Far intrared LMR spectra of monobromomethyl radicals. *Chemical Physics Letters, 214*, 305-309.

Mares, J. A., Boulon, G., Brenier, A., Lou, L, & Rotman, S. R. (1994). Spectroscopy of multisites Cr^{3} and Nd^{3+} in yttrium aluminum perovskite. *Chemical Physics Letters, 217*, 105-111.

Ortiz, J. V. (1993). Electron propagator calculations on the electron affinity of C_5. *Chemical Physics Letters, 216*, 319-323.

Smirnov, S. N., & Braun, C. L. (1994). Light polarization effects in the transient dc conductivity response to excitation of dipolar solutes. *Chemical Physics Letters, 217*, 167-172.

Journal of Chromatography

Colli, M., Zabarini, L., Melze d'Eril, G. V., & Marchetti, R. (1993). Evaluation of a modified Marcali technique with high-performance liquid chromotography-ultraviolet detection for the determination of 2, 4-toluene diisocyanate in air. *Journal of Chromotography, 643*, 51-54.

Dankers, J., Groenenboom, M., Scholtis, L.H.A., & van der Heiden, C. (1993). High-spped supercritical fluid extraction method for routine measurement of polycyclic aromatic hydrocarbons in environmental soils with dichloromethane as a static modifier. *Journal of Chromotography, 641*, 352-362.

Meiring, H. D., den Engelsman, G., & de Jong, A. P. J. M. (1993). Determination of polar pesticides by phase-transfer catalysed derivatization and negative-ion chemical ionization gas chromotography-mass spectrometry. *Journal of Chromotography, 644*, 357-364.

Silvestro, L, Da Col, R., Scappaticci, E., Libertucci, D., Biancone, L, & Camussi, G. (1993). Development of a high-performance liquid chromotographic-mass spectrometric technique, with an ionspray interface, for the determination of platelet-activating factor (PAF) and lyso-PAF in biological samples. *Journal of Chromography, 647*, 261-269.

Appendix B:
Citations for Study 2 Article Adaptations

Rhodes, S. (1988). *Counterfactual reasoning in civil liability cases.* Paper presented at the University of Washington Psychology Research Festival, June 1988.24 pp.

Parrott, W.G., & Sabini, J. (1990). Mood and memory under natural conditions: Evidence for mood incongruent recall. *Journal of Personality and Social Psychology, 59*, 231-336.

Forthman, D. L., Elder, S. D., Bakeman, R., Kurkowski, T. W., Noble, C. C., & Winslow, S. W. (1992). Effects of feeding enrichment on behavior of three species of captive bears. *Zoo Biology, 11*, 187-195.

Saetre, G. P., Fossnes, T., & Sladsvold, T. Food provisioning in the pied flycatcher: do females gain direct benefits from choosing bright-colored males? *Journal of Animal Ecology, 64*, 21-30.

Appendix C:
Active and Passive Page Excerpts
From One Study 2 Article

On the following two pages is the first page of the *Feeding Bears* articles (Study 2b). Voice manipulations are bolded.

Article B — Feeding Bears in Zoos (active voice version)

From the public's perspective, bears are among the most popular zoo animals. They are active, large, impressive, and the subject of considerable folklore. In the recent trend toward naturalizing zoo exhibits, however, zoo managers **have generally left the family *Ursidae* behind.** Most captive bears inhabit outdoor moated grottos constructed of gunite, with a pool and occasionally logs. Such environments **do not adequately display** the normal behavioral repertoire and natural history of the family (Melisso & Savarese, 1990). Zoo bears exhibit unnatural, stereotypic behaviors that we **seldom see** in the wild, e.g., pacing, rocking, and swaying—patterns which in humans reflect a maladaptive pattern of adjustment to environment stresses (Forthman *et al.*, 1992). Researchers studying captive animals have often linked such behaviors to conflict and frustration (Forthman *et al.*, 1992; Wechsler, 1991). In addition to exhibiting stereotypic behavior, captive bears are also prone to engage in begging or masturbation (Bayermeister, 1995).

Thus, the manner in which **we exhibit** most captive bears fails to promote behavior characteristic of bears in the wild. In particular, traditional feeding procedures—simply putting out food once or twice a day—do not promote the kind of hunting and foraging behavior seen in wild bears. Two negative consequences ensue: this management approach (1) **compromises** the health and well-being of the bears (Bayermeister & Puff, 1994; Melisso & Savarese, 1990) and (2) hampers efforts to conserve bears in the wild by educating the zoo-going public about bear behavior because the public **too often sees** zoo bears doing things that the bears would never do in the wild (Forthman & Elder, 1992).

However, it is difficult to provide bears with opportunities to perform species-typical feeding behaviors in standard exhibits. Devising representative feeding exhibits for bears **has become** more difficult because of the varied ways in which bears feed in the wild. Wild bears are omnivorous. They spend much of their time climbing, digging for small mammals and tubers, tearing apart rotting logs, and consuming seasonal fruits (Graber & White, 1983; Mace & Jonkel, 1986). These activities, combined with a large body size, make it difficult to display bears in exhibits with natural substrates and vegetation (Tsai & Puff, 1991).

There have been two key studies of feeding enrichment in captive bears. In one recent study (Markowitz, 1982), researchers **made** efforts to reduce the level of combative interactions between a male and female polar bear and to provide them with alternative behavioral opportunities. They **designed** a sound-activated fish catapult to enable the bears to "order" fish by vocalizing into a microphone, thereby giving the bears a more active role in the feeding process. This treatment resulted in a decreased incidence of combative interactions by both bears and a decreased incidence of pacing by the male.

Article B — Feeding Bears in Zoos (passive voice version)

From the public's perspective, bears are among the most popular zoo animals. They are active, large, impressive, and the subject of considerable folklore. In the recent trend toward naturalizing zoo exhibits, however, the family *Ursidae* generally **has been left** behind. Most captive bears inhabit outdoor moated grottos constructed of gunite, with a pool and occasionally logs. In such environments, the normal behavioral repertoire and natural history of the family **is not adequately displayed** (Melisso & Savarese, 1990). Zoo bears exhibit unnatural, stereotypic behaviors that **are seldom seen** in thewild, e.g., pacing, rocking, and swaying—patterns which in humans reflect a maladaptive pattern of adjustment to environment stresses (Forthman *et al.*, 1992). In captive animals, such behaviors have often been linked with conflict and frustration (Forthman *et al.*, 1992; Wechsler, 1991). In addition to exhibiting stereotypic behavior, captive bears are also prone to engage in begging or masturbation (Bayermeister, 1995).

Thus, the manner in which most captive bears **are exhibited** fails to promote behavior characteristic of bears in the wild. In particular, traditional feeding procedures—simply putting out food once or twice a day—do not promote the kind of hunting and foraging behavior seen in wild bears. Two negative consequences ensue: (1) the health and well-being of the bears **is compromised** (Bayermeister & Puff, 1994; Melisso & Savarese, 1990) and (2) efforts to conserve bears in the wild by educating the zoo-going public about bear behavior are hampered because zoo bears **are too often seen** doing things they would never do in the wild (Forthman & Elder, 1992).

However, it is difficult to provide bears with opportunities to perform species-typical feeding behaviors in standard exhibits. Devising representative feeding exhibits for bears **has been made** more difficult by the varied ways in which bears feed in the wild. Wild bears are omnivorous. Much of their time is spent climbing, digging for small mammals and tubers, tearing apart rotting logs, and consuming seasonal fruits (Graber and White, 1983; Mace and Jonkel, 1986). These activities, combined with a large body size, make it difficult to display bears in exhibits with natural substrates and vegetation (Tsai & Puff, 1991).

There have been two key studies of feeding enrichment in captive bears. In one recent study (Markowitz, 1982), efforts **were made** to reduce the level of combative interactions between a male and female polar bear and to provide them with alternative behavioral opportunities. A sound-activated fish catapult **was designed** to enable the bears to "order" fish by vocalizing into a microphone, thereby giving the bears a more active role in the feeding process. This treatment resulted in a decreased incidence of combative interactions by both bears and a decreased incidence of pacing by the male.

Appendix D:
Entire Texts of Study 2 Articles

On the following pages are the entire texts of the articles and comprehension questions used for Study a and b. The articles and one version (either active or passive) of each article and its multiple choice questions appears in the following order:

Study 2a:
Article X: *Counterfactual Thinking* (active voice)
Article Y: *Mood and Memory* (passive voice)

Study 2b:
Article A: *Parenting in Pied Flycatchers* (passive voice version)
Article B: *Feeding Bears in Zoos* (active voice version)

Article X: Counterfactual Thinking (active voice version)

Consider the following situations:

> Ted was waiting in a checkout line to purchase groceries and try his luck on a scratch-and-win ticket given with each purchase. As he was about to check his groceries, Ted let a woman go before him because she had only two items. Her ticket won $5,000.
>
> Jan, a busy mother of three and elementary school teacher, had elective surgery during her summer vacation. To be prudent, she had initially decided to donate her own blood for the operation but couldn't quite fit the appointments into her busy schedule. The operation went well but a six-month checkup revealed that Jan was HIV-positive, the result of having received AIDS-contaminated blood.

Scenarios like these tend to evoke counterfactual thinking—"if only" thinking. Counterfactual thinking refers to the way that we alter unfortunate or abnormal outcomes and substitute preferred alternatives in their place. In the checkout line example, we tend to feel that the winning ticket really belongs to Ted, and thus to imagine that he—rather than the woman—won the $5,000. In the tainted blood scenario, we can easily imagine Jan having donated her own blood before the operation. If only she had, she could have avoided the AIDS-contaminated transfusion.

The most noticeable characteristic of counterfactual thinking is its *irresistibility*. Almost before we can help it, we tend to mentally re-write Jan's story so that Jan *does* donate her own blood, and thus doesn't receive the contaminated blood. We easily generate such alternative scenarios—in fact, they often seem to spring to mind unbidden. These unbidden thoughts about what might have been—if only some event had been different—often arise despite our best efforts to block them out.

Such counterfactual thinking is commonplace. But we don't yet understand how it works psychologically. Some have suggested that we make sense of past events by constructing causal scenarios based on what we know about similar situations (Read, 1987). If information is missing, we use inferences to fill in the missing details so that the story makes sense. For example, in the case of court testimony, jurors must construct plausible stories about what might have actually occurred to make sense of different versions of the "facts" (Pennington and Hastie, 1986) .

Wells and Gavanski (1989) have studied the effects of counterfactual *mental acts* on subjects attributions of causality and responsibility. They assessed how subjects would react to what someone considers doing versus what they actually do in a study in which they presented subjects with the following scenario:

> Mr. Carlson took his assistant editor Karen to an elegant French restaurant to celebrate her promotion. Because he had been there before, he ordered for both of them. He considered ordering the Coquilles Saint-Jacques, but decided at the last moment to order Moules Mariniere. Soon after finishing her meal, Karen collapsed. Efforts to revive her were unsuccessful. It turned out that she was extremely allergic to wine and had died of anaphylactic shock after eating the Moules Mariniere, which contained wine.

The experimenters divided the subjects into two groups. They presented subjects in the "two wine" condition with the above scenario and informed them that Coquilles Saint-Jacques uses wine. They presented subjects in the "one wine" condition with a scenario identical to the one above except that it was not Coquilles Saint Jacques that Carlson almost ordered but Scallops Meuniere, a wine-free dish. Thus, in the "two-wine" condition, Carlson made an initial decision to order a dish containing wine, but at the last minute switched to yet another dish containing wine. In the "one-wine" condition, he initially decided to order a wine-free dish but at the last minute switched to a dish containing wine.

Subjects listed four things that could have been different in the story to prevent Karen's death and also listed the four most important causes of Karen's death. They also rated on a nine point scale how much of a contribution Mr. Carlson's decision to order the Moules Mariniere played in Karen's death. 86% of the subjects in the "one-wine" condition listed the ordering decision as a cause; only 49% of the subjects in the "two-wine" condition listed it as a cause. Subjects in the "one-wine" condition also thought that Mr. Carlson's decision to order the Moules Mariniere played more of a contribution to Karen's death than did subjects in the "two-wine" condition.

Wells and Gavanski concluded that thoughts of what might have been can influence causal reasoning. In this experiment, when they made desirable counterfactual alternatives explicit, subjects tended to think that what *could* have happened, *should* have happened. Consequently, they assigned greater responsibility and causality to someone who considered doing something that would have changed an

unfortunate outcome—but then *failed to follow through* on this consideration—than to the person who considered doing something that would *not* have changed the outcome.

We conducted the present study to assess the effects of counterfactual thinking on causal attribution and damage awards in civil suits. We designed the study to elicit the reaction of mock jurors to the *actual* acts versus the *considered* acts of the defendant in a lawsuit. In the Wells and Gavanski study, subjects responded to counterfactual information presented in a very short scenario. Our experiment presented counterfactual information so that this information would be less easy to spot, as part of a simulated court transcript. If counterfactual information was embedded within a longer, more diverse series of passages, would it influence judgments of liability, judgments of causality and responsibility, and damage awards?

We used a relatively simple fact pattern for the case of Potter versus Shrackle: Driver Charles Shrackle takes his usual route through town. Unfortunately, he hits and kills Catherine Potter, a pedestrian. In the control condition, he merely drives through town and does not consider taking an alternative route. In the experimental condition, he takes his usual route through town, but almost decides to take an alternative scenic route prior to the accident. Her estate sues, claiming that Potter's death resulted from Shrackle's negligence.

Method

Dependent Measures and Hypotheses. We assessed the effect of condition on four variables: defendant liability (liable/not liable), size of damage award to the plaintiff, causal role of the defendant's route decision (0 - 8 scale, where 0 = none and 8 = maximum role), and the defendant's feelings of responsibility (0 - 8 scale, where 0 = none and 8 = maximum). We predicted that defendant liability would be greater for subjects in the experimental condition and that the mean award size would be larger in the experimental condition. We also predicted that causal role of the route decision and defendant feelings of responsibility would be greater for experimental subjects.

Procedure. We gave subjects a written transcript of the key parts of the trial testimony, relevant sections of the Washington Pattern Jury Instructions and a questionnaire to fill out after reading the transcript and jury instructions. We mentioned Shrackle's consideration of an alternative route three places in the transcript: in the opening address of the defense attorney; in the testimony of the decedent's husband, Jeffrey Potter; and in the closing address of the defense attorney.

Subjects. We randomly selected 214 University of Washington students enrolled in introductory psychology courses to participate in this study.

Results

60% of subjects rated the defendant as liable; 40% of subjects rated the defendant as not liable. However, there was no difference in liability ratings by condition. 68% of subjects made damage awards (145/214). For those instances in which subjects made damage awards, subjects in the experimental condition made significantly higher mean damage awards than subjects in the control condition, $t(213) = 1.85$, $p < .05$. We predicted that both the causal role of route decision and defendant feelings of responsibility would be higher for experimental subjects. But we did not confirm either of these predictions: subjects in the experimental condition rated neither causal role nor defendant's feelings of responsibility as higher.

Discussion

In this experiment, when we emphasized the fleeting thoughts of a defendant, mock jurors tended to make higher mean damage awards. This finding is consistent with the results of Wells and Gavanski (1989) in that we demonstrated that fleeting thoughts have a measurable effect on subject judgments.

However, these higher awards did not seem to arise as the result of the perception that the defendant's route decision was more causal of the accident or that the defendant should feel more responsible for the accident. Why then did subjects in the experimental condition award higher damages? One possibility is that these subjects feel greater sympathy for the victim because the consideration of an alternative route made the accident seemed more poignant, more regrettable. But

because we did not ask subjects to indicate their own feelings of either sympathy or regret, such a conclusion is only conjecture at this point. Including this measure in future studies may permit a better understanding of the reasons why awards in this experiment were higher for accidents in which subjects did not consider counterfactual alternatives.

If mock jurors do indeed tend to award higher amounts when they have been presented with easy-to-imagine alternatives that would have prevented the situation from arising, such a finding has major implications for the way that attorneys present legal evidence to the jury. Plaintiffs' attorneys might boost monetary awards by dropping frequent hints that the unfortunate situation could easily have been different had it occurred five minutes earlier, had a traffic light been green instead of red, etc. Conversely, defense attorneys might hold awards to a minimum by rigorously reminding the jurors what the facts actually were, not what they might have been.

Article X - Multiple-choice Questions

1. According to the authors of Article X, which of the following statements about counterfactual thoughts is FALSE?

a. They are hard to block out
b. They tend to arise during periods of high stress
c. They require mental concentration to generate
d. b and c only
e. All of the above

2. You have purchased a state lottery ticket. Your number is 9374889. Which of the following would be the best example of counterfactual thinking as defined in the article?

a. You mentally concentrate for several hours on 9374889, hoping that number will be drawn.
b. You imagine in great detail all the things you will do with the lottery money if you win.
c. The winning number is 9374888; you can't help imagining the winning number being 9374889 instead.
d. The winning number is 005563; you strongly regret the number was not 9374889.
e. None of the above

3. What did Wells and Gavanski (1989) study?

a. the effect of attributions of causality and responsibility on mental concentration
b. the effect of mental acts on attributions of causality and responsibility
c. the effect of mental acts on the future outcomes of physical events
d. the effect of desirable outcomes on attributions of causality and feelings of regret
e. the effect of attributions of causality and responsibility on damage awards

4. Which of the following scenarios would make the BEST substitute for one of the two conditions described in Wells & Gavanski's "one wine, two wine" experiment?

a. At a dinner party, a woman was served Steak Tartare (a raw beef patty), one of her favorite dishes. She ate every bite. But unknown to her the dish contained E Coli bacteria. She became very ill and quickly died of kidney complications.

b. A ticketseller surprised a friend with tickets to a basketball game. He initially considered selecting tickets on the west side of the stadium, but at the last minute, he decided on east side tickets. During the first quarter, a violent earthquake collapsed all the seats on the east and south sides. The friend was crushed to death.

c. A woman took the last seat on a bus, a handicapped seat at the front. When a man on crutches boarded, the driver asked her to vacate her seat and move to the rear. When a large truck rear-ended the bus, the man was uninjured but the woman was killed.

d. A man with an early morning flight forget to set his alarm and awoke late. Although he rushed to the airport, he barely missed his flight. He took the next flight instead. He was killed when his plane crashed shortly after takeoff.

e. Two girls, Selma and Louise, were skating on a frozen pond. Selma thought about skating over to the log but decided not to. Louise also thought about skating over to the log and decided to do it. She fell through the ice and unfortunately drowned.

5. For Wells and Gavanski's "one wine, two wine" experiment, which of the following statements is true?

a. Scallops Meuniere contains wine but Coquilles Saint-Jacques does not contain wine

b. Scallops Meuniere contains wine but Moules Mariniere does not contain wine

c. Moules Mariniere contains wine but Coquilles Saint-Jacques does not contain wine

d. Coquilles Saint-Jacques contains wine but Moules Mariniere does not contain wine

e. Coquilles Saint-Jacques contains wine but Scallops Meuniere does not contain wine

6. Wells and Gavanski found that:

a. subjects in the "one wine" condition were much more likely than subjects in the "two wine" condition to list the ordering decision as a cause of Karen's death

b. subjects in the "two wine" condition were much more likely than subjects in the "one wine" condition to list the ordering decision as a cause of Karen's death

c. subjects in the "two wine" condition were not more likely than subjects in the "one wine" condition to list the ordering decision as a cause of Karen's death

d. over 80% of the subjects in both of the groups listed the ordering decision as a cause of Karen's death

e. a and d only

7. How did the design of the Article X experiment differ from the design of previous experiments on counterfactual thinking?

a. It was shorter and used a legal scenario involving a civil suit.
b. It was shorter and used a legal scenario involving a criminal case.
c. It was longer and used a legal scenario involving a civil suit.
d. It was longer and used a legal scenario involving a criminal case.
e. It was longer and used a legal scenario involving both a civil suit and a criminal case.

8. In the Article X experiment, subjects received the following:

a. a written transcript
b. relevant sections of the Washington Pattern Jury Instructions
c. a written questionnaire
d. a and c only
e. all of the above

9. For the Article X experiment, the transcript mentioned defendant Shrackle's route decision in how many places?

a. one
b. two
c. three
d. four
e. five

10. How many subjects participated in the Article X experiment?

a. 48
b. 104
c. 138
d. 148
e. 214

11. Which of the following best describes the two conditions in the Article X experiment?

a. Experimental: Defendant (driver) considers taking an alternative route
Control: Defendant does not consider taking an alternative route
b. Experimental: Plaintiff (pedestrian) considers taking an alternative route
Control: Plaintiff does not consider taking an alternative route
c. Experimental: Both defendant and plaintiff consider taking an alternative route
Control: Only defendant considers taking an alternative route
d. Experimental: Both defendant and plaintiff consider taking an alternative route
Control: Only plaintiff consider taking an alternative route
e. Experimental: Both defendant and plaintiff consider taking an alternative route
Control: Neither considers taking an alternative route

12. Which of the following was among the variables that the Article X experiment assessed? (note: The defendant is the driver, the plaintiff is the pedestrian)
a. plaintiff liability
b. size of the damage awards made to the plaintiff
c. sympathy for the plaintiff
d. a and b only
e. b and c only

13. Subjects in the EXPERIMENTAL group of the Article X experiment:
a. rated the defendant as liable more often than subjects in the control group
b. made higher damage awards to plaintiff than subjects in the control group
c. rated the causal role of route decision as greater that subjects in the control group
d. a and b only
e. b and c only

14. According to the authors of Article X, the results of the Article X experiment:
a. were consistent with the findings of Wells & Gavanski
b. did not seem to arise from viewing the defendant's route decision as a causal factor in the accident
c. were consistent with the idea that subjects in the experimental group perceived the accident as more poignant than subjects in the control condition
d. all of the above
e. none of the above

15. Which of the following implications for presenting of legal evidence did the authors of Article X present in the Discussion section?
a. plaintiffs' attorneys should emphasize factual information
b. defense attorneys should emphasize counterfactual alternatives
c. defense attorneys should emphasize factual information
d. juries should remain unswayed by counterfactual statements
e. judges should refuse to allow the introduction of counterfactual statements

16. According to the authors of Article X, what shortcoming of the Article X experiment could a future study remedy?
a. it could assess feelings of sympathy for the victim
b. it could perform the same study using subjects drawn from actual jury pools
c. it could assess feelings of responsibility of the defendant
d. it could vary the amount of time elapsed between the consideration of a counterfactual route and the accident
e. it could assess causal role of the plaintiff

Article Y: Mood and Memory (passive voice version)

How does mood affect memory? Although this question is relevant for how memory operates in everyday life, it has been little investigated outside the psychology laboratory, although the general relationship between affect and cognition has been investigated (see Blaney, 1986, for a summary). Most research to date has focused on testing the theory that mood semantically primes those memories with which it is associated (Bower, 1981; Clark & Isen, 1982). According to this theory, people in a negative mood tend to recall negative memories. Support for this theory has been shown by a number of studies (Bower, 1981; Bower, Gilligan, & Monteiro, 1981; Johnson, Petzel, Hartney, and Morgan, 1983). One of the attractions of this memory theory is that it fits nicely within the framework of a larger associative network theory of semantic memory, a theory in which memory processing is viewed primarily from a cognitive, rather than a social, perspective.

In the Bower, Gilligan, & Monteiro study, subject mood was induced via a post-hypnotic suggestion to feel either happy or sad (subjects were tested one at a time). Both groups were asked to read a fictional narrative of approximately 1000 words about two characters, Andre and Jack, who meet to play a friendly game of tennis. The story is about their discussion as they meet, drive to the tennis court, warm up and play a game of tennis, and get dressed. Both characters discuss their recent activities. One of them, happy Andre, has been enjoying life and is in a buoyant mood. His companion, sad Jack, has had a number of recent setbacks and is rather glum. After the subjects finished reading this story, they were told "it's over." This "trigger" phrase was a pre-arranged signal designed to negate the effects of the mood-inducing post-hypnotic suggestion. Subjects were then questioned about each character in the story. Although both groups recalled approximately the same amount of information, the happy subjects tended to recall more information about happy Andre; sad subjects tended to recalled more about sad Jack.

In another study (Johnson, Petzel, Hartney, and Morgan, 1983), subjects were asked to complete a series of tasks. Task outcome—successful versus unsuccessful—was controlled by the experimenters so that each group had identical success and failure rates. The subjects were undergraduates whose mood—depressed or non-depressed—was ascertained prior to the experiment by having subjects fill out a self-report questionnaire. After subjects finished working on the series of tasks, they

were asked to recall details about the tasks just performed. As hypothesized, depressed subjects were better at recalling failed tasks; non-depressed subjects were better at recalling successfully completed tasks.

Such results demonstrate support for the theory of *mood congruent recall.* Despite these results, an antithetical theory has been proposed, the theory of *mood incongruent recall.* According to supporters of this theory, memories are often used in daily life to regulate mood, i.e., to maintain a balanced emotional state. Prolonged mania and depression are both undesirable states; when depressed, one desires to feel happier; when overly elated, one desires to feel calmer. The mood incongruent theory of recall could account for a psychological regulatory mechanism set in motion to regulate internal states, a mechanism that prompts happy people to recall negative experiences and depressed people to recall positive experiences (Isen, 1984; Singer & Salovey, 1988) .

The theory of mood incongruent recall appears to directly contradict the theory of mood congruent recall. At present, the evidence in favor of mood congruent recall is currently considered to be more compelling. However, this may be because standard laboratory procedures tend to encourage mood congruent recall while discouraging mood incongruent recall. Subjects in most laboratory experiments try to cooperate with the experimenter, to do what they think is expected of them; they may be on their "best" behavior during laboratory sessions. If laboratory procedures are used to subtly encourage the recall of congruent memories, mood incongruent memories may be inhibited; only mood congruent memories may be verbalized. But whether people react this way outside the lab can been questioned.

The present experiment was designed to test the effect of mood upon memory in a naturalistic setting. A reliable elicitor of moods for undergraduate students was employed: the return of midterm exams. The mood-inducing exams were returned at the beginning of class. The subsequent class lecture included a class demonstration involving an autobiographical memory task. Comparing moods with memory responses enabled us to see which theory would prevail in a study simulating a common experience in everyday life.

Method

Subjects. 113 students from a University of Pennsylvania social psychology class were selected to participate the study.

Procedure and Materials. At the beginning of class, a confederate posing as a visiting professor doing research on student perceptions of exams handed out blank cards and asked students to write the grade they expected to receive on the exam and a personal code of their own choosing. After the cards were collected, the graded midterms were returned by the course instructor. Students were asked by the confederate to fill out a one-page exam questionnaire, indicating their actual grade, their code, and their responses to questions about the test. He then left the room.

The second part of the study was designed to appear unrelated to the first. The course instructor announced he would begin the lecture by conducting an experiment to determine the relationship between memory and personality. First students would be asked to recall a memory from a specified time period. Then they would be asked to fill out a brief personality assessment questionnaire.

After the demonstration was explained to the class, the students received packets containing instructions to record an event that occurred during their high school years. They were told their responses would be entirely confidential. They were then asked for the personality test to indicate their present mood, a step that was actually intended to provide a manipulation check on the effect of midterm grade on mood. This step required students to fill out a seven-point scale (where 1 = not at all and 7 = extremely) for Happy, Sad, Emotional, and Confused.

After writing down the memory recalled, students were asked to indicate in writing how they thought getting their exams back affected which memory was recalled.

Coding of Memory. The memory was rated for affective tone by two raters who assessed both positive and negative mood on a seven point scale (where 1 = none and 7 = intense). The inter-rater reliability was quite good ($r = .87$). The ratings of the two judges were averaged for the data analysis.

Coding of Students. Students were divided into two groups based on the relationship between the midterm grade received and their grade expectation. Those receiving a lower-than-expected grade were assigned to the Dissatisfied group; those receiving a grade equal to or higher than expected were assigned to the Satisfied group.

Accordingly, 60% of participants were assigned to the Satisfied group; 40% were assigned to the Dissatisfied group.

Results

The manipulation check was analyzed by comparing the results of the students' self-reported affect on the four dimensions (Happy, Sad, Emotional, and Confused) with student grade performance. As predicted, students in the Satisfied group were, in fact, happier, less sad, less emotional, and less confused than students in the Dissatisfied group. The results of an analysis of variance supported the theory of memory incongruent recall; students in the Satisfied group tended to recall negative memories, $t(112) = 2.29$, $p < .01$; those in the Dissatisfied group tended to recall positive memories, $t(112) = 1.88$, $p < .05$.

However, this inverse effect of mood on memory was not readily perceived; 63% of the students denied any effect and 15% believed their memories to be biased congruently; the remaining 22% made irrelevant comments. The idea that positive memories might be elicited by a negative mood (or vice-versa) was not expressed by any student.

Discussion

In this experiment, students with unsatisfactory grades became depressed while those with satisfactory grades did not. In a subsequent memory recall task, depressed students tended to recall happy memories while non-depressed students tended to recall more negative memories. Thus, these results are consistent with the incongruent mood recall theory, the idea that in everyday life idea, recall is often used to maintain a balanced emotional state. Surprisingly, students appeared to be completely unaware that a negative mood might trigger the retrieval of positive memories, suggesting a rather highly automated mood regulation mechanism.

However, this study has several important limitations that must be considered before firm conclusions can be made. Subjects were not randomly assigned to treatment groups because this would have required the random assignment of midterm grades. For ethical reasons, grades could not be randomly assigned. Thus, it is possible that grade was correlated with other variables (e.g., personality variables, such as motivation). A related problem is that students who performed well on the midterm

have had different high school experiences than those who performed less well (e.g., good students may have had less fun); thus, differences in recall may reflect actual differences in high school experiences rather than an attempt at mood adjustment.

The robustness of these findings could be assessed in future studies. Mood changes could be induced using alternative methods, e.g., performing the experiment on sunny vs. cloudy days (Schwarz & Clore, 1983) or exposing subjects to happy vs. sad music (Kenealy, 1988). Also, the results of the present study could be compared with those obtained in future studies designed to permit random assignment of subjects.

Article Y - Multiple-choice Questions

1. According to the authors of Article Y, which of the following is a characteristic of the theory of mood congruent recall?

a. It supports cognitive theories of memory processing
b. It supports social theories of memory processing
c. It is inconsistent with the network theory of semantic memory
d. a and b only
e. b and c only

2. Which of the following is the BEST example of mood incongruent memory recall?

a. When Sarah was eight, she was sexually abused by a neighbor. She soon repressed any memory of the incident. But when she was 24, she suddenly recalled the incident during a hypnosis session with a therapist. She was deeply shocked by the memory of the incident.
b. On Saturday night, Kim's boyfriend Sam broke off their relationship. Kim became very depressed. All the next day, she kept thinking about the awful moment when he told her the news.
c. Dan witnessed a bank robbery. He initially thought that the robber had red hair and a brown hat but upon further reflection decided the man had actually had brown hair and a red hat.
d. Teresa was in a very bad mood when she took her SATs. As a result, she expected to score poorly. But when she received her scores, she was pleased to find that she had actually performed quite well.
e. Mark was fired after receiving three poor work evaluations in a row. He was understandably discouraged, but consoled himself with the thought that he had only wanted to work until he had enough money to buy a new car.

3. In Bower, Gilligan, & Monteiro's (1981) study involving happy Andre and sad Jack, after the experimenters told the subjects "it's over," what did the subjects do next?

a. read instructions for the second half of the experiment
b. answered questions about story characters Andre and Jack
c. rewrote the story from memory as accurately as possible
d. wrote a one-sentence moral for the story
e. stopped what they were doing and awaited further instructions

4. What did Bower, Gilligan, & Monteiro's (1981) study involving happy Andre and sad Jack demonstrate?

a. strong support for mood congruent memory recall
b. strong support for mood incongruent memory recall
c. weak support for congruent memory recall
d. weak support for incongruent memory recall
e. no support for either incongruent and congruent memory recall

5. In the study by Johnson, Petzel, Hartney, and Morgan (1983), the experimenters asked the subjects to complete a series of tasks and then to recall task details. Which of the following most accurately describes the procedure for this experiment?

a. The experimenters induced subject mood but did not control task outcome.
b. The experimenters assessed subject mood but did control task outcome.
c. The experimenters assessed subject mood but did not control task outcome.
d. The experimenters induced subject mood and controlled task outcome.
e. None of the above

6. Which of the following did Article Y's authors mention as a shortcoming of current laboratory research on mood and memory?

a. Subjects tend to try too hard to please the experimenter
b. Laboratory studies tend to encourage mood incongruent recall
c. Laboratory studies tend to use biased samples
d. a and b only
e. all of the above

7. The experiment done by the authors of Article Y tested the effects of mood on memory:

a. when there was an equal number of subjects in each mood condition
b. when subjects were told that memories tend to be incongruently biased
c. when subjects were told that memories tend to be congruently biased
d. as part of a larger experiment on personality
e. in a naturalistic setting

8. For the first phase of the Article Y experiment, who asked students to fill out blank cards indicating the grade they expected to receive and a personal code?

a. the course instructor
b. the course teaching assistant
c. a professor from another psychology course
d. one of the course peer tutors
e. none of the above

9. For the Article Y experiment, what was the real purpose of asking students to take a personality test?

a. It enabled the experimenters to assess the relationship between personality profile and mood
b. It enabled the experimenters to assess the effect of personality profile and type of memory recalled
c. It had no purpose other than to serve as a "buffer" task between the first and second phases of the experiment
d. It had no purpose other than providing a more plausible cover story for the experiment
e. It provided a manipulation check on the effect of midterm grade on mood

10. For the second phase of the Article Y experiment, the instructor asked students to write down:
a. a memory from junior high school
b. a memory from high school
c. a memory from their first year in college
d. a memory from the previous 12 months
e. a memory from the previous summer

11. For the Article Y experiment, the experimenters divided students into two groups on the basis of:
a. actual grade only
b. expected grade only
c. the expected grade versus actual grade
d. the expected grade versus class average
e. the actual grade versus the class average

12. What percentage of participants in the Article Y experiment were in the SATISFIED group?
a. 30%
b. 40%
c. 50%
d. 60%
e. 70%

13. For the Article Y experiment, students in the DISSATISFIED group:
a. tended to write down a negative memory
b. were more likely to believe their memories were congruently biased
c. tended to write down a positive memory
d. tended to take less time to recall a memory
e. none of the above

14. For the Article Y experiment, how aware were students of the relationship between their mood and the memory they produced?
a. completely aware
b. fairly aware
c. barely aware
d. completely unaware
e. Students were not asked to rate their awareness of the relationship between mood and memory

15. According to the authors of Article Y, which of the following was a limitation of the Article Y experiment?
a. midterm grades were not randomly assigned
b. the manipulation check was inadequate
c. some students failed to write down a memory
d. a and b only
e. all of the above

16. Which of the following did the authors of Article Y mention as a possible activity for future studies?

a. to replicate the current experiment
b. to randomly assign subjects to groups
c. to use alternative methods to induce mood changes
d. a and c only
e. b and c only

Article A — Parenting in Pied Flycatchers (passive voice version)

Animals vary considerably in the degree of responsibility that each parent assumes in caring for and feeding their young. When responsibility for raising the young is shared by both parents, as is the case with many birds and mammals, it seems natural that mate selection would be based to a large extent on a potential mate's perceived parenting ability (Nisbet, 1973; Kirkpatrick, 1985).

Research has demonstrated that breeding success tends to improve with age or previous breeding experience (Curio, 1983; Clutton-Brock, 1988; Newton, 1989). The positive correlation between breeding success and age or experience could be due to a number of factors, the most likely being that breeding by older animals is usually initiated earlier in the season than breeding by younger animals (Lundy, 1995). However, lower rates of breeding success among younger animals could be attributed to a number of other factors: less experience in raising the young (Lack, 1966; Curio, 1983); less effort exerted in raising the young (Gadgil & Bossert, 1970; Pugesek, 1990; Pärt, Gustafsson & Moreno, 1992); existence of a higher proportion of low-quality individuals that die before aging (Curio, 1983; Clutton-Brock, 1988); or inability to attract a suitable mate (Lundy 1995).

In long-lived species, females often remain with their present partners or change partners, depending upon the success of past breeding cycles (Coulson, 1966, Brooke, 1978). However, the female in short-lived species frequently has little opportunity to select a mate based upon past experience. She may have to rely on indirect indicators of breeding success (secondary traits), she cannot evaluate male parental performance prior to mating.

By choosing males with secondary traits that tend to predict parental performance, females are more likely to obtain a mate in good physical condition capable of providing high-quality parental care (e.g., Hoeizer, 1989; Hill, 1991; Lundy, 1995) and siring young of high genetic quality (e.g. Hiligarth, 1990). For example, in the case of the common tern, the female selects her mate based upon the quantity of food that is delivered by the male during courtship (Nisbet, 1973). However, the tendency of females to choose mates with secondary traits that may predict parental performance is not universal; in some species (e.g., western gulls), mate selection is based upon arbitrary factors that appear entirely unrelated to breeding success (Hunt, 1995).

Method

The study was conducted during April-July, 1992 and, 1993, in a mixed deciduous-coniferous woodland area near Oslo, Norway. A deciduous habitat is preferred by pied-flycatchers; in this environment, females tend to lay more nestlings (Lundberg & Alatalo, 1992).

A total of 277 nestboxes of similar size and shape that had been erected for a previous study were used; the nest boxes were 1- 5 meters above the ground. As the males arrived in the nestboxes, they were trapped and ringed with unique combinations of color rings. Their age (one-year versus older than one-year) and breeding experience (first-time breeders versus experienced breeders) were recorded; plumage coloration was also noted (to the nearest the 0-5 value, on a 7-point scale, where 1 = bright, black-and-white upperparts, score 7 = dull, brown upperparts; Drost, 1936). The nestboxes were inspected on a daily basis to assess the males' mating status and the progress of breeding.

Subjects. Male pairs were used for two different experiments on male parental care. In the first experiment, we compared 23 bright-colored male pairs to 23 dull-colored male pairs. In the second experiment, compared 18 male pairs with previous breeding experience to 18 male pairs without previous breeding experience. In the latter experiment, we chose male pairs with similar plumage color. In most cases, the bright-colored males were older than the dull-colored males; similarly, the experienced males were older than the inexperienced males.

Procedure. The same procedure was employed for both experiments. When the broods of the chosen pairs were 8-9 days old (i.e., when the nestlings no longer required continual care by the female), the males were trapped and their body mass was recorded. (The males were trapped on the day before female removal to avoid unnecessary disturbance.) Males were trapped from each matched pair simultaneously. Then nestlings were exchanged between some of the pairs of broods to create broods of equal size (five, six or seven young) and to match brood mass for each pair as closely as possible. The following day the females of each male pair used for the study were trapped; then initial brood mass was recorded.

The females were kept in captivity for 24 hours. During this period, the males were left to feed the nestlings alone. On the third day of the experiment, the males were again trapped. Both male body mass and brood mass (terminal brood mass) were

recorded for the second time (all pairs of birds and broods were measured at the same time of day). Then the females were released.

The change in brood mass during the 24-hour period of female absence (initial brood mass minus terminal brood mass) was used as the primary measure of male parental care. The changes in male body mass was used as a secondary measure of male parental care.

Results

Brood mass for the bright- versus dull-colored males did not differ at the time of female removal. However, during the period of female absence, the broods cared for by bright-colored males were found to have a significantly greater brood mass than the broods cared for by dull-colored males ($z = 2.37$, $n = 23$, $p = .009$, Wilcoxon matched-pairs signed-ranks test, one-tailed). In addition, although all males lost weight during the experiment, greater weight loss was seen for dull-colored males (.33 grams) than for bright-colored males (.22 grams).

In the comparison of similarly colored males having or not having previous breeding experience, the broods cared for by the experienced males were found to have significantly greater brood mass than the broods cared for by the inexperienced males ($n = 18$, $p = .03$, Wilcoxon matched-pairs signed-ranks test, one-tailed). Again, there had been no difference in initial brood mass between the two groups of males ($z = -.96$, $n = 18$, $p > .05$, Wilcoxon matched-pairs signed-ranks test, two-tailed). In both experiments, older males were both more likely to be bright-colored and to have previous breeding experience. Thus, most of the variation in male performance of parental care may be explained by differences in age. The older males sustained their broods significantly better than the younger males ($z = 2.90$, $n = 29$, $p < .05$, Wilcoxon matched-pairs signed-ranks test, one-tailed). Broods of older males also managed better than broods of younger males when plumage color was similar ($z = 1.88$, $n = 12$, $p < .05$, Wilcoxon matched-pairs signed-ranks test, two-tailed).

Interestingly, although males in both experiments tried to compensate for the absence of the female by increasing the number of trips made to the nest with food (feeding effort), feeding effort was not a significant predictor of brood mass.

Discussion

We have presented experimental evidence that male pied flycatchers differ in their ability to provide parental care. In this study, older, more experienced, more brightly-colored males provided better care than younger, less experienced, less brightly-colored males. Thus, at least one male phenotypic characteristic—plumage color—is a potential indicator of a high-quality partner.

A number of positive effects may be produced by the presence of a high-quality partner: (a) higher egg production (Burley, 1988; Norris, 1990); (b) larger broods (Harvey *et al.*, 1985, 1988); (c) high brood survival rates due to superior feeding; and (d) less exertion by the female (Burley, 1988; Norris, 1990).

These results support the position that plumage color reflects male quality in that (a) bright-colored males generally are older than dull-colored males and (b) older males take better care of the brood. Although it could be argued that our methodology—examining brood mass gain after the female was removed from the nest—constitutes an unnatural intervention, the loss of a parent is not un infrequent occurence in nature (Kluijver, 1950). We suggest that this approach provides a more precise measure of parental care, which results in more clear-cut findings than those of previous studies conducted on this species (Alatalo, Lundberg & Stahlbrandt; 1984, Jarvi *et al.*, 1987; Slagsvold & Lifjeld, 1988).

Article A Questions

1. If we compare the breeding practices of older versus younger animals, older animals tend to initiate breeding:
a. earlier in the breeding season
b. later in the breeding season
c. more times than younger animals
d. in a more favorable environment
e. in a less favorable environment

2. In the authors' discussion of factors that might account for differences in breeding success between young and old animals, which of the following was NOT mentioned?
a. differences in experience in raising the young
b. differences in quality of individual (inferior individuals die before they age)
c. differences in ability to attract a mate
d. differences in effort exerted to raise the young
e. differences in the ability to successfully impregnate the female

3. Common terns choose their mates mainly on the basis of _____; western gulls choose their mates mainly on the basis of _____;
a. elaborateness of the male courtship dance; male plumage
b. amount of food the male delivers to the female during courtship; arbitrary characteristics
c. male plumage; arbitrary characteristics
d. arbitrary characteristics; amount of food the male delivers to the female during courtship
e. previous breeding experience; male plumage

4. How do pied-flycatchers normally feed their young?
a. Both parents provide food to the young
b. The female is the main food provider
c. The male is the main food provider
d. The male feeds the young in the mornings; the female feeds the young in the evenings
e. The male feeds the young until nestlings are 8-9 days old; then both parents feed the young

5. Pied-flycatcher females choose mates primarily on the basis of:
a. past breeding success of the male
b. amount of food received from the male
c. male body mass
d. male plumage and territory quality
e. male body mass and plumage

6. In the study done by the authors of this article, the experimenters removed the females from the nest for _____.

a. 6 hours
b. 12 hours
c. 1 day
d. 2 days
e. 3 days

7. The authors used a _____ design for both experiments in their study.

a. multiple baseline
b. case study
c. single subject
d. matched pairs
e. non-experimental

8. This study took place in:

a. the United States
b. England
c. Norway
d. Canada
e. France

9. In what kind of woodland did this study take place?

a. coastal pine
b. mixed hemlock-fir
c. old growth fir
d. deciduous
e. mixed pine and deciduous

10. What is the correct order of the procedure for both experiments in this study?

1 removal of females from the nest
2 trapping of males and recording of male body mass
3 recording of initial brood mass
4 exchange of nestlings

a. 1, 2, 4, 3
b. 1, 4, 2, 3
c. 2, 1, 3, 4
d. 2, 4, 1, 3
e. 3, 2, 1, 4

11. What was the purpose of exchanging nestlings in some broods?

a. to ensure brood pairs of equal size and mass
b. to ensure brood pairs with the same number of males and females
c. to ensure brood pairs with similar plumage
d. to introduce an additional manipulation: trauma versus no trauma
e. to introduce an additional manipulation: strong versus weak broods

12. What was the primary measure of how well males cared for the brood during the female's absence?
a. male body mass
b. male viability
c. brood mass
d. brood viability
e. brood feeding rate upon return of the female

13. In the experiment involving bright versus dull-colored males:
a. both dull-colored and bright-colored males maintained body mass to the same extent
b. dull-colored males lost their body mass; bright-colored males maintained their body mass
c. dull-colored gained body mass; bright-colored males maintained their body mass
d. both dull-colored and bright-colored males lost body mass, but dull-colored males lost more
e. both dull-colored and bright-colored males lost about the same amount of body mass

14. This experiment provided support for the idea that _____ is the best single predictor of breeding success in a male pied-flycatchers.
a. body mass
b. body dimensions
c. breeding experience
d. age
e. plumage

15. The authors discuss potential benefits of females' selecting a superior mate. Which of the following did they NOT discuss?
a. larger broods
b. higher production of eggs
c. less female effort
d. better feeding of the brood
e. less brood attrition due to predation

16. Why did the authors think the results of this study are more clear-cut than the results of similar studies in the past?
a. they used more subjects
b. they studied more than one species of bird
c. they studied characteristics of females as well as males
d. they used an improved measure of parental care
e. they used better-trained observers

Article B — Feeding Bears in Zoos (active voice version)

From the public's perspective, bears are among the most popular zoo animals. They are active, large, impressive, and the subject of considerable folklore. In the recent trend toward naturalizing zoo exhibits, however, zoo managers have generally left the family *Ursidae* behind. Most captive bears inhabit outdoor moated grottos constructed of gunite, with a pool and occasionally logs. Such environments do not adequately display the normal behavioral repertoire and natural history of the family (Melisso & Savarese, 1990). Zoo bears exhibit unnatural, stereotypic behaviors that we seldom see in the wild, e.g., pacing, rocking, and swaying—patterns which in humans reflect a maladaptive pattern of adjustment to environment stresses (Forthman *et al.*, 1992). Researchers studying captive animals have often linked such behaviors to conflict and frustration (Forthman *et al.*, 1992; Wechsler, 1991). In addition to exhibiting stereotypic behavior, captive bears are also prone to engage in begging or masturbation (Bayermeister, 1995).

Thus, the manner in which we exhibit most captive bears fails to promote behavior characteristic of bears in the wild. In particular, traditional feeding procedures—simply putting out food once or twice a day—do not promote the kind of hunting and foraging behavior seen in wild bears. Two negative consequences ensue: this management approach (1) compromises the health and well-being of the bears (Bayermeister & Puff, 1994; Melisso & Savarese, 1990) and (2) hampers efforts to conserve bears in the wild by educating the zoo-going public about bear behavior because the public too often sees zoo bears doing things that the bears would never do in the wild (Forthman & Elder, 1992).

However, it is difficult to provide bears with opportunities to perform species-typical feeding behaviors in standard exhibits. Devising representative feeding exhibits for bears has become more difficult because of the varied ways in which bears feed in the wild. Wild bears are omnivorous. They spend much of their time climbing, digging for small mammals and tubers, tearing apart rotting logs, and consuming seasonal fruits (Graber & White, 1983; Mace & Jonkel, 1986). These activities, combined with a large body size, make it difficult to display bears in exhibits with natural substrates and vegetation (Tsai & Puff, 1991).

There have been two key studies of feeding enrichment in captive bears. In one recent study (Markowitz, 1982), researchers made efforts to reduce the level of combative interactions between a male and female polar bear and to provide them with

alternative behavioral opportunities. They designed a sound-activated fish catapult to enable the bears to "order" fish by vocalizing into a microphone, thereby giving the bears a more active role in the feeding process. This treatment resulted in a decreased incidence of combative interactions by both bears and a decreased incidence of pacing by the male.

Carlstead *et al.* (1991) provided three species of bear—black, brown, and sloth bears—with honey-filled logs. One species—the sloth bears—paced less often as a result. In a second experiment with just the black bears, they compared three feeding regimes: (1) food provided once daily in den; or (2) food provided once daily with additional foods provided from a mechanical tree; or (3) food scattered in the exhibit and hidden in manipulable objects. The third method reduced the bears' stereotypic pacing dramatically when compared to methods one and two.

We performed the present study to provide more species-appropriate behavioral opportunities for four solitary bears: a male Kodiak, a male Asiatic black, and a male and female polar bear at Zoo Atlanta. All these bears exhibited stereotypic behavior, particularly the Asiatic black bear and the female polar bear. Furthermore, visitors frequently expressed concern about the bears, particularly during the summer, when the animals appeared hot and bored.

We provided feeding enrichment to the bears during the summers of 1989 and 1990. We hypothesized that feeding enrichment would increase the amount of time the bears spent active, decrease the amount of time spent passive, and decrease the frequency of undesirable stereotypic behaviors.

Method

Enclosures. The zoo individually housed each bear in an outdoor moated grotto with free access to dens; two of the exhibits were 610 meters^2 in area; the remaining two were 762 meters^2. Each contained a small pool (approximately 137 meters^3) and a variety of logs as "deadfalls." The substrate was composed of stone and concrete.

Subjects. We studied four bears in 1989, and three in 1990: RN, a male Kodiak bear born 1/75; BU, a male Asiatic black bear born 12/85; LA, a female Polar bear born 4/67; and TR, a wild-born male Polar bear, born 1/65, who we euthanized in September of 1989 due to arthritis and severe muscular disease.

Procedure. We conducted this study during two consecutive years. The first phase took place in 1989 between June and October; the second phase, between June and September 1990, with the three remaining bears. We used the same procedure during both phases of the study.

Trained coders observed the bears on three consecutive days per week, during two separate 30-minute observation periods each day. The first observation period (period one) occurred immediately after keepers fed the bears in the morning (approximately 9:30 - 10:00 am); keepers provided bears with enrichment items (described below) at this time. The second observation (period two) occurred approximately two hours later; its purpose was to ascertain whether potential enrichment effects persisted after the keeper removed the enrichment items.

We coded three categories of behavior: (a) passive behavior (e.g., resting; resting but alert; self-maintenance behavior such as grooming; passive orienting to humans); (b) active behavior (object investigation, foraging/ingesting, locomoting, play, swimming/diving, active orienting to humans); and (c) abnormal (masturbation, swaying, repetitive pacing/swimming).

The approach to feeding enrichment had to meet several important requirements: it had to be a regime that (a) we could easily incorporate into the keepers' routines; (b) we could implement without entering the exhibits; and (c) did not exacerbate fire-ant activity, a serious problem during the summer. We were able to devise a feeding regime that met these criteria: for the Kodiak and polar bears, we provided enrichment in form of plain ice blocks and "fishcicles"—whole mackerel frozen into ice blocks. For the Asiatic black bear, we provided fresh browse (tender plant shoots and leaves). During 1990, we varied the stimuli. In addition to fish, we introduced other foods, including ice blocks containing peanuts, apples, raisins, peanut butter, and jelly, plus scattered peanuts, bread, raisins, and sunflower seeds.

We provided enrichment to only one bear per week; thus, during any given week, we collected enrichment data on one bear, while we collected baseline data on the others. We froze ice and fish in 5-7 gallon plastic containers. The keepers threw these into the bear's pool prior to the beginning of the first observation period.

Results

General Effects of Enrichment. When receiving enriched feedings, all the bears were more active ($X^2 = 11.34$, $df = 3$, $p < .01$), less passive ($X^2 = 8.21$, $df = 3$,

$p < .05$), and less likely to engage in abnormal behaviors ($X^2 = 9.33$, $df = 3$, $p < .05$), but only during period one (during the presence of enrichment). However, this relationship did not persist after the bears consumed the enrichment items—during period two, we observed no significant behavioral differences ($X^2 = 4.55$, $df = 3$, $p > .05$).

Comparison by Year. We compared the effects of enrichment in year one (1989) versus (1990) to see whether the greater variety of enrichment provided in year two would result in bigger improvements in behavior than those observed in year one. For this analysis, we included only period one tallies (we excluded period two tallies, as well as the period one tallies for the male polar bear, TR, who we observed only during 1989). Our analysis indicated that, although we observed positive effects due to food enrichment during both years, these positive effects were more pronounced during year two (1990), when we provided a greater variety of food enrichment ($G^2 = 23.7$, $df = 4$, $p < .001$).

Discussion

One of the strengths of zoos as research and educational institutions is their ability to provide access, both to scientists and to the public, to species that are difficult to study or enjoy in the wild, due to low population density, nocturnal habits or remote habitats. It is difficult, however, to conduct valid zoo biology research or to educate the public properly about wild animals, when captive animals inhabit facilities that suppress natural behavior patterns or severely skew their distribution. Moreover, the chronic display of stereotypic behaviors may indicate high stress levels that may compromise the health of the animals; it may also partially account for the low reproduction rates seen in many bear species in captivity (Bloxam, 1977).

The results of this study were similar to those of other studies (e.g., Carlstead *et al.*, 1991; Markowitz, 1992; Wechsler & Puff, 1989). Examining all the bears over both years of the study, we confirmed both the hypotheses that active behavior would increase and that passive and abnormal behavior would decrease in the enriched condition.

This study demonstrated that quite simple and inexpensive methods of feeding enrichment may have significant positive influences on the behavior of captive bears. The bears displayed more natural, more characteristic behaviors, despite the following constraints imposed by the management routine: the keepers fed bears the same

amounts year round, (thus suppressing normal seasonal fluctuations in activity) and they housed the bears alone (eliminating social behavior). Also, the public did not observe bears during those hours when bears are the most active. Thus, behavior that we observed in the presence of food enrichment, while still different from the behavior patterns of wild bears, more closely approximated natural behaviors. The bears' interactions with the enrichment items elicited both an increased incidence in complex manipulative behaviors characteristic of bears and a decrease in maladaptive behaviors such as pacing and swaying.

We speculate that we would have prolonged the enrichment effect had we been able to use larger containers for the ice blocks, so that they lasted longer; those provided lasted only 5-15 minutes. Given sufficient freezer space, we could provide more than one ice block during the day. We would also expect frequent scattering or hiding of small food items throughout the day to prolong effects of feeding enrichment, particularly if the keepers could hide items in the exhibit, in logs, straw or natural substrates. We believe that the simple approach of food scattering would enhance both the well-being of the bears and public understanding of how bears behave in the wild.

Article B Questions

1. Which of the following observations about bear care in zoos does did NOT occur in the article you just read?

a. Current practices do not encourage bears to behave as they do in the wild
b. The costs of upgrading bear exhibits is often prohibitive
c. The public may receive a false impression about how wild bears behave after viewing zoo bears
d. Zoos have not updated bear exhibits to the extent that they have updated other exhibits
e. Much of the problem with developing bear exhibits that show the zoo-going public how bears feed stems from the hard-to-duplicate feeding patterns of wild bears

2. Which of the following are the best examples of stereotypic behaviors?

a. begging and pacing
b. begging and masturbation
c. pacing and swaying
d. resting and sleeping
e. active orienting to humans and pacing

3. Researchers have observed that stereotypic behaviors often occur in the presence of:

a. frustration
b. sexual arousal
c. fear
d. alertness
e. lethargy

4. In an experiment by Markowitz (1982), bears could "order" fish by vocalizing into a microphone. Which of the following statements about this experiment is completely FALSE?

a. The subjects were two polar bears
b. The frequency of combative behavior by the male decreased
c. The frequency of combative behavior by the female decreased
d. The male's pacing decreased
e. The female's pacing decreased

5. Carlstead et al. (1991) performed 2 experiments. In experiment 1, three species of bear received honey-filled logs. In this experiment 2, the experimenters compared three feeding regimes for one species of bear: (1) food given once daily in den; (2) food given once daily with additional foods provided from a mechanical tree; and (3) food scattered in the exhibit and hidden in manipulable objects. Which of the following statements about the results of this study is completely TRUE?

a. In experiment 1, providing honey-filled logs reduced stereotypic pacing in all species
b. In experiment 1, providing honey-filled logs failed to reduce the incidence of pacing in the sloth bear
c. In experiment 2, both the second and third method reduced the incidence of stereotypic pacing in the black bear
d. In experiment 2, the third method reduced the incidence of stereotypic pacing in the black bear
e. In experiment 1, providing honey-filled logs reduced the incidence of pacing in the sloth bear; in experiment 2, the second and third method reduced the incidence of stereotypic pacing in the black bear

6. During year one of the experiment (1989), the experimenters used different methods for feeding the bears and then observed their behavior during two periods. Period one occurred ____; Period two occurred _____.

a. during food enrichment; 1 hour after food enrichment
b. during food enrichment; 2 hours after food enrichment
c. 1 hour after food enrichment; 2 hours after food enrichment
d. 1 hour after food enrichment; 1 day after food enrichment
e. 2 hours after food enrichment; 4 hours after food enrichment

7. Which bears served as subjects during year two of the study (1990)?

a. two polar bears, one Kodiak bear
b. one polar bear, one black bear, one Kodiak bear
c. two brown bears, one polar bear
d. two black bears, one polar bear, one brown bear
e. two black bears, one polar bear, one Kodiak bear

8. Observers categorized bear behaviors during two observation periods. Which of the following belong to the same category?

a. resting but alert, active orienting to humans, self-maintenance behavior
b. foraging/ingesting, locomoting, active orienting to humans
c. resting, passive orienting to humans, swaying
d. swaying, masturbation, self-maintenance
e. grooming, active orienting to humans, resting but alert

9. Which of the following types of infestation did the authors cite as a serious problem for zoos in the summer?

a. mice
b. lice
c. fly
d. termite
e. ant

10. In year one (1989), _____ bear(s) received enrichment for _____ week(s) while experimenters collected baseline data on the remainder of the bears.
a. one; one
b. one; three
c. two; one
d. two; two
e. two; three

11. In year one (1989), food enrichment consisted of:
a. fish in ice chunks
b. fish in ice chunks and fresh plant food
c. fresh plant food in ice chunks
d. frozen fish and fresh plant food
e. fish and fresh plant food in ice chunks, fresh fish, and fresh plant food

12. During year one (1989), when the researchers compared bears that received enrichment to bears that did not, bears that received enrichment:
a. demonstrated fewer abnormal behaviors during both periods one and two
b. demonstrated fewer abnormal behaviors, but only during period two
c. demonstrated higher levels of activity during period one and fewer abnormal behaviors during period two
d. demonstrated fewer abnormal behaviors during period one and more activity/less passivity during period two
e. demonstrated more activity/less passivity and fewer abnormal behaviors, but only during period one

13. Which of the following statements comparing 1989 versus 1990 observations is completely TRUE?
a. The stimuli were identical; the effects of the stimuli were identical
b. The stimuli were identical; the effects of the stimuli were stronger during 1989
c. The stimuli were identical; the effects of the stimuli were stronger during 1990
d. The stimuli were not identical; the effects of the stimuli were identical
e. The stimuli were not identical; the effects of the stimuli were stronger in 1990

14. The authors describe the bears' living conditions and treatment during the study. Which of the following statements was NOT part of this description?
a. Each bear lived in a separate enclosure
b. The bears received the same amounts of food each day
c. The bears received regular vitamin supplements each day
d. Each enclosure contained a pool and some logs; a moat surrounded each enclosure
e. The public did not view the bears during the hours when bears are the most active

15. When the authors discussed why they observed behavioral improvements only during period one (not period two), they speculated that the effects would have lassted longer if:

a. there had been more subjects
b. public viewing of the bears had not been permitted
c. the weather had not been so hot and humid
d. they had provided honey-filled logs as enrichment
e. they had used larger chunks of ice

16. What did the authors recommend after completing their study?

a. feeding zoo bears a variety of fresh plants enclosed in ice chunks
b. trying to varying the amount of food zoo bears receive throughout the year in order to more closely match seasonal fluctuations seen in the wild
c. placing food tidbits in zoo bears' enclosures throughout the day
d. conducting a follow-up study with a larger number of bears
e. conducting a follow-up study with the same bears to look at the long-term effects of food enrichment

Appendix E:

Rating Sheet for Student Lab Reports

Rating Sheet for Scientific Lab Reports

Scientific Content - quality rating of instructions and discussions

Rating of title____

0=v poor	title missing or nonsensical
1=poor	title vague, ambiguous or confusing
2=mediocre	title somewhat vague or ambiguous, less specific than it should be
3=good title	clear, specific, and easy to understand
4=v good	title highly descriptive, easy to understand, professional-sounding

Introduction:
____gen'l research question
____research done
____research implications
____current study: specific hypothesis or research question

Discussion:
____tie to hypothesis
____meaning of results
____limitations of study
____applications/future research

0=v poor	very few ideas or highly inaccurate ideas; ideas very poorly developed
1=poor	ideas poorly developed; very few details, ideas not clearly delineated
2=mediocre	main ideas present but details only partially developed, could be learer
3=good	both main ideas and details clearly developed
4=v good	ideas exceptionally well developed; main ideas clearly presented and details extensively developed

Organization: Degree to which (1) content sections are in same order as content sections above, (2) ideas within content sections are well-organized, (3) transitions between content sections are easy to follow.

____intro organization ____discussion organization

0=v poor	nonexistent sequence of sections, ideas within section incoherent, no transitions
1=poor	major weaknesses in sequence, org within sections and transitions
2=mediocre	some weaknesses in sequence, org within sections and transitions
3=good	acceptable seequence, org within sections and transitions
4=v good	excellent sequence, org within sections and transitions

Logical Connections between Sentences (indicate paragraph breaks with *slashes*)

Sentence connections—Introduction

— — — — — — — — — —

— — — — — — — — — —

Total____ average score____

Sentence connections—Discussions

— — — — — — — — — —

— — — — — — — — — —

Total____ average score____

0=v poor	sentences totally unconnected; nonexistent relationship between ideas
1=poor	sentences poorly connected; relationship between ideas is not at all clear
2=mediocre	sentences connected, but specific relationship between idas is not very clear
3=good	sentences well connected; relationship between ideas is clear
4=v good	sentences extremely well-connected; relationship between ideas is crystal-clear

"For staff use only"
Unique ID_____ Course_____ Instructor_____ Classification_____

Appendix F:

Correlations Between Lab Report Ratings

Variable	Correlation between Raters	Prob.
Content variable—paper title	.59	.006
Content variables—Introduction:		
Rating of research question	.70	.001
Rating of past research	.60	.005
Rating of implications	.63	.003
Rating of hypothesis	.65	.002
Content variables—Discussion:		
Rating of tie to hypothesis	.57	.009
Rating of results interpretation	.55	.012
Ratings of study limitations	.69	.001
Ratings of future study	.69	.001
Organization variables:		
Organization—Introduction	.67	.001
Organization—Discussion	.68	.001
Sentence cohesion variables:		
Sentence cohesion—Introduction	.24	.31
Sentence cohesion—Discussion	.35	.13

Appendix G:

Examples of Weak and Strong Student Lab Reports

On the following pages are two student lab reports, one weak and one strong. Both were written for Psych 231 (Laboratory in Cognitive Psychology) and were one the same research topic. Figures and references are not included. All formatting and errors from the originals have been preserved.

Example of a weak Psych 231 paper:

How Short Term Memory is Used

Prior to Sternberg paradigm, was the Donders (1968) experiment. In Donders experiment, they used the subtraction method to determine the reaction time of certain tasks. Donders would simple subtract some reaction time with the overall time and called that the reaction time that the subjected needed for the specific task. This paradigm relied on certain assumption that when a subject did a task, the subject would do the task sequentially in memory. In this experiment, we are trying to find the way the subjects will search through memory, when given a simple task of remembering two to six digits. Three different ways a subject might search their short term memory, Parallel Self-Terminating Search, Serial Self Terminating Search, and Serial Exhaustive Search.

In the Sternberg experiment, he was giving subjects a memory set size that was a series of random digits of zero to nine. The size of the set size varied from two to six digits in length. After memory set size was displayed, the subjects would see the probe, and ask whether or not the probe was in the memory set. The reaction time was recorded as how long it would take each subject to answer the question. Sternberg found that there was a correlation between the set size and the reaction time and also the slopes of the lines were parallel. Sternberg found in his experiment that the subject would use the serial exhaustive search. We will try to replicate the Sternberg experiment. With our collected data, we will be able to agree or disagree with Sternberg conclusion.

Methods

Subjects

Nine undergraduate subjects from the University of Washington Psychology class participated in the experiment.

Apparatus

The stimuli presented by the program, Psyscope, on a Macintosh computer with a 14 inch color monitor. The subject used a box with three buttons that is connected to the Macintosh with a cable. The computer was used to record the subjects responses and reaction time.

Design

The experiment was completely within subjects design. The two independent variables are size of the memory set and the probe type. The size of memory set had five levels 2,3,4,5,6 and the probe type had two levels, positive and negative. The dependent variable was the reaction time required for each subject to determine how long it would take to make a decision. The design was a X 5 factorial (probe type X set size).

Stimuli

Fifty trials are in this experiment. They had a set size ranging from two to six digits. Both the memory set and the probe ranged from zero to nine. After the digits were displayed one other digit was displayed in the same location. The subjects were to remember if that last digit was in the previous set of numbers.

Procedure

The experiment consisted of 5 parts. The first section was the instructions. The second section was the ten practice trials, one per set size. The third section reminded the subject what the three buttons were used. The fourth section was the fifty trials (five per condition). The final section was the closing comments that thanked the subject for participating in the experiment.

The subjects are led into the computer room and were sitting in front of the computer A. The subjects were told to read the instruction on the screen and to ask if they had any further question about this experiment. Once they read all the instructions and had any questions answered, they were ready to start the experiment. The subjects were given ten practice trials, one trial for each condition to familiarize the subjects with the experiment. No data was collected at this time. The sequence of events is listed below.

First the participants are led through a demonstration trial.

In this experiment you will see a list of digits, one digit at a time.

The list will vary in size from 2 to 6 digits.

Your task is to remember all the digits in the list.

Press the Green button when you are ready to memorize the first list.

If the probe was in the list, push the red button.

If you press the wrong button, you will hear a beep.

Again, at the first of each trial you will press the GREEN button to begin.

Use your middle and index fingers of the hand you write with. Place one finger on the RED button and one finger on the YELLOW button.

Remember, your responses are being timed and you should respond as FAST and as ACCURATELY as you can.

When you are ready to begin press either the yellow or the red button.

Results

We did a regression analysis on both the negative and positive reaction times. For all of the reaction times, we converted all of the times greater than 2000 ms to 2000 ms. We found that the relationship between the two probes was reliable, the negative probe $R(214)=.225$, $p < .05$ and for the positive probe $R(216)=.312$, $p < .05$. See figure 1.

This indicated that the reaction time and the set size are correlated to each other. To make sure these two slopes were the same, we did a T-test on the slope. The result of this T-test was that the $t(433)=.00598$. The reaction time is shown below:

	POSITIVE (in ms)	NEGATIVE (in ms)
Minimum	446.043	496.280
Maximum	896.600	1108.400
Range	450.557	612.120
Mean	701.041	799.639
Standard Dev	157.032	206.510
Std. Error	52.344	68.837

Another T-test, that shown the positive probe was reported more quickly then the negative probe, $t(8)=2.611$, $p < .04$.

Discussion

The results of the experiment support the hypothesis that subjects use a serial exhaustive search. If we looked at the diagrams Sternberg did on how people search through their memory, we would see our graph would be very similar to Sternberg's graph. The slopes of both the negative and positive lines are parallel to each other, where the negative lines is higher than the positive. Which means no matter what the memory set size is there is a direct correlation between the negative and positive probe reaction time. On the average, the subject took 100 ms longer to respond to the negative probe than did they did with positive probe.

Some possible directions for future research on short-term memory would be to investigate the amount of time it takes the probe to be shown and if there is any relationship to the reaction time. For example, instead of having the probe show up right after the last digit, allow some fixed interval of time before till showing the probe.

They following researchers that have found evidence for the other two searches. Neisser, Jonides, and Wall (1972) found evidence for parallel search self-terminating search through long-term memory. It indicates that if different task is used in the experiment, we can and will get different results from the data.

Example of a strong Psych 231 paper:

A Serial Exhaustive Search is Used in the Sternberg Paradigm

There has been a great deal of research in cognitive psychology regarding the speed of mental processing. The subtraction method, originating with F.C. Donders in 1868, makes use of the concept that mental tasks involve a number of processes and that these processes take a measurable amount of time to complete (Donders, 1969). What makes this concept complicated is that it is impossible to empirically measure separate times for processes that depend on each other (i.e. measuring time to respond to a stimulus separate from time to encode stimulus, since the former depends on the latter.) The subtraction method is an attempt to tease out the various processes by inserting steps into a simple reaction task and then subtracting the time to do the simpler task to isolate the time required to do the added task. Combinations of this model allowed Donders to isolate discrimination time and motor choice time. Because the subtraction method relied on certain questionable assumptions, the additive factors method was developed by S. Sternberg (1966). Instead of adding or deleting steps in a reaction, this method alters the length of processing for a particular step.

The Sternberg memory-scanning paradigm was first used to demonstrate the additive factors method (Sternberg, 1966). This involved presenting a visual memory set and then a probe. The memory set was a series of random numbers (0 to 9) presented sequentially and was a specific size (2 to 6 numbers). The probe was a number that was either contained in the memory set (positive probe) or was not contained in the set (negative set).

Participants chose whether or not the probe was contained in the memory set by pressing a yes or no button. The reaction time was recorded as the time from the presentation of the probe to the button press. Using this method, Sternberg found that reaction time was positively correlated with set size for both positive and negative probe conditions. He also found that for both the positive and negative conditions, the degree of these linear correlations was very similar. Using the latter result, he was able to make assumptions about how short term memory is scanned and concluded that participants used a serial exhaustive search. The main assumption was that for both the negative and positive conditions, a subject must serially search the complete memory set before being able to respond with the correct answer. This is supported by (a) the

positive linear correlation between set size and reaction time for both probes and (b) the similarity of these two correlations.

From these results, Sternberg came up with four hypothesized cognitive stages thought to exist between presentation of the probe and response: stimulus encoding, serial comparison, binary decision, and response organization and execution. The slopes of the two linear regression lines, which are the same for the positive and negative probes, are useful for determining the times it takes to make one serial comparison. This is because for each increase in set size, there is a corresponding increase in reaction time. The differences between the two intercepts of these lines, (negative is greater than positive) is also useful for isolating the effects of positive versus negative probe conditions in the binary decision stage. Thus, instead of using the Donders subtraction method where whole stages were deleted or added to the task, Sternberg was able to alter the time spent in each stage by manipulating the level of each variable.

The present experiment was a replication of Sternberg's original paradigm. Based on his previous research, it was hypothesized that (a) set size and reaction time would be positively correlated, (b) this correlation would be linear, and (c) the correlation for both positive and negative probe conditions would be similar. Furthermore, by satisfying these three elements, participants would demonstrate use of the serial exhaustive search.

Methods

Subjects

Nine undergraduate students participated in the experiment as part of a psychology course requirement at the University of Washington.

Apparatus

The program Psyscope on a color Macintosh computer was used to present stimuli and record data. A box with three buttons was attached to the computer and was used to record subjects responses and measure reaction time.

Design

This experiment was a within-subjects design. There were two independent variables, probe type and memory set size. The former had two levels, positive and negative; and the latter had five levels, 2, 3, 4, 5, and 6. The dependent variable was the reaction time needed to respond to the probe.

Stimuli

There were 50 total trials, 10 per each memory set size, of which half were followed by a positive probe and half followed by a negative probe. Both the memory set and probe consisted of the numbers ranging from 0 to 9.

Procedure

At the computer terminal, subjects were told to follow the instructions on the screen and were led through a demonstration of the proper use of the recording box. Once completed, they began a 10 trial randomized practice session consisting of 1 positive trial and 1 negative trial for each memory set size. No data was collected for these first 10 trials. Following this practice session, subjects completed the 50 experimental trials.

The following is a sequential list of the experimental protocol:

1. To begin each trial, subjects were prompted by the computer to press the green button as an indication of readiness.
2. The screen then went blank and digits were presented one at a time. A minimum of 2 and a maximum of 6 digits were presented at this time.
3. Again, the screen goes blank and a single digit is presented with this message: Was this a digit in the list? The digit, the probe, was either a positive probe, i.e., contained in the previous presentation of digits, or a negative probe, i.e. not contained in it.
4. Subjects then pressed one of two buttons to answer this question. If the probe was in the presentation, they were to press the red button and if it was not in the presentation, they were to press the red button and if it was not in the presentation, they were to press the yellow button. Time between presentation of probe and answer was recorded by the computer.
5. Subjects repeated 1-4 until all trials were completed.

Results

Data for any incorrect responses was thrown out. Regression analysis was completed to address the relationship between set size and reaction time for both positive and negative probe conditions across all subjects (see Figure 1). This was found to be statistically significant for both the positive condition, $R(216)=.31$, $p<.05$, and negative condition, $R(214)=.23$, $p<.05$, indicating that set size and reaction time are positively correlated. To determine whether these two correlation values were statistically the same, a t-test was used and there was no statistical difference, $t(433)$

=.00598, ns, indicating that the degree of correlation between set size and reaction time for both positive and negative conditions is the same.

The mean reaction time for all subjects was 70ms in the positive condition and 800ms for the negative condition (see Figure 2). A correlated t-test was used and a statistically significant difference was found, $t(8)=2.6112$, $p<.05$, indicating that reaction time was greater for the negative condition than for the positive condition, irrespective of set size.

Discussion

The results of this experiment support the hypothesis that participants use a serial exhaustive search through two main findings: (a) The correlation between set size and reaction time is linear and positive for both probes, and (b) the two correlations for positive and negative probe conditions are the statistically the same. Furthermore, by applying Sternberg's method of using the slope of the regression line as a determination of how long it takes for one serial comparison, it is found that the time to make one comparison is approximately 55ms.

Another finding is that changing the level of the probe altered the reaction time. On average, it took 99ms longer to respond when there was a negative probe than when there was a positive one. No reasonable explanation could be found for the existence of this phenomenon.

The present experiment exactly replicated the findings of Sternberg (1966). Although this is impressive, it is important to note that by using other tasks, researchers have been able to demonstrate the use of different short term memory searching methods other than the serial exhaustive search (Egeth, Jonides, & Wall, 1972; Neisser, 1963; Neisser, Novick, & Lazar, 1963). Hopefully in the future, examining a multitude of research paradigms will elucidate the nature of short term memory searching.

Susan Rhodes
2816 Northeast 117th Street
Seattle, Washington 98125
206/364-7007

Education

Ph.D. in Cognitive Psychology, December 1997
University of Washington College of Arts and Sciences
Seattle, Washington

M.S. in Scientific and Technical Communication, June 1988
University of Washington College of Engineering, May 1988

M.Libr. in Librarianship and Information Science, June 1986
University of Washington Graduate School of Librarianship and Information Science,

B.A. in Speech Pathology and Audiology, June 1973
University of Denver, Denver, Colorado

Research and Teaching Experience

Instructor for U.W. Extension Communication in the Workplace Certificate Program (1989-present). *Have developed and taught two courses on effective writing in the workplace (one on business writing basics, one on writing persuasive documents).*

Instructor in Technical Communication Department, Technical Communication Department, University of Washington (1986-present). *Currently an instructor for TC 401C (Style in Technical Communication), correspondence version, for which I designed a 207-page Study Guide. I previously designed a correspondence Study Guide for ENGR 331 (Advanced Technical Writing) which resulted in the immediate tripling of course enrollment. Have also taught ENGR 331 for the Technical Communication Department.*

Teaching Associate, Psychology Department, University of Washington (1993-present). *Currently PSYCH 231 (Human Performance Laboratory) Assisted in all phases of course organization, instruction (quiz sections), and evaluating student performance in four large undergraduate psychology courses: PSYCH 101 (Introduction to Psychology), PSYCH 205 (Personality), PSYCH 209 (Research Methods in Psychology), and PSYCH 345 (Social Psychology).*

Research Associate, Human Performance Laboratory, University of Washington Psychology Department (1992-93). *Analyzed conversational discourse patterns of pilots during 747-400 Boeing simulator flights to determine the effects of glass cockpit innovations on pilot communication.*

Research Associate, National Center on Postsecondary Teaching, Learning & Assessment, The Pennsylvania State University, University Park, Pennsylvania (1991-92). *On-site coordinator for research assessing the effects of Freshman Interest Groups at the University of Washington.*

Research Associate, University of Washington Psychology Department (Spring 1991). *Developed a successful proposal for funding a departmental writing center for the Psychology Department, University of Washington, Seattle, Washington. Staffed by a half-time faculty member and several graduate psychology students, the Psychology Writing Center has provided individual writing feedback and writing workshops to psychology students since 1991 to the present.*

Research Associate, Software Usability Project, Microsoft Corporation (Summer 1987). *Administered software usability tests and led videotaped group discussions with test subjects.*

Freelance Writer, Editor, Indexer and Document Designer (1980-present). *Have written, edited, and indexed a variety of publications including manuals, book reviews, technical articles, theses. Designed and produced a 55-page course catalog for the U.W. School of Engineering (Summer 1988). Editor-in-Chief for* ***Postcom****, the U.W. Technical Communication Department alumni publication (1989).*

VISTA Volunteer (1981-82). *Researched and wrote a Request for Proposal soliciting bids for supplying fuel oil to low-income families, Portland Action Committees Together, Portland, OR. Also set up a drug information center for Oregon Free From Drug Abuse, Tigard, OR (state-wide organization of parents combating drug abuse by children).*

Conference Manager, Cascadian Regional Library, Eugene, Oregon (1977-79). *Directed all phases of conference organization (e.g., budgeting, publicity, housing, and food services) for conferences and retreats of 50 - 600 participants. Also served as Assistant Editor for* ***Cascade: Journal of the Northwest.***

Publications

The Active and Passive Voice are Equally Comprehensible in Scientific Writing. **Doctoral Dissertation, Psychology Department, University of Washington (1997).** *A series of studies examining the use of passive voice in scientific writing and its effect on reading comprehension.*

University of Washington Distance Learning Study Guide, ENGR 331 - Advanced Technical Writing. **Seattle: University of Washington Extension (1989).** *108-page study guide developed for students in the correspondence version of ENGR 331, Advanced Technical Writing.*

Mandatory Plain Language Statutes: Their Real and Imagined Effects. **Master's Thesis, Scientific and Technical Communication Program, University of Washington (1988).** *Psycholinguistic critique of laws that require plain language for consumer contracts, 1988.*

University of Washington Distance Learning Study Guide, TC 401 - Style in Technical Communication. *Seattle: University of Washington Extension (1989). 202-page study guide developed for students in the correspondence version of TC 401, Style in Technical Communication.*

"The Purpose of Consumer Contracts: Communication or Performance?" ***International Professional Communication Conference Proceedings*** **(October 1988), Seattle, Washington, 313-317.** *Article based on Master's Thesis (see above).*

Oral Presentations

"The Effects of Data Link, Workload Stress, and Relevant Experience on Crew Performance in a B747-400 Simulator, Phase I" (Human Factors Conference, 1994). Seattle: University of Washington, Organizational Research Group. *Analysis of pilot conversations to study the effects of glass cockpit technology on cockpit communication patterns and other crew performance variables.*

"How to Write Your Resume and Cover Letter" (1993). *Two-hour workshop on resume writing for undergraduate and graduate psychology students sponsored by the U.W. Psychology Writing Center. The written script, overheads, and handouts developed for this workshop were designed for ongoing use by U.W. Psychology Writing Center staffers.*

"Writing Centers as Discourse Communities: Departmental Writing Centers at the University of Washington," College Composition and Communication Conference, Cincinnati, Ohio (March 1992). *Discussion of U.W. departmental writing centers as supportive writing environments within discourse communities.*

"The Dilemma of Discourse Conventions: Can Style and Substance Be Mastered at the Same Time?" (July 1991). *One of three papers on a panel discussion on the rhetoric of psychology, 10th Annual Penn State Conference on Rhetoric and Composition, State College, Pennsylvania.*

"Disposable Prose and the Ethos of Behaviorism: Writing for the APA," with Norm Thompson (July 1991). *One of three papers on a panel discussion on the rhetoric of psychology, 10th Annual Penn State Conference on Rhetoric and Composition, State College, Pennsylvania.*

"Rhetorical Devices in Legal Argumentation," with Norm Thompson (July 1990). *Paper presented at the 9th Annual Penn State Conference on Rhetoric and Composition, State College, Pennsylvania.*

"Writing the Power Memo" (May 1991). *Featured presentation for the Washington State Association of Home Economists, Seattle, Washington.*

"How University of Washington Teaching Assistants Perceive TA Training Programs," International TA Training Conference, Seattle, Washington (November 1989). *Survey of 214 teaching assistants on TA training at the University of Washington.*

"Creating Writing Communities in a Large University," West Coast Writing Centers Association Conference, Tacoma, Washington (October 1989). *Discussion of writing centers at the University of Washington comparing discipline-based versus interdisciplinary writing centers.*

"The Purpose of Consumer Contracts: Communication or Performance?" International Professional Communication Conference, Seattle, Washington (October 1988). *Presentation based on Master's Thesis (see above).*

Honors

Psi Chi Psychology Honor Society member (1995-1997).
Hayek Fund for Scholars (Summer 1991). *$1,000 travel award from the Institute for Humane Studies to present two papers at the Penn State Conference for Rhetoric and Composition, State Park, Pennsylvania.*

Claude R. Lambe Fellowship (1990-91). *$4,000 award from the Institute for Humane Studies for academic excellence and interest in classical liberal thought.*

Scholarship for attending "Career Development, Strategies, and Opportunities: A Seminar for Graduate Students," (November 1990). *Travel and living expenses for a three-day invitation-only career advancement seminar sponsored by the Institute for Humane Studies, Washington, DC.*

Pew Teaching Leadership Award (1989). *Scholarship from the Pew Charitable Trust to attend the Second National Conference on the Training and Employment of Teaching Assistants, Seattle, Washington.*

Claude R. Lambe Fellowship Runner-Up Award (1989-90). *$2,000 award from the Institute for Humane Studies for academic excellence and interest in classical liberal thought.*

Liberty and Society Seminar Scholarship (1987). *$400 award from the Institute for Humane Studies to attend Liberty and Society Summer Seminar, Belmont, California.*

Graduate and Professional Senate Student Senator (1986-88). *Elected to represent Scientific and Technical Communication Program in the Graduate and Professional Student Senate, University of Washington, Seattle, Washington.*

Graduated *With Distinction* from University of Washington School of Librarianship (1986).

1986 Student Paper Award, Northwest Chapter of the American Society for Information Science. *Wrote the winning entry for a competitive student paper contest. My article was "Manufacturers as information providers: Meeting the information needs of microcomputer users."*

1984-85 Genevieve C. Cobb Scholarship. *$1,000 award given to U.W. first-year library students with academic promise.*