DISCOURSE PROCESSES 18, 271–296 (1994)

# In Search of Complete Comprehension: Getting "Minimalists" to Work

JULIE FOERTSCH
MORTON ANN GERNSBACHER
*University of Wisconsin–Madison*

Three experiments illustrated that readers will not completely comprehend the sentences they read unless sufficiently motivated by situational demands. Complete comprehension of a topic is defined as the ability to accurately redescribe that topic in one's own words, and it entails three separate yet interdependent processing tasks: (a) activating the information contained in a topic, (b) resolving the topic as a new topic or as an anaphor referring to an old topic, and (c) modifying one's mental structures to organize the additional information that is received. Each process hinges on the outcome of those that preceded it, and comprehenders are not expected to initiate the next process in the sequence unless it is required or motivated by task demands. To test these predictions, three experiments were conducted in which participants were prompted to engage in one, two, or all three comprehension processes after reading two-clause conjunctive sentences. The results suggested that experimental participants had a strategy of minimal task satisfaction: They did not resolve anaphors, build structures, or draw inferences unless it was necessary for completion of the experiment.

Human discourse revolves around topics. Whether we are listening to conversation or reading text, comprehending the material involves identifying sentence topics and associating new information with them. What or whom is being discussed? Is the sentence topic one we already know something about, or is it a new and unfamiliar one? In short, to what or whom should we attach the information we are currently receiving? Our understanding of the material will be greatly influenced by how we interpret the topic of a sentence and how much we believe we already know about that person, place, or thing. For this reason, it is important to examine how processing sentence topics influences the comprehension of discourse.

---

This research was funded by grants from the National Institute of Health, NIH Grant No. ROI NS 29926, and the Army Research Institute, Grant No. DASW0194-K-004. We thank Caroline Bollinger, Maureen Marron, Stephanie Gomez, Paul Schultz, and Rachel Robertson for help in testing our subjects and coding data. We also thank Jen Deaton, Matt Traxler, Peter Kruley, Rachel Robertson, Art Graesser, and Rolf Zwaan for their helpful comments on earlier drafts of this manuscript.

Correspondence and requests for reprints should be sent to Julie Foertsch, Department of Psychology, 1202 W. Johnson Street, University of Wisconsin–Madison, Madison, WI 53706–1611.

But what exactly is comprehension? Here we define *complete comprehension* of a topic as comprising three separate yet interdependent processing tasks: (a) activating the information contained in a topic, which involves computing a meaning of the topic that is relatively context-independent (e.g., "the woman" is a definite, singular, female person); (b) resolving the topic as either a new discourse topic or as an anaphor—an old discourse topic whose referent must be found (e.g., deciding whether "the woman" refers to someone who was mentioned previously); and (c) structure building, as proposed by Gernsbacher (1990), which involves modifying one's mental structures to accommodate the new information that is received (e.g., mapping information about "the woman" onto a previously created or newly created mental structure). Each of these comprehension tasks—activation, anaphoric resolution, and structure building—hinges on the outcomes of the tasks that preceded it, and each task adds to the amount of time and effort needed to complete the comprehension process. *Partial comprehension,* on the other hand, occurs when only one or two tasks—but not all three—have been used in processing the information.

Our three-stage model of comprehension is similar to the "given–new" strategy proposed by Haviland and Clark (1974). Like Haviland and Clark (1974), we believe that comprehenders must decide which information is given and which information is new and must build and modify mental structures to attach information onto the appropriate antecedent. However, we extend Haviland and Clark's proposal by distinguishing between the tasks of anaphoric resolution (deciding whether two topics are the same or different) and the task of structure building (attaching information to old structures or laying the foundations for new ones). We also suggest that these three stages do not always run to completion, as if comprehension were always the same unitary process. Rather, we suggest that comprehension, in some situations, is only partial and does not include every stage. Some experimental participants may proceed only as far as is necessary to satisfactorily complete a given reading task and might stop after the first or second stage if that is all that the experiment requires.

Let us look at each of these tasks in more detail in order to understand their separate yet interdependent effects on the comprehension of sentence subjects. Let us assume that true comprehension involves understanding a topic well enough that one can redescribe it accurately in one's own words. Because building structures to accommodate our knowledge seems to be such an important task in truly understanding the topics we encounter, this task shall be discussed first, even though it may be the last comprehension task to occur during sentence processing.

According to the structure-building framework described by Gernsbacher (1990), comprehenders build mental models or structures to organize information about various topics as they are encountered. These topic-grounded structures must be updated throughout discourse. Information that coheres with the current topic may be attached or "mapped onto" that topic's structure, but whenever a new discourse topic is introduced, the comprehender must shift from the current topic's structure and begin laying the foundation for a new one. The proposed process of shifting explains an increase in reading time for sentences or clauses that introduce a new topic (Anderson, Garrod, & Sanford, 1983; Dee-Lucas, Just, Carpenter, & Daneman, 1980; Haberlandt, Berian, & Sandson, 1980; Mandler & Goodman, 1982; Olsen, Duffy, & Mack, 1984). The structure-building framework further supposes that it is easiest to access information on the structure we are currently processing. Thus, questions about old discourse topics, whose mental structures are less fully-activated, will be responded to more slowly and with less accuracy than questions about current topics (Anderson et al., 1983; Clements, 1979; Gernsbacher, 1989; Mandler & Goodman, 1982). In short, either shifting to an old topic's structure or creating a new topic's structure will increase processing time and slow comprehension.



Deciding whether a topic is old or new is not as straightforward as it seems, and this is where the use of anaphors comes into play. Anaphors, such as the pronoun *she,* are linguistic devices that refer back to previously-mentioned concepts, and they are one way of connecting ideas. Although pronouns are the most prototypical anaphors, repeated nouns, synonyms, and common nouns are also used. When an anaphor appears in discourse, comprehenders must decide to what or whom the anaphor refers. In other words, they must uniquely identify the anaphor's referent before they know how to modify their mental structure (Gernsbacher, 1990).

An effective anaphor is one that will enhance the activation of its referent and suppress the activation of nonreferents. In situations where an anaphor has more than one possible referent, anaphors that are more specific boost the activation of their intended referents better than anaphors that are less specific (Chang, 1980; Corbett & Chang, 1983; Gernsbacher, 1989; Haviland & Clark, 1974; Yekovich & Walker, 1978). For example, if *John* and *Greg* are both current topics, a nonspecific anaphor like *he* can cause confusion and slow comprehension. However, we need a specific anaphor only when it is necessary to distinguish among several likely referents. If only one topic is the focus of discourse, as occurs early in discourse or whenever old topics are out of focus, then a highly specific anaphor may actually suppress its intended referent. Consider the following sentences:

(1) *Mike* called for a taxi, and *he* waited downstairs.

(2) *Mike* called for a taxi, and *the man* waited downstairs.

(3) *Mike* called for a taxi, and *the agent* waited downstairs.

(4) *Mike* called for a taxi, and *Tom* waited downstairs.

The pronoun *he* used in Example (1) is the least specific anaphor, *the man* is somewhat more specific, and *the agent* and *Tom* are more specific still. In

sentences of this type, pronouns make the most effective anaphors precisely because they are nonspecific and are treated as "given information" (Chafe, 1974; Karmiloff-Smith, 1980). In other words, comprehenders are biased toward treating pronouns as old topics and try to map clauses containing them onto preexisting structures. In this way, pronouns help unify structures in memory (Lesgold, 1972).

On the other hand, using a more specific noun, as in Examples (2), (3), and (4), may prompt comprehenders to assume a new topic has been introduced, and the more specific the noun, the higher this probability. For example, some comprehenders will probably still think that the category noun *the man* in Example (2) refers to *Mike*, so these comprehenders will map those items onto the same person's structure. But in Example (3), where the more particular instance noun *the agent* appears, the likelihood that comprehenders will treat this word as an anaphor for *Mike* is greatly reduced. Finally, in Example (4) it should be clear that a new person, *Tom*, has been introduced, and comprehenders who do not shift to start a new structure will be in error.

In sum, for situations where only one potential referent has been mentioned, using a specific noun as an anaphor may actually suppress the activation of an intended referent by causing comprehenders to assume a new topic and shift. This is one reason why pronouns, which are the least specific of nouns, often make the most effective anaphors. In the same way, comprehension of common noun anaphors is generally faster when those nouns are less specific than their referents. For example, category nouns like *the bird* make more effective anaphors for instance nouns like *the robin* than vice versa (Garnham, 1981, 1984; Garrod & Sanford, 1977; Sanford & Garrod, 1980).

The experiments by Garnham, Garrod, and Sanford also illustrate the importance of activating previously unknown information from the topics we encounter. This processing task occurs very early on the route to complete comprehension. Indeed, in all likelihood, it is the first task to be initiated, occurring prior to anaphoric resolution, which itself precedes structure building. Our reasoning for this claim is that the amount of unfamiliar information contained in a sentence subject aids in determining whether that topic is old or new. If a subject contains information that is new to us, that subject can be perceived either as an elaboration of a previously mentioned topic (an anaphor that is more specific than its referent) or as a new topic altogether, depending on how much new information there is and whether that information can logically be attached to a previous referent. In short, activating the information from a sentence subject aids in determining whether that subject is old or new, which aids in the task of anaphoric resolution. Anaphoric resolution cannot proceed until the amount of new information in a potential anaphor has been assessed. As a result, the assumed order of processing tasks, the last of which may not be completed, is (a) activate subject information; (b) resolve the subject as a new noun or an anaphor; and (c) build an addition onto either the current mental structure or an old mental structure, or build the foundation for a new one, depending on how the potential anaphor is resolved. Performance of later tasks depends on the outcome of earlier tasks. Hence, the final goal of this comprehension chain, structure building, may not occur at all in some contexts.



The complete chain of events will be discussed in more detail later. First, how does this initial task of activating subject information influence processing time? Obviously, the more information a subject contains, the more time it should take to process. Hence, the information from richer, more detailed subjects takes longer to activate than the information from sketchy, generic subjects. The effect of an anaphor's information load on reading times has been studied extensively by Garnham (1981, 1984) and his colleagues (Garnham & Oakhill, 1985). Garnham suggested that the reason less specific nouns make more effective anaphors than more specific nouns do is because they contain less new information to be activated. For example, consider these test sentences:

(5) A *robin* would sometimes wander into the house.
The *bird* was attracted by the larder.

(6) A *bird* would sometimes wander into the house.
The *robin* was attracted by the larder.

In Example (5), a more general category noun, *bird*, is used as an anaphor for a more specific instance noun, *robin*. In Example (6), the instance noun is used as the anaphor, and the category noun is used as the referent. Comprehenders were faster reading the second sentence of a pair when it contained a category noun preceded by an instance noun, as in Example (5), than when it contained an instance noun preceded by a category noun, as in Example (6). In Example (5), we already know that a robin is a kind of bird, so the anaphor that begins the second sentence is redundant and takes little time to process. In Example (6), however, we do not know what kind of bird we are dealing with until the second sentence, so the anaphor *robin* provides us with new information that takes time to activate.

In sum, a more informative anaphor causes slower reading times than a less informative anaphor. Of the four noun types we have discussed (pronouns, category nouns, instance nouns, and names), which are the most informative? Certainly pronouns, which are the least specific, also contain the least amount of new information. They are redundant in almost every case, unless referring to a gender-ambiguous name such as *Chris*, in which case we can use the pronoun to determine the gender of the subject. Category nouns like *the man* are somewhat more informative and more specific. However, instance nouns like *the agent* are by far the most informative. An anaphor like *the agent* is not only quite specific, it also adds a lot more information to our topic's developing structure than an anaphor like *he* or *the man*, so it should take significantly longer to process. Then there are names, which, though highly specific, often contain less actual infor-

mation than instance nouns. We may have a stereotyped conception of what a *Fred* or *Mabel* is like, and when those names are used in a known context, we may even visualize particular persons, but names are merely labels and are not inherently meaningful in and of themselves. At the same time, names can be more informative than either pronouns or category nouns, or at the very least, they can be more salient. So, although specificity increases from pronouns to category nouns to instance nouns to names, information content is at its peak with instance nouns.

To summarize, for the initial task of activating topic information, more informative topics take longer to activate. If we have a second-clause anaphor that adds information to a name seen in the first clause, we can expect reading times to increase. Hence, we expect longer reading times for clauses that contain instance-noun anaphors, like *the agent,* than for any other type.

What about the effect of anaphoric resolution on processing time? As illustrated earlier, anaphoric specificity will affect resolution times differently depending on the context. An effective anaphor is one that clears up ambiguity. Sometimes this ambiguity is over which of several nouns is the intended referent (e.g., whether *Jane* or *Sue* is the intended referent for *she*). In such cases, more specific nouns make the most effective anaphors because they help us pick out the right referent (Gernsbacher, 1989). However, in cases where there is only one potential referent, the ambiguity is over whether a given noun refers back to that topic or to a new topic altogether. When a comprehender must decide whether a topic is old or new, highly specific nouns are less likely to be treated as anaphors. With a repeated or synonymous noun it is clear we have an old topic, and with a mutually exclusive noun it is clear we have a new topic, so decisions of old or new will be made quickly. However, in all other instances, we can expect less specific nouns (like pronouns) to be resolved faster than more specific nouns (like category and instance nouns).

Another way of viewing this relationship is to consider the "set size" or "specificity level" of a given anaphor. Pronouns are at the top of the hierarchy in terms of their set size and lack of specificity. In other words, a pronoun like *he* has a very large number of potential referents. Category nouns are one level more specific, so the number of potential referents for a category noun like *the man* is smaller. Instance nouns are more specific still, so an instance noun like *the agent* refers to a much smaller set of people. Finally, names are the most specific of all and as a result have the smallest number of potential referents. The bigger a noun's set size, the higher the probability that it will be used as an anaphor.

When the potential anaphor and its referent are equally specific, resolving the potential anaphor is fairly simple. If the two nouns in question are mutually exclusive, like *Mike* and *Bill* or *he* and *she,* a comprehender can quickly conclude that the nouns refer to different people. If the two nouns in question are synonyms or repeated nouns, a comprehender can quickly conclude that the nouns refer to the same person. A special case occurs when equally specific



nouns are neither mutually exclusive nor synonymous, as in *the teacher* and *the driver.* Though it is certainly true that many teachers drive, using these two nouns to refer to the same person is confusing, especially in the absence of a context that makes their relationship clear. In such cases, comprehenders seem biased to assume that nonsynonymous anaphors will be less specific than their referents. Anaphors are, after all, a form of shorthand in discourse. As a result, it would be highly unusual to use equally specific yet nonsynonymous nouns to refer to the same person, especially within the same sentence. What matters is that the second noun is no less specific than the first noun which significantly reduces the probability that it is intended as an anaphor.

In sum, we can expect slower reading times when potential anaphors are ambiguous and hard to resolve, and faster reading times when resolution is simplified (Matthews & Chodorow, 1988). Consider the sentences about Mike calling for a taxi [Examples (1)–(4)]. Resolution decisions will be made quickly in cases where it is highly probable that we have an anaphor referring to an old noun (*he* will almost always be perceived as an anaphor for *Mike*) or where we clearly have a new noun (*Mike* and *Tom* are obviously different people). However, resolution will be more difficult and more time-consuming when readers are unsure if the second-clause noun is meant to be an anaphor or a new subject, as occurs in Examples (2) and (3), where *the man* and *the agent* may be perceived as either anaphors or new nouns unrelated to Mike. In cases where category and instance nouns begin the second clause, how that noun is actually resolved will be biased by its context in the sentence. Considering this context takes additional processing.

Finally, according to the structure-building framework, we can expect to see longer reading times for sentences or clauses in which new topics are introduced. In the sentences about Mike, for example, readers should take longer to read the words *waited downstairs* whenever these words are associated with a new topic (someone other than Mike). We predict the fastest reading times for pronoun clauses like in Example (1), because readers will be mapping onto an old structure instead of creating a new one. Reading times should be slower for category-noun clauses like in Example (2), because many readers will shift to lay a new foundation. They should be slower still for instance-noun clauses like in Example (3), where the vast majority of readers will presumably shift. The structure-building framework suggests that the slowest reading times will be for clauses like that in Example (4), where the subject is clearly new and all readers should have to shift. The higher the probability of a shift, the slower the average reading time for a given clause across readers.

Note that these predictions refer to the separate effects of each process on reading comprehension times. Because one, two, or all three of these processes may affect comprehension time, depending on the nature of the experimental task, the impact of each separate task must somehow be incorporated into our overall predictions for clause reading times. How might this be done? One

straightforward method would be to sum the effect of each processing task that is believed to come into play during a particular comprehension activity.[1] By adding the assumed effects of each task, we can predict how long it should take to process each type of clause for complete comprehension. Recall that complete comprehension occurs only when all three subject-comprehension tasks are actually performed. If, for some reason, readers do not proceed to the final task of structure building, or even to the transitional task of anaphoric resolution, then only partial comprehension is demonstrated.

Of course, for comprehension activities that require only partial comprehension, only those tasks that are actually performed will affect reading times. In other words, if just the first task of information activation is completed, our predictions for clause reading times should look like the predictions for information activation alone. If readers make it past this first step to perform anaphor resolution as well (yet stop short of building a structure), we must sum the effects of the first two tasks to make our predictions for reading times. Finally, if readers are required to form a mental representation of the information contained in a sentence, they will proceed to the very end of the complete comprehension chain, and their reading times should look like a sum of all three patterns combined. In short, reading-time data will depend on how far along the subject-comprehension chain a reader must proceed in order to satisfy experimental demands.

What we are suggesting, in other words, is that experimental participants are often unenthused by the experimental texts given them and will take comprehension only as far as they are motivated or required. If the demands of a given comprehension task do not include building a structure, they may not bother to build one. If it is not necessary to resolve a given anaphor to complete the experiment, the anaphor may go unresolved. This is hardly a new suggestion in psychological literature. Indeed, in some respects this position sounds like the "minimalist" perspective put forth by McKoon and Ratcliff (1992), but it is actually more in line with what may be called the "satisficing subject hypothesis" preferred by Zwaan and Graesser (1993a, 1993b).

Concurring with Zwaan and Graesser, we suggest that all readers—even experimental subjects—have strategies and goals when they read. Readers are motivated to process texts for a variety of reasons. In most situations, these motivations probably lead to structure building and "complete comprehension." However, the reading done by subjects in experimental situations is hardly typical. Unlike most readers outside the lab, experimental participants have little

[1]Summing the effects of each separate subject-comprehension task suggests that these tasks are performed serially, rather than in parallel. Serial processing of subject-based information is a reasonable assumption, given that tasks later in the comprehension chain are dependent on the outcomes of earlier tasks and cannot be initiated until earlier tasks are nearing completion.



motivation to engage in complete comprehension of the texts given to them. But that does not mean that these subjects go into experimental reading tasks with no strategies whatsoever. Rather, experimental subjects can be assumed to have what Zwaan and Graesser (1993a, 1993b) have called a "satisficing" strategy: Readers converge on strategies that are good enough to enable them to satisfactorily complete experimental tasks, but that do not entail any unnecessary processing. In other words, a reader's strategy is to do only what is required and no more. This means that if we want our subjects to perform a particular processing task, we must sufficiently motivate them to do so. Subjects are quick to discover what is required in an experimental task, and they adjust the amount of their processing accordingly in order to achieve maximal results with minimal effort.

In the three experiments presented here, the stimuli were essentially the same: two-clause conjunctive sentences, like the *Mike* sentences seen earlier, in which the subject of the second clause was a potential anaphor for the subject of the first clause. This potential anaphor was either a pronoun, category noun, instance noun, or new name. The following sentences demonstrate these four conditions. One of these sentences appeared in each of four test versions.

Stephanie watched the kids in the park, and she ate lunch. (pronoun)
Stephanie watched the kids in the park, and the woman ate lunch. (category noun)
Stephanie watched the kids in the park, and the teacher ate lunch. (instance noun)
Stephanie watched the kids in the park, and Jill ate lunch. (new name)

The task demands for the experiments varied. In the first experiment, we simply told experimental participants to read the sentences and pay attention to the information each contained. For the satisficing participant, these instructions should have prompted information activation and nothing more. In the second experiment, experimental participants had to answer a subject-focused question immediately after each sentence (e.g., "Did Stephanie eat lunch?"). Answering correctly required not only activating subject information, but resolving the subject that headed the second clause as well. Finally, in the third experiment, we asked readers to rewrite each sentence in their own words shortly after reading it. Because the readers knew ahead of time that they would have to reconstruct the information with which they were presented, we expected them to process the sentences with structure building in mind. This was the only experiment in which we expected our satisficing readers to reach the end of the complete comprehension chain.

One purpose of these experiments was to test the satisficing subject hypothesis that experimental participants often do only what is required of them and no more. If this hypothesis is valid, if our underlying assumptions about the three comprehension tasks and their separate effects on reading time are essentially correct, and if our "pattern-summing" procedure is an appropriate way to account

for each task's effect, then our predicted reading-time patterns should conform to the observed results in each of the three experiments. This finding in itself would justify the tedious comprehension tasks our undergraduate subjects must begrudgingly perform.

## EXPERIMENT 1

In the first experiment, our objective was to see if varying the information content of the second-clause subject influenced how quickly comprehenders read the words that followed it. We believed that if comprehenders were reading at the most basic level and weren't expecting to be tested on the sentences they read, the only thing that would affect their reading times for the second clause would be how much information they had to activate from the subject that appeared in it. Hence, words appearing after the second-clause subject should be read most slowly in the instance-noun condition where the subject is highly informative. They should be read most quickly in the pronoun and category-noun conditions where the information provided is redundant. For the name condition, reading times may be somewhat elevated if comprehenders have a particular person or stereotype in mind after reading the second-clause name. Note that we measured reading times for every word that appeared after the second-clause subject (i.e., the predicate of the clause), but we did not record the reading time for the subject itself. This was to avoid the variations in reading time caused by the varied length and frequency of the four second-clause subject types. Pronouns may be read the most quickly, not only because they are the least informative, but because they have the fewest letters and the greatest printed word frequency of the four noun types.

Recall for actual words and main ideas seen in the second clause of the sentences was tested after all sentences had been read. Experimental participants were cued with the first clause of a sentence they had read and had to write the second clause of that sentence to the best of their recollection. Prior to the recall portion of the experiment, experimental participants were not told that their recall of sentences would be tested. Hence, their clause reading times should not have been affected by anticipation of recall.

### Method

***Subjects.*** Eighty-four undergraduates at the University of Wisconsin–Madison participated in this experiment to receive extra credit for an introductory psychology course. All participants were native English speakers.

***Materials.*** The materials were 48 two-clause test sentences of the type mentioned earlier. The first clause of each sentence had a named subject performing some action. The second clause of each sentence had a subject of the same gender performing another action in the same setting. The two clauses were joined by the conjunction *and,* and they were written in such a way that the actions in the two clauses could reasonably have been performed by either the same or different people.

In each version, a sentence had a different noun type as the subject of the second clause: This second subject appeared as either a pronoun, a category noun, an instance noun, or a new (not-yet-seen) name. Category-noun and instance-noun subjects were always modified by the definite article (*the* man, *the* agent). The action in the second clause usually consisted of a verb and a direct object, or a verb modified by an adverb or short prepositional phrase. The number of words appearing after the second-clause subject varied from two words to four, but was most often three words. The 48 test sentences appeared in the same order in each of the four versions. Readers were randomly assigned to one of four test versions, but each version contained sentences in all four conditions. Hence, this experiment, like all of those reported here, was a repeated-measures design, with noun condition as the within-subject factor.[2]

For the delayed-recall task, 36 of the 48 sentences were selected at random (9 sentences from each condition). The first clause of each of these sentences was used to cue recall of the second clause. For example, if experimental participants had read the sentence "Seth vacuumed the carpet, and Victor did the wash," they would be cued with the clause "Seth vacuumed the carpet. . . ." They were then expected to write, "and Victor did the wash." First-clause cues appeared in the same order as their corresponding sentences had appeared during the reading portion.

***Apparatus.*** The experiment was run on an Apple II computer system. Each terminal had a two-key response board. By pressing either of the keys, participants could advance through the text they were supposed to read.

***Procedure.*** Participants were tested in groups of 1 to 3, seated at separate terminals in a quiet room. Test stimuli were presented using a "moving-window" technique. When participants were first seated at a terminal, they saw a screen filled with dashes with each cluster of dashes representing a word in a paragraph of instructions. Participants were told that by repeatedly pressing one of the keys on the response board in front of them, they could transfer the dashes into letters and thus advance through the text. Each key press turned the next cluster of dashes into a word. Pressing the key again would cause this word to revert back to dashes and transform the next set of dashes into a word. In this way, the reading time for every word in a sentence could be measured. Participants were instructed to "try to read along at your normal pace, but pay careful attention to

[2]Version served as a between-subjects control factor in all three experiments and never demonstrated a significant effect.

what the sentence is saying. Once you press the key to move on to the next word, you will not be able to go back to the words you read earlier, so it is important to understand the words you just read before you move on." They were not warned about the cued-recall test afterwards. Presenting the experimental instructions with the moving-window technique gave participants a chance to practice reading text in this fashion before the experimental stimuli were presented.

After reading the instructions, participants read four practice sentences of the type to be seen in the experiment. Each sentence consisted of a line of dashes that appeared center-screen with each cluster of dashes representing a word in a two-clause sentence. As with the instructions, sentences were read using consecutive key presses. There was a 2.5-s pause after the last key press of each sentence before the next line of dashes appeared on screen.

Once participants finished reading the 48 test sentences, a paragraph of instructions appeared on the screen directing them to recall the sentences they had just read. Participants were handed a sheet of lined paper with numbered spaces for 36 sentences. Each key press caused the first half of one of the sentences to appear onscreen. They were supposed to write the second half of that sentence in the numbered space provided, word-for-word if possible, or using any ideas or phrases that came to mind. The first-clause cue remained on the screen until participants finished writing their recollections and pressed the key again. The recall portion of the experiment was entirely self-paced, but participants could not return to earlier sentences or change any of their answers once they had moved on to the next first-clause cue. The experimental session lasted about 30 min.

### Results and Conclusions

First, we analyzed the participants' reading times for words that appeared after the second-clause subject. These individual word reading times were then summed to give a predicate-clause reading time. The more information a clause's subject contained, the longer it should have taken to process. Because instance nouns are far more informative than any other noun type, we expected to see significantly elevated reading times for the instance-noun clauses. That is precisely what our data show. The mean reading time for instance-noun predicates was 1,167 ms ($SE$ = 61 ms), compared to a mean of 1,082 ms ($SE$ = 55 ms) for new-name predicates, 1,051 ms ($SE$ = 56 ms) for category-nouns predicates, and 1,044 ms ($SE$ = 54 ms) for pronoun predicates. The differences in these predicate reading times were significant, $F_1(3, 240) = 12.24$, $p < .0001$, $F_2(3, 138) = 3.58$, $p < .016$, min $F'(3, 220) = 2.77$, $p < .05$. As shown by post-hoc Tukey tests performed with subject random and then with items random, readers were significantly slower reading instance-noun clauses than any other type, with $p < .05$ for both types of analyses. The reading times for new-name clauses were also somewhat elevated, but post-hoc Tukey tests showed that neither this elevation nor any of the remaining differences in reading times were statistically significant.



The delayed cued-recall data were scored in the following way: Participants were given points for any bits of information *after* the second-clause subject that they recalled accurately, with 4 being the highest number of points awarded per clause.[3] For words occurring after the second-clause subject, readers were given 2 points for a verbatim recall of the clause's verb and 1 point for a synonym. Similarly, subjects were given 2 points for a verbatim recall of the object of the clause and 1 point for recall of a synonym. Recall sheets were scored by two independent raters who agreed on 95% of the clauses they scored. In instances where they differed, it was only by 1 point and the average of the two raters' scores was then used.

Because readers were not expecting the recall task, overall recall was quite low—less than 11% of all the information presented in the second clauses. However, readers were significantly more successful at recalling information from pronoun clauses (14%) than from any other type of clause (category = 9%, instance = 10%, new name = 10%). As expected, the amount of information recalled depended on the type of noun that headed the second clause, $F_1(3, 240) = 8.32$, $p < .0001$, $F_2(3, 105) = 7.52$, $p < .0001$, min $F'(3, 280) = 3.95$, $p < .05$. Post-hoc Tukey tests with subjects random and then items random showed that, although pronouns caused significantly higher recall than any other noun type, the remaining three conditions did not significantly differ from one another. This pattern of results conforms to our expectations about the unifying effect of pronouns. Far more than any other noun type, pronouns are perceived as cues to map clauses together.

## EXPERIMENT 2

In this experiment, we prompted readers to move on to the second process in the complete comprehension chain and to resolve the potential anaphors that appeared. This step must be completed before the structure-building decision to map or shift can be made. Will experimental participants also initiate the third and final step of structure building if they can complete the experiment without it? Probably not, if the satisficing position is correct. The satisficing subject hypothesis suggests that laboratory subjects seldom perform unnecessary comprehension tasks. They will do what they must to properly complete the experiment, but cannot be counted on to do more.

In the first experiment, all we instructed experimental participants to do was read the sentences and "pay attention to" the information they contained. Recall

[3]We opted not to give experimental participants points for recalling the subject of the second clause because clauses that started with a pronoun would have been given an inherent advantage. Participants often began their recall with a gender-appropriate pronoun whether this pronoun had appeared in the original sentence or not. Hence, sentences in the pronoun condition would have ended up with scores 2 points higher than in any of the other conditions, even in cases where readers hadn't actually remembered the subject of the second clause.

data and reading-time data suggest readers generally did not go beyond this initial process of information activation. Recall, in most cases, was very low, and reading-time patterns looked like those expected for information activation alone.

In the second experiment, we hoped readers would take comprehension one step further. To prompt this, we asked a comprehension question after each sentence read, a question that required them to resolve the potential anaphor in the second clause. The test sentences were of the same form seen in Experiment 1: two-clause, past-tense conjunctive sentences in which there was a subject performing an action in each clause. The question probed whether or not readers thought the subject of the first clause and the subject of the second clause were the same person. In other words, was the second-clause subject an anaphor referring back to the first-clause subject, or was it a new subject altogether? This question was always of the form: Did Subject A perform Action B? For example, if the test sentence was, "Susan opened the window, and the woman removed her coat," the question was, "Did Susan remove her coat?" If readers answered "yes," to this question, it suggested they thought Subject A (Susan) and Subject B (the woman) were the same people. If they answered "no," it suggested they thought two different people were being discussed.

Again, one purpose of this experiment was to see if readers bother to build a structure when it is not required of them. If the satisficing position is correct, the subjects should try to resolve the anaphor because answering the question correctly requires it, but they should not necessarily build a structure because complete comprehension is unnecessary in this experimental situation. All that readers need to "figure out" is whether the second-clause subject seems like a possible or probable anaphor for the first-clause subject. Hence, reading-time patterns for the words after the second-clause subject should look like the patterns for information activation and anaphor resolution summed. If this is the case, the slowest reading times should be for instance-noun clauses, since instance nouns are both more informative and more ambiguous to resolve as old topic or new. The second slowest reading times should be for category-noun clauses. The quickest reading times should be seen for the pronoun and new-name clauses, because both noun types are low in information content and easily resolvable (pronouns are quickly resolved as old topics, and new names are quickly resolved as new topics).

## Method

***Subjects.*** A total of 161 undergraduates from the University of Oregon participated in this experiment as one means of fulfilling an introductory psychology course requirement. All participants were native English speakers.

***Apparatus.*** The experiment was run on an Apple II computer system. Each terminal had a two-key response board which readers used to call up sentences, proceed through the text, and answer questions. The left key was marked *yes;* the right key was marked *no*. The system was designed so that 4 readers seated at separate terminals could participate in each session, with all participants working at their own pace.



***Procedure.*** Participants were tested 4 at a time in separate cubicles in a quiet room. As in Experiment 1, test sentences and instructions were presented with a moving-window technique, and readers were given practice sentences before the actual experiment (one from each of seven experimental conditions). After participants read a sentence and pressed the key one last time, there was a 2.5-s pause before the test sentence disappeared and the comprehension question appeared. The question appeared 2 in. below where the sentence had appeared and was in regular print, not in dashes. Experimental participants responded to the question by pressing the *yes* or *no* key. Each response was followed by another 2.5-s pause before the next series of dashes appeared.

After a seven-sentence practice session, each participant filled out an informed consent sheet and then pressed a response key to begin the experiment. They were given three "warm-up" trials before proceeding to the 70 experimental stimuli. These three trials were not included in the results. As participants read the test sentences, the computer recorded the reading time for each word after the second-clause subject, as in Experiment 1. Responses for each comprehension question and the time it took to answer it were also recorded. The entire experimental session lasted about 20 min.[4]

## Results and Conclusions

There were three parts to the analysis for this experiment. The first part examined readers' responses to the yes/no question to see whether they treated the second-clause subject as an anaphor or as a new subject. The second part examined the reading times for second-clause words to see how reading speed was jointly affected by the assumed tasks of this experiment: information activation and anaphoric resolution. The third part examined the time to answer the yes/no question to see if the noun type of the second-clause subject affected comprehension of the question as well.

For each of the noun-type conditions, the percentage of questions to which readers responded "yes" (indicating they had considered the second-clause sub-

[4]Participants were given 70 test sentences (rather than 48), and there were 7 noun conditions (rather than 4). As in Experiment 1, the second-clause subject could be a pronoun, category noun, instance noun, or new name, and the noun type that appeared in a given sentence was different in each version of the test. However, in Experiment 2, the subject of the first clause also varied. For our purposes here, we will only consider the sentences that were of the form seen in Experiment 1: those with a name as the subject of the first clause and one of the four noun types as the subject of the second clause. We will disregard the results from test sentences where the first subject was a pronoun, category noun, or instance noun and the second subject was a name, although such sentences did appear amidst the stimuli of interest.

ject as an anaphor) was calculated. Whereas 90% of the pronouns were used as anaphors, less than 12% of the other noun types were used this way (category nouns = 11%, instance nouns = 10%, new names = 6%). A repeated-measures analysis of variance (ANOVA) indicated a significant condition effect for the percentage used as anaphors, $F(3, 462) = 1611.95$, $p < .0001$. This pattern of data confirms the expectation that pronouns would almost always be viewed as anaphors because of their prototypicality and lack of unnecessary specificity. The data also coincide with our prediction that increasingly more specific nouns are less likely to be resolved as anaphors. A post-hoc Tukey comparison showed that the mean for pronouns was significantly greater than all other means, and the mean for new names was significantly smaller than all other means, $p < .05$, but the category-noun and instance-noun means did not differ significantly from one another.

As in the other experiments reported here, we analyzed the reading times for the predicate of each sentence's second clause. The means for the predicates were 1,149 ms ($SE$ = 26 ms) for pronoun, 1,323 ms ($SE$ = 33 ms) for category noun, 1,419 ms ($SE$ = 32 ms) for instance noun, and 1,179 ms ($SE$ = 26 ms) for new name. Repeated-measures ANOVAs showed a significant main effect for condition, $F_1(3, 462) = 33.08$, $p < .0001$, $F_2(3, 207) = 36.25$, $p < .0001$, min $F'(3, 591) = 17.30$, $p < .05$. Post-hoc Tukey tests showed that, as predicted, predicate reading times were significantly faster following a pronoun or a new name than following a category noun or an instance noun, with $p < .0001$ for both subject and item analyses. The reading times following pronouns did not significantly differ from those following new names, but the reading times following category nouns did differ from those following instance nouns, $p < .002$.

Using the predicate-clause reading times, we examined whether activating subject information and resolving the potential anaphor jointly affected the speed at which clauses were read. As in Experiment 1, the high information content of instance nouns was expected to slow down reading times for instance-noun clauses relative to other types of clauses. Now, consider the additional effect of trying to resolve the second-clause subject as an anaphor. Recall that we believed this resolution would proceed quickly in cases where the subject was clearly old (a pronoun) or clearly new (a new name), but that resolution would take longer in cases where it was unclear whether the subject was an anaphor or a new subject (as is the case with "middle-ground" category nouns and instance nouns). Hence, we expected the task of anaphoric resolution to cause faster reading times in the pronoun and new-name conditions and slower reading times in the category-noun and instance-noun conditions. When the expected information-activation effect is combined with the expected anaphoric-resolution effect, the outcome is a graph that looks remarkably like our data for Experiment 2. This pattern of results looks more like the expected pattern for information activation and anaphor resolution alone than like the pattern for all three processes summed (which is discussed in Experiment 3). The pattern suggests that, although the second-clause subjects were resolved in this experiment, new structures for nonanaphors were not typically built.



Finally, the mean time to answer the yes/no question for each condition was calculated. Questions about sentences containing pronouns, which almost always elicited a *yes* answer, were answered more rapidly than any other type ($M$ = 1,500 ms, $SE$ = 35 ms), whereas question-answering times for the sentences containing category nouns ($M$ = 1,773 ms, $SE$ = 40 ms), instance nouns ($M$ = 1,827 ms, $SE$ = 41 ms), and new names ($M$ = 1,839 ms, $SE$ = 38 ms), which typically elicited a *no* answer, were about the same. An ANOVA for these means showed a significant condition effect, $F(3, 462) = 37.42$, $p < .0001$. Post hoc Tukeys revealed that the pronoun-related questions were answered significantly faster than any other type $p < .0001$, but the other question-answering times did not significantly differ from one another. In sum, readers were about 300 ms faster responding to questions about pronoun clauses than to any other type. This result corroborates Lesgold's (1972) finding that readers are better at recalling the words in a sentence when the second clause contains a "unifying" pronoun. However, the faster responses to questions about pronoun clauses could also be due to the fact that such responses were always affirmative as opposed to disconfirmative.

## EXPERIMENT 3

In Experiment 1, participants were not warned about the delayed-recall test and were told only "to pay careful attention to what the sentence is saying." Because participants were not given any indication that anaphor resolution and structure building would be necessary in order to complete the experiment, we expected them to do only a minimal processing of the sentences they read. As anticipated, only the information content of the second-clause subject seemed to affect reading times across the four noun conditions.

In Experiment 2, we prodded the participants to move one link further down the chain of complete comprehension. In this second experiment, experimental participants were required to resolve the potential anaphor that headed the second clause. For this experiment, we expected both subject informativeness and difficulties in anaphor resolution to have an effect on clause reading times, and the data suggest that this expectation was fulfilled.

In Experiment 3, we believed participants would reach the third and final link in the complete comprehension chain. In other words, we expected them to build mental structures that would enable them to "truly comprehend" each sentence they read because we required them to rewrite those sentences in their own words. For Experiment 3, we assumed all three processing tasks (information activation, anaphor resolution, and structure building) would affect reading times. Hence, the pattern of results should look like the patterns of all three processing tasks summed. This would again lead to the slowest reading times for

the instance-noun clauses and the fastest reading times for the pronoun clauses. The reading times for the new-name and category-noun clauses should fall somewhere between these two extremes, but should be essentially equivalent to one another.

Experiment 3 used the same test sentences as Experiment 1: 48 two-clause conjunctive sentences in which the type of noun used as the second-clause subject in a given sentence varied across the four versions. Although the stimuli were exactly the same as in Experiment 1, what we expected readers to do with them was more elaborate. In this experiment, we instructed readers to rewrite each sentence in their own words shortly after reading it. They were given a fixed amount of time (25 s) in which to do this. This procedure, unlike those in the previous two experiments, should prompt participants to build mental structures for the sentences they read. These structures may then be used as a guide for rewriting the sentences in different words.

## Method

***Subjects.*** Eighty-four undergraduates at the University of Wisconsin–Madison participated in this experiment to get credit for an introductory psychology course. Of the 84 students who completed the experiment, the data for 12 were discarded because they failed to follow instructions for a large number of experimental trials. If participants rewrote more than 50% of the test sentences verbatim instead of in their own words, or if on more than 25% of the trials they started to rewrite the sentence before they had pressed the key to remove the last word (thus contaminating their reading-time data for the last word), those participants' data were excluded from the results. As in the other experiments reported here, all participants were native English speakers.

***Materials.*** The 48 test sentences were the same ones used in Experiment 1. The computer systems, response key set-up, and experimental setting were also identical to those in Experiment 1.

***Procedure.*** As in the previous experiments, the instructions and stimuli were presented with a moving-window technique that transformed clusters of dashes into words. In this experiment, however, readers were also asked to rewrite the sentence they had just read. Experimental participants were given four practice trials in order to familiarize themselves with this procedure. During these practice trials (and during the experimental trials that followed), participants read through a line of dashes by repeatedly pressing a key, as described earlier. Three seconds after the key press that indicated the last word of the sentence had been read, the words "please rewrite the last sentence" appeared where the sentence had been. Participants then rewrote the sentence in their own words on numbered lines provided for that purpose.



Participants were given 25 s in which to rewrite the sentence before the next series of dashes appeared onscreen. However, because reading through the next sentence did not begin until the participant pressed a key, they had as much time as needed to finish writing. Most rewrites were completed well before the next series of dashes appeared. The experimental sessions for Experiment 3 lasted about 40 min.

## Results and Conclusions

In Experiment 3, participants read sentences knowing that they would have to rewrite them in their own words soon afterwards. This forced them to comprehend the situation depicted in the sentence more fully than in the previous experiments. Hence, it was the only experiment in our series for which participants were expected to build mental representations of the sentences they read, and the only one in which the slow-down associated with shifting to a new topic should have an effect. We predicted that predicate reading times for this experiment would look like the patterns of all three subject-comprehension tasks summed. In other words, if readers must engage in all three comprehension tasks in order to build a mental representation of the sentence they are rewriting, and if the effects of each of these tasks on reading times is summed, the pronoun condition should still show the fastest reading times and the instance-noun condition should show the slowest reading times. Reading times for the category-noun and new-name conditions should fall somewhere between the other two: significantly slower than the pronoun condition, but significantly faster than the instance-noun condition. This is precisely the pattern that resulted. The pronoun condition was the fastest ($M$ = 1,657 ms, $SE$ = 109 ms); the category noun, second fastest ($M$ = 1,817 ms, $SE$ = 114 ms); the new name, third fastest ($M$ = 1,905 ms, $SE$ = 120 ms); and the instance-noun condition, slowest of all ($M$ = 2,075 ms, $SE$ = 143 ms). These predicate reading times were submitted to repeated-measures ANOVAs, which showed a significant effect of noun type, $F_1(3, 204) = 11.72, p < .0001, F_2(3, 141) = 4.98, p < .003$, min $F'(3, 254) = 3.49, p < .05$. Post-hoc Tukey tests for both subject and item analyses showed that, as expected, the pronoun condition was significantly faster than all others, whereas the instance-noun condition was significantly slower, $p < .05$. The category-noun and new-name conditions did not differ significantly.

Examining the sentence rewrites should be of help in determining which of the second-clause subjects were perceived as anaphors. We expected this analysis to replicate results from the yes/no question in Experiment 2 which suggested that the vast majority of pronouns were perceived as anaphors (90%), yet few category nouns (11%), instance nouns (9%), or new names (6%) were perceived as such.

Independent raters examined the rewrites and recorded whether they thought one actor or two was indicated. We expected that in cases where readers thought the actor in the first clause and the actor in the second clause were the same

person, they would use a null subject or a pronoun to refer to the subject of the second clause. In cases where they thought the sentence contained two separate actors, they were expected to keep these actors separate in their rewrites, referring to each in the sentence's original words. For example, if the original sentence was: "Mike called the taxi, and the man waited downstairs," and the reader used both "Mike" and "the man" in his rewrite, a rater would score this rewrite as a 2, indicating two actors. If the rewrite used the word *he* or a null subject to refer to *the man,* ("Mike called the cab and waited for it downstairs"), a rater would score this rewrite as 1, indicating one actor.[5] A rewrite that contained only one actor suggested that the second-clause subject was perceived as an anaphor for the first-clause subject. Raters agreed on more than 99% of their classifications, and where they did not, an average rating was used.

Sentence rewrites suggested that 100% of pronoun subjects were perceived as anaphors. Category-noun subjects were perceived as anaphors 25% of the time, whereas instance-noun subjects were perceived as anaphors only 18% of the time. New-name subjects were perceived as anaphors less than 3% of the time. Repeated-measures ANOVAs on the total number of anaphors per condition showed a strong effect of noun type, $F_1(3, 204) = 383.92$, $p < .0001$, $F_2(3, 141) = 1994.22$, $p < .0001$, min $F'(3, 275) = 321.94$, $p < .0001$. Post-hoc Tukey tests showed that pronouns were used as anaphors far more than any of the other noun types, $p < .0001$. Category nouns and new names also differed significantly in the number of times they were used as anaphors, $p < .0001$. No other differences between conditions were significant.

## GENERAL DISCUSSION

One purpose of this study was to replicate past results in anaphoric resolution which suggest that pronouns are the most effective type of anaphor precisely because they are nonspecific and suggest given information (Chafe, 1974). Anaphors must be specific enough to enhance their referents and suppress all nonreferents (Gernsbacher, 1989), but there is such a thing as being *too* specific. When only one apparent referent has been mentioned, using a highly specific anaphor will actually bias comprehenders into thinking a new topic is being introduced. After all, why use a highly specific noun to refer back to an old topic when a short and simple pronoun will do the trick?

The sentences used in this study were constructed so that only one potential referent had been mentioned for each possible anaphor that occurred. The specificity of the possible anaphor in the second clause varied. Ranging from the least specific to the most specific, the second-clause subject was either a pronoun, a

[5]Sentences which originally contained pronouns would be scored as containing only one actor even if the reader was just rewriting the subjects exactly as seen. This is why 100% of rewrites in the pronoun condition were scored as 1.



category noun, an instance noun, or a new name. As expected, comprehenders treated pronouns like anaphors in the vast majority of cases (90% or greater). In general, the more specific the noun, the lower the probability that it would be resolved as an anaphor. In Experiments 2 and 3, pronouns were almost always treated as anaphors, category nouns ranked second in this regard, instance nouns ranked third, and new names—which couldn't logically be used as anaphors for a name seen in the first clause—ranked last. In short, these experiments demonstrated that highly specific nonrepeated nouns will usually be resolved not as anaphors, but as new subjects. That is why pronouns, which are the least specific of nouns, often make the best anaphors.

Experiment 1 also showed that pronouns are significantly better at unifying related clauses in memory. Because pronouns are resolved as anaphors in the vast majority of cases, whereas other noun types are not, pronouns will lead comprehenders to attach the information from a pronoun clause to the pronoun's referent. Accordingly, participants in Experiment 1 had the easiest time associating the first and second clauses of the sentences they read when the second clause was headed by a pronoun. This gave pronouns a significant advantage in facilitating recall of second clauses when the first clause of a sentence was presented in the delayed-recall task. The three other noun types that headed the second clause did not significantly differ in their facilitation of recall.

Finally, as all three experiments demonstrated, clauses that followed a pronoun were read faster than any other type. Considering that all three comprehension processes favored pronouns as far as reading time was concerned, it isn't any wonder that pronouns showed this decided advantage. For the first comprehension task of information activation, pronouns are the least informative of the four types of nouns and hence the easiest to activate, whereas instance nouns are the most informative and hardest to activate. Accordingly, in Experiment 1, where participants thought they merely had to read the sentences, pronoun clauses were read the fastest and instance-noun clauses were read the slowest.

For the second comprehension task of anaphor resolution, pronouns are easiest to resolve as anaphors, because experienced readers know that pronouns almost always have referents. New names, on the other hand, are the most specific and the easiest to resolve as new topics, because experienced readers know that a sudden switch of personal names usually means the person being discussed has also changed. As a result, in Experiment 2, where participants were prompted to resolve subjects as either anaphors or new nouns, we expected to see the fastest reading times in the pronoun and new-noun conditions. This is precisely what occurred. Readers were fastest at the extremes of the specificity spectrum (pronouns and new nouns) where their resolution tasks were simple, and they were slowest toward the middle of the specificity spectrum (category nouns and instance nouns) where it was ambiguous whether a new noun had been introduced or not.

The third and final comprehension task—structure building—was prompted by the instructions to "rewrite the sentence in your own words" in Experiment 3. The structure-building framework proposed by Gernsbacher (1990) suggests that it takes more time to lay down the foundation for a new topic than it does to add information to an old topic's structure. Again, this would mean pronouns, which are almost always treated as old subjects, would have a decided advantage in reading times. If structure building could be considered in isolation, we would expect the fastest reading times to occur in clauses that followed pronouns, the second fastest in category-noun clauses, the third fastest in instance-noun clauses, and the slowest in new-name clauses, since the probability of having to shift to a new topic increases as the second-clause subject becomes more specific. However, because the other two comprehension tasks that precede structure building are also affecting reading times, the pattern expected should look like a sum of all three tasks' patterns. This still leads to an advantage for pronouns, but it means that instance-noun clauses, not new-name clauses, should show the slowest reading times. The reading-time data from Experiment 3 confirm this.

These experiments lend support to our theory that complete comprehension of subject-based information is made up of three separate yet interdependent processing tasks. These tasks occur in a fixed order: (a) activating information contained in the subject; (b) resolving the subject as either a new topic or an anaphor referring back to an old topic; and (c) building a new structure to accommodate subject-related information if the topic is new, or attaching this same information to a previous structure if the topic is old. This order is necessary because each task hinges upon the outcomes of those that precede it. Table 1 summarizes the effects of these three tasks on reading times. Our expectations about the different conditions' reading times were fulfilled in each of the three experiments. In addition, the mean time to read the second clause increased across the three experiments, as was expected if additional processing tasks were being prompted with each new experiment.

We have defined *true comprehension* as the ability to accurately redescribe a

**TABLE 1**
**A Comparison of Reading Times for Experiments 1, 2, and 3**

| Task | Pronoun | Category | Instance | New Name | *M* |
|---|---|---|---|---|---|
| Experiment 1 | 1,044 | 1,051 | 1,167 | 1,082 | 1,086 |
| Experiment 2 | 1,149 | 1,323 | 1,419 | 1,179 | 1,268 |
| Experiment 3 | 1,657 | 1,817 | 2,076 | 1,905 | 1,864 |

*Note.* The first experiment involved only information activation, the second experiment added the process of anaphoric resolution, and the third experiment added the process of structure building to the other two.



situation, action, or concept in one's own words. Other types of knowledge can be gained and retained short of this type of comprehension, but such knowledge is usually less lasting and less useful than understanding the material well enough to redescribe and reapply it. For example, we all know that students can do well on certain types of knowledge tests without really comprehending the material that is covered. We also know that once the semester is over, these students will forget most of the "knowledge" they have supposedly gained. But what do they really learn when their goal is to do well on the typical multiple-choice or fill-in-the-blank exam? Simple recognition or regurgitation of memorized facts may demonstrate that students have temporarily memorized the material in their notes and textbooks or can recognize it if they see it, but do they really understand it? Are they merely repeating the concepts to which they have been exposed, or do they comprehend the concepts well enough that they could actually re-explain them, fitting the facts they have learned into new contexts? Being able to re-explain ideas and integrate them with past and future knowledge requires something more than rote memorization of facts. It requires building a mental structure that links the new concepts to things we already know and that organizes the information in a clear and tractable fashion. Such structures take time and effort to build, and—as with buildings in the physical world—the more time and thought we spend on a structure, the longer it is likely to last.

Because mental structures take time and effort to construct, a comprehender must be sufficiently motivated to build them. The question is, in what comprehension situation *is* motivation sufficient? Readers in real-world contexts probably build structures for much of the material they read, because readers are often motivated to try to enjoy, remember, or learn something new from the various texts that they encounter. However, it would be naive for us to assume that the majority of experimental participants are similarly motivated. It is more prudent to assume that most experimental participants are satisficers—that they will do only what is required of them in a given situation and no more. This is especially the case in the typical reading experiment, where undergraduate participants perform often-tedious comprehension tasks for small amounts of money or course credit. How well participants perform in these experiments rarely influences the amount of compensation they receive, so participants are poorly motivated to do unnecessary processing of experimental materials. If the experimental task actually requires that they build a structure in order to answer questions, they will do so. Otherwise, they probably will not. Similarly, participants can only be expected to resolve the anaphors they encounter if resolution is actually required for comprehension of experimental materials. If the subjects can get through the experiment just fine without resolving the anaphors, why bother doing it?

If believing that experimental participants are "lazy" seems uncharitable, consider it from this perspective: Even if someone is motivated to do well in an experiment, it is expected that he or she will *not* engage in processing tasks that are unrelated to—or even inhibit—that person's performance on the experimen-

tal task. Hence, if we want experimental participants to engage in a given process, we must make that process integral to their performance in the experiment. As we have already mentioned, structure building takes time and cognitive effort to perform, effort which, in some cases, may actually detract from a subject's performance. If subjects in Experiment 2 can answer the anaphor-based question by merely resolving the anaphor, proceeding to build a structure is a waste of their time. Efficiency-minded subjects should pick up on this very quickly. Call it lazy or call it efficient—the human mind doesn't believe in wasted effort.

This same effect was demonstrated in a recent study by Wilson, Rinck, McNamara, Bower, and Morrow (1993). In a series of four experiments, it was shown that readers would not construct highly detailed spatial representations of a narrative situation unless those spatial representations could be used to improve the readers' performance on the experimental task. Subjects studied detailed spatial layouts of a research center until they could reproduce the layout from memory. They then read narratives in which a protagonist moved through this layout while performing various activities. At certain points during the narrative, subjects were primed with the names of objects located somewhere in the memorized layout. In the first two experiments, subjects had to respond whether a probe object and a target object were located in the same or different rooms. In the latter two experiments, subjects instead had to indicate whether the protagonist and the target object were in the same room at a given point in the narrative. Response times from all four experiments indicated that subjects did not bother to access their memorized spatial representations unless they were forced by task demands to follow the protagonist through the learned layout. In Experiments 1 and 2, it was not necessary to follow the protagonist through the narrative to answer whether two objects were in the same room, so response times did not increase with increased distance between the protagonist and the target object. However, in Experiments 3 and 4, the farther the protagonist was from the room in which the target object was located, the longer it took subjects to respond to it.

One might criticize the present study for its use of one-sentence "textoids" which have practically no context and do not resemble the texts that people read in real life. We admit that these short experimental stimuli are rather deficient as texts, but they were used for precisely that reason. In truth, many of the texts used in reading research are unnatural, and experimental subjects are not inherently inclined to read them. The need for experimental control usually results in texts that are short, bland, oddly written, and unfamiliar or irrelevant to the students who must read them. Our textoids were designed to mirror this fact. Decontextualized and tedious, they were almost guaranteed to be boring and of little relevance to our subjects, and this served to discourage our more motivated subjects from engaging in complete comprehension of their own accord. As a matter of control, we wanted our subjects' motivation to comprehend the texts to be prompted by the task instructions, not by features of the text itself.



Using short textoids enabled us to demonstrate what happens when a subject's motivation to read is very low, as it often is in many experiments. With more natural and interesting texts, readers are less likely to stop short of complete comprehension. However, because we cannot always predict whether an experimental text is interesting to a particular subject, we must require our subjects to perform tasks that specifically prompt the comprehension processes being studied. In other words, the only way language researchers can guarantee that experimental subjects will perform all of the comprehension tasks under investigation is to include experimental manipulations and procedures that force subjects to do what is desired. As Wilson et al. (1993) discovered in Experiment 2 of their study, even strong encouragement to perform a given mental operation will fall on deaf ears if this operation is not integral to the experimental task. This has long been an underlying assumption of cognitive research, and yet it is an assumption that, in many cases, is not entirely met. We know that we should run checks to insure that our subjects are actually performing the comprehension processes under investigation, yet, for some reason, we too often fail to do so. Stranger yet, when we find that our subjects are not performing mental operations that are both complicated and unnecessary for success in a given task, we seem surprised.

So what is the cautious researcher to do in order to ensure full subject cooperation? Certainly, it is in the best interests of the researcher to use questions and procedures that probe the specific types of knowledge one hopes to investigate. If studying anaphoric resolution, ask questions that require the anaphor be resolved. If studying structure building, ask questions that require subjects to comprehend the material fully and make structurally based inferences. Furthermore, using "natural" texts that are inherently interesting and enjoyable to read increases the chance that experimental subjects will be internally motivated to process those texts. With sufficient prodding, even the "minimalists" among us can be pushed down the path toward complete comprehension.

## REFERENCES

Anderson, A., Garrod, S.C., & Sanford, A.J. (1983). The accessibility of pronominal antecedents as a function of episode shifts in narrative text. *Quarterly Journal of Experimental Psychology, 36A*, 1–12.

Chafe, W.L. (1974). Language and consciousness. *Language, 50*, 111–113.

Chang, F.R. (1980). Active memory processes in visual sentence comprehension: Clause effects and pronominal reference. *Memory & Cognition, 8*, 58–64.

Clements, P. (1979). The effects of staging on recall from prose. In R.O. Freedle (Ed.), *New directions in discourse processing* (pp. 287–330). Norwood, NJ: Ablex.

Corbett, A.T., & Chang, F.R. (1983). Pronoun disambiguation: Accessing potential antecedents. *Memory & Cognition, 11*, 283–294.

Dee-Lucas, D., Just, M.A., Carpenter, P.A., & Daneman, M. (1980). What eye fixations tell us about the time course of text integration. In R. Groner & P. Fraisse (Eds.), *Cognition and eye movements* (pp. 155–168). Amsterdam: North-Holland.

Garnham, A. (1981). Anaphoric reference to instances, instantiated and non-instantiated categories: A reading time study. *British Journal of Psychology, 72,* 377–384.

Garnham, A. (1984). Effects of specificity on the interpretation of anaphoric noun phrases. *Quarterly Journal of Experimental Psychology, 36A,* 1–12.

Garnham, A., & Oakhill, J. (1985). On-line resolution of anaphoric pronouns: Effects of inference making and verb semantics. *British Journal of Psychology, 76,* 385–393.

Garrod, S., & Sanford, A. (1977). Interpreting anaphoric relations: The integration of semantic information while reading. *Journal of Verbal Learning and Verbal Behavior, 16,* 77–90.

Gernsbacher, M.A. (1989). Mechanisms that improve refrential access. *Cognition, 32,* 99–156.

Gernsbacher, M.A. (1990). *Language comprehension as structure building.* Hillsdale, NJ: Erlbaum.

Haberlandt, K., Berian, C., & Sandson, J. (1980). The episode schema in story processing. *Journal of Verbal Learning and Verbal Behavior, 19,* 635–650.

Haviland, S.E., & Clark, H.H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior, 30,* 495–504.

Karmiloff-Smith, A. (1980). Psychological processes underlying pronominalization and nonpronominalization in children's connected discourse. In J. Kreiman & A.E. Ojeda (Eds.), *Papers from the parasession on pronouns and anaphora* (pp. 231–250). Chicago: Chicago Linguistic Society.

Lesgold, A.M. (1972). Pronominalization: A device for unifying sentences in memory. *Journal of Verbal Learning and Verbal Behavior, 11,* 316–323.

Mandler, J.M., & Goodman, M.S. (1982). On the psychological validity of story structure. *Journal of Verbal Learning and Verbal Behavior, 21,* 507–523.

Matthews, A., & Chodorow, M.S. (1988). Pronoun resolution in two-clause sentences: Effects of ambiguity, antecedent location, and depth of embedding. *Journal of Memory and Language, 27,* 245–260.

McKoon, G., & Ratcliff, R. (1992). Inferences during reading. *Psychological Review, 99,* 440–446.

Olsen, G.M., Duffy, S.A., & Mack, R.L. (1984). Thinking-out-loud as a method for studying real-time comprehension processes. In D.E. Kieras & M.A. Just (Eds.), *New methods in reading comprehension research* (pp. 253–286). Hillsdale, NJ: Erlbaum.

Sanford, A.J., & Garrod, S.C. (1980). Memory and attention in text comprehension: The problem of reference. In R.S. Nickerson (Ed.), *Attention and performance VIII* (pp. 459–474). Hillsdale, NJ: Erlbaum.

Wilson, G.W., Rinck, M., McNamara, T.P., Bower, G.H., & Morrow, D.G. (1993). Mental models and narrative comprehension: Some qualifications. *Journal of Memory and Language, 32,* 141–154.

Yekovich, F.R., & Walker, C.H. (1978). Identifying and using referents in sentence comprehension. *Journal of Verbal Learning and Verbal Behavior, 17,* 265–277.

Zwaan, R.A., & Graesser, A.C. (1993a). Reading goals and situation models. *Psycholoquy, 4*(3), reading-inference.4.

Zwaan, R.A., & Graesser, A.C. (1993b). There is no empirical evidence that some inferences are automatically or partially encoded in text comprehension. *Psycholoquy, 4*(5), reading-inference.6.