Journal of Experimental Psychology:
Learning, Memory, and Cognition
1992, Vol. 18, No. 2, 266-283

0278-7393/92/$3.00

Pronoun Resolution and Discourse Models

Steven B. Greene
Stanford University

Gail McKoon and Roger Ratcliff
Northwestern University

Psychological investigations of pronoun resolution have implicitly assumed that the processes involved automatically provide a unique referent for every pronoun. We challenge this assumption and propose a new framework for studying pronoun resolution. Drawing on advances in discourse representation and global memory modeling, this framework suggests that automatic processes may not always identify a unique referent for a pronoun. In 9 experiments, we demonstrate that, unlike noun anaphors, pronouns sometimes do not produce relative facilitation of their referents in comparison with nonreferents. We argue that research on pronoun resolution must consider the discourse contexts in which pronouns are likely to occur.

When we encounter a pronoun in a discourse, we usually feel as if we understand its referent immediately (cf. Clark & Sengul, 1979). We are not consciously aware of any pronoun resolution mechanism operating or of any disambiguation strategies that we might use. Because of this unawareness, most psycholinguists studying pronominal reference have been tempted to assume that the psychological process involved is automatic. That is, researchers implicitly assumed that the process under investigation in studies of pronoun resolution is always triggered when a reader encounters a pronoun and that the process is always carried through to completion: the identification of a unique referent for every pronoun. The questions for recent research have been how soon after the occurrence of the pronoun is the process triggered and how many possible referents are considered (cf. Chang, 1980; Corbett & Chang, 1983; Gernsbacher, 1989). Unfortunately, 15 years of research based on the belief that pronominal referents are always automatically identified have so far failed to produce a satisfactory account of the process of pronoun resolution.

In this article, we propose a new framework within which to view the process of pronoun resolution. This framework is motivated by both empirical and theoretical considerations. First, we take seriously the notion of an automatic process (Neely, 1977; Posner & Snyder, 1975; Ratcliff & McKoon, 1981). Previous research on pronoun resolution has left the assumption of automaticity implicit and, thus, untested. One goal of the present research is to state explicitly what is automatic and what is strategic in pronoun resolution and to subject these claims to empirical verification. More important, our theoretical framework draws from contemporary work in discourse representation and in global memory models. Whereas early theories of discourse comprehension were based on the verbal learning tradition and modeled discourse as a single dimensioned list of clauses or propositions ordered serially or hierarchically (e.g., Clark & Sengul, 1979; Jarvella, 1971; Kintsch, 1974), recent discourse models organize information in multidimensional ways that more strongly reflect local context (e.g., Grosz, Joshi, & Weinstein, 1983; Webber, 1983). Similarly, most of the early process models for identifying referents of pronouns used either explicitly or implicitly a serial linear or hierarchical search (e.g., Clark & Sengul, 1979; Corbett & Chang, 1983; Hobbs, 1978; van Dijk & Kintsch, 1983; see Matthews & Chodorow, 1988, for a review). These models were inspired by the memory scanning retrieval models of the time (e.g., serial scanning models; Murdock, 1974), which have now largely been replaced by global parallel retrieval models (e.g., Gillund & Shiffrin, 1984; Hintzman, 1988; Murdock, 1982; Ratcliff, 1978). Hence, we replace the metaphor of the pronoun as a trigger initiating a serial search through a minimally structured textual representation with that of the pronoun as a cue to the most likely entity in a rich discourse representation.

Viewed in this way, the problem for research is not to investigate the mechanics of how a search process triggered by a pronoun might proceed but instead to investigate how a discourse model is constructed during comprehension so as to make the use of pronouns felicitous. In current conceptions, a discourse model represents the entities and events evoked by a discourse and the relationships among them (Grosz, 1981; Grosz et al., 1983; Grosz & Sidner, 1986; Sidner, 1983a, 1983b; Webber, 1983). Each entity is assumed to have some degree of accessibility, which is determined in part by the syntactic and semantic structures in which it is linguistically expressed. Accessibility is measured relative to the local environment, that is, relative to the other entities introduced in nearby clauses and sentences. As the reader or listener moves through a discourse, the accessibility of entities changes as the local environment changes. The entity or entities that are most accessible at any point are what the discourse is about at that point, a notion that various authors attempted to

This research was supported by National Science Foundation (NSF) grant 85-16350 and Air Force Office of Scientific Research grant 90-0246 (jointly funded by NSF) to Gail McKoon and by National Institute of Mental Health grants MH44640 and MH00871 to Roger Ratcliff. Steven Greene, who is now at Princeton University, was supported by an NSF graduate fellowship.

We thank Susan Brennan, Herbert Clark, Morty Gernsbacher, Richard Gerrig, Ray Gibbs, Brian MacWhinney, and Gregory Ward for helpful comments on the research presented here.

Correspondence concerning this article should be addressed to Gail McKoon, Psychology Department, Northwestern University, Evanston, Illinois 60208.

capture in the concepts of a discourse segment's "focus" (Sidner, 1983a), "center(s)" (Grosz et al., 1983), or "topic" (Reinhart, 1982) and which we refer to by the term "focus of attention."

One indicator of the relative accessibility of various entities in a discourse model is provided by syntax. Different syntactic structures can be used to emphasize some entities and de-emphasize others (cf. Sidner, 1983a). For example, compare

Barry saw Harriet.

and

It was Barry who saw Harriet.

In contrast to the first sentence, the second sentence makes it clear that the discourse is more about Barry than Harriet, with the consequence that Barry will be more accessible for future reference than Harriet. Empirical evidence confirms that the syntactic structures used to describe an entity affect the accessibility of that entity. For example, Matthews and Chodorow (1988) reported that reading times for the final word of the following sentence:

When the food was prepared by the owner of the restaurant, it was always delicious.

are shorter than those for this sentence:

When the owner of the restaurant prepared the food, it was always delicious.

This suggests that readers have less trouble identifying a referent for the pronoun *it* when the referent is introduced in subject position than when it is introduced in object position (even though the referent is more recent in object position). In a similar vein, McKoon, Ward, Ratcliff, and Sproat (1991; see also Rothkopf, Koether, & Billington, 1988) found that a modifying property is more accessible if it is introduced as a predicate than as a prenominal adjective; for example, the modifier *hostile* was more accessible in the sentence "His intolerant aunt was hostile" than in the sentence "His hostile aunt was intolerant." McKoon et al. (1991) also showed that a noun is more accessible if introduced in a verbal complement (*hunting deer*) than in a nominal compound (*deer hunting*).

Semantic and pragmatic factors also contribute to the relative accessibilities of discourse entities. For example, the perceived causal agent of a verb may be more accessible than its other arguments (Hudson, Tanenhaus, & Dell, 1986), and a discourse entity may be more accessible if it is more closely related to the topic of its discourse (McKoon et al., 1991). In addition, changes in relative accessibility can be signaled by certain conventional words and phrases that are used to indicate a shift in discourse focus (Grosz, 1981).

The accessibility of entities in a discourse is determined not only by the local environment at the time they are initially introduced but also by subsequent reference to them or to objects or properties associated with them. For example, noun anaphors can increase the accessibility not only of the concept to which they refer but also of other concepts that were mentioned in the same clause as the noun with which they corefer (Dell, McKoon, & Ratcliff, 1983). Certain concepts also permit the use of "associative anaphora" (Hawkins, 1977): After introducing the topic of a car, a reference to "*the* steering wheel" is felicitous. The initial reference to the car makes its parts accessible enough that they can be referred to using the definite article, usually reserved for previously mentioned entities (see also Chafe, 1976; Clark & Marshall, 1981; Prince, 1981).

The framework we put forward here is intended to suggest how referents for pronouns can be identified in the context of a highly structured discourse model rather than the simple linear representation implicit in previous research (e.g., Clark & Sengul, 1979; Corbett & Chang, 1983). In our framework, a pronoun must be evaluated against the rich and complex structure established by the syntactic, semantic, and pragmatic factors that determine the relative accessibilities of the different entities in the discourse. We propose that a pronoun can be completely and correctly understood if its intended referent is sufficiently more highly accessible in the comprehender's discourse model relative to the pronoun as a cue than all other discourse entities. We base the process by which a pronoun is matched against possible referents on current global memory models (Gillund & Shiffrin, 1984; Hintzman, 1988; Murdock, 1982; Ratcliff, 1978; see also Gernsbacher, 1989). In the proposed process, the semantic and grammatical features provided by an anaphor (as a retrieval cue) are matched automatically and in parallel against the semantic features of all entities in the current discourse model. A particular entity will match the anaphor to some degree depending on how accessible the entity is from the anaphor as a cue. Both the features of the entity (e.g., gender and number) and its accessibility will contribute to a determination of the degree to which it matches. If the degree of match for a single discourse entity is sufficiently high and better than the match for all other entities, that entity is automatically identified as the anaphor's referent. If there is no entity that matches sufficiently well, then no referent is identified, and selection of a referent is postponed or some kind of strategic (problem-solving) process can be invoked. If more than one entity matches sufficiently, then again selection is postponed to wait for more content from the discourse, or strategic problem solving can be attempted. In the usual case, when one entity matches sufficiently better than all others, the information in the propositions that include the anaphor is combined with the information from the propositions that include the referent entity.

Hence, in this framework, pronouns are resolved either by an automatic matching process or, if that process fails to produce a discourse entity that matches the pronoun sufficiently better than all other entities, an optional strategic process. This account of the mechanism by which pronouns cue potential referents can be applied to a variety of different discourse contexts. Most often, a pronoun is used to refer to a single discourse entity that is already easily accessible based on the syntactic and semantic context in which it was introduced: an entity that is in the reader's or listener's focus of attention (Brennan, 1989; Chafe, 1974; Fletcher, 1984; see

also Givon, 1976). In this situation, the pronoun matches a focused entity to a high degree and sufficiently better than all other entities in memory. As a result, the propositions that include the pronoun can be simply and automatically attached to the entity that is in focus at the time of the pronoun's use with the consequence that the accessibility of the focused entity is maintained or enhanced. Pronouns are usually used when the focus of attention of the discourse has not shifted (Grosz et al., 1983), so the default procedure of attaching new propositions to focused entities may have little processing cost.

Although pronouns may often be used to refer to a single, most accessible entity, a processing model in which a pronoun can vary in the degree to which it matches previously evoked entities leads directly to the possibility that sometimes there may be no discourse entity that matches sufficiently better than all others. This could come about either because no entity matches well or because several entities match about equally well. In these cases, no referent is automatically and uniquely identified for the pronoun. Various factors, such as the reader's or speaker's speed, the reader's or listener's comprehension goals, and the surrounding discourse context may conspire to make this possibility more or less likely. Variations in these factors can affect the degree to which a pronoun evokes its intended referent so that in some contextual conditions a pronoun will succeed in matching its intended referent, whereas in others it may fail to do so. In the case in which no discourse entity matches sufficiently well and strategic processes are not invoked, then no referent will be identified, and there may be no effect on the relative accessibilities of discourse entities as a result of reading the pronoun. When several entities are simultaneously in the focus of attention, they may all match the pronoun about equally well, and none of them would be singled out as the unique referent. Information about the pronoun would be attached to them jointly as the focus with the consequence that their relative accessibilities would not change as a result of reading the pronoun.

The possibility that people might sometimes fail to identify unique referents for pronouns has been suggested in the linguistic literature. Emphasizing the need to take the comprehender's purposes into account, Yule (1982) argued that comprehenders will sometimes interpret the discourse "in terms of some information marked for attention predicated of some individual or group, the referential identity of which is not an issue" (p. 319). Webber (1983) made a similar point: If there is no single best matching discourse entity for an anaphor, and if there is no immediate need to choose a referent for the anaphor, then the comprehender may simply leave the reference unresolved. If readers or listeners have little inducement to identify the referent of a pronoun, they may simply associate the information from the propositions that include the pronoun with whatever entities are currently accessible.

Our proposal—that anaphoric processing involves an automatic matching process that may sometimes fail to produce a referent—cannot be evaluated with respect to past research in any simple way. In the earliest studies of anaphoric reference (cf. Clark & Sengul, 1979; Haviland & Clark, 1974), it was assumed that the referents of pronouns would always be identified (probably a correct assumption for the texts that were used), and the exact point at which identification took place was not at issue. The only question was how difficult the identification process would be, and difficulty was measured by reading time. The more difficult the identification process for a pronoun in a sentence, the longer the reading time for the sentence. In more recent studies, the questions at issue have changed to focus on whether, and when in the time course of processing, a referent for an anaphor is understood (Chang, 1980; Corbett & Chang, 1983; Dell et al., 1983; Ehrlich & Rayner, 1983; McKoon & Ratcliff, 1980, 1981, 1984; Nicol & Swinney, 1989; Tanenhaus, Carlson, & Trueswell, 1989). The results of these studies still do not lead to a direct test of our proposal, but the studies do offer an appropriate methodology. We first explain the methodology and then consider the possible implications of previous results.

The procedure introduced by Chang (1980; also Caplan, 1972) was a probe task in which possible referents of a pronoun are presented as test words for recognition. Subjects read or listen to a short discourse that describes two characters and then refers unambiguously to one of them with a pronoun. At some point after the pronoun, the subject is shown a character's name and is asked to verify that the character was mentioned in the discourse just presented. The tested name can be either the intended referent, the other character, or some name that was not in the discourse at all. For example, in the final sentence in Table 1, the pronoun *she* is intended to refer to Mary, and either Mary, John, or some other name could be presented as a test word. For the character names that are in the discourse, the correct response is "Yes, the name was mentioned in the discourse." The result that was always expected by previous researchers is that responses to the name of the intended referent, *Mary* in Table 1, will be faster and more accurate than responses to the name of the other character, *John.* The reasoning is that the processes by which the pronoun is understood leave the intended referent in a more accessible state than the other possible referent, and this increased accessibility leads to relative facilitation for the referent as a test word.

Our proposed framework differs from previous views in the claim that the unique referent of a pronoun may or may not be identified depending on contextual conditions. Under some conditions, the automatic process of matching the features of a pronoun against the features of entities in memory will succeed in producing a discourse entity that matches the pronoun sufficiently better than other entities, and so the referent of a pronoun will be uniquely identified. The result will be to leave the identified referent in a state of high accessibility that will, in turn, lead to relative facilitation when the referent is presented as a test word. However, under other conditions, the process may fail to identify uniquely the intended referent, and then its accessibility will not be high relative to the accessibilities of other possible referents with no resulting facilitation for the intended referent relative to other test words.

Tests of this proposed framework depend critically on the assumption that the matching process of pronoun resolution is relatively fast and automatic. This assumption is adopted

Table 1
Example of the Experimental Texts

Mary and John were doing the dishes after dinner.
One of them was washing while the other dried.
Mary accidentally scratched John with a knife
and then$_1$ she dropped$_2$ it on the counter.$_3$
Test words
Referent: Mary
Nonreferent: John
Control: dishes

because it accords with our intuition that pronouns are normally processed quickly and effortlessly. We make this assumption explicitly to distinguish the automatic matching process from other, more strategic, and usually slower processes that might come into play if a single, best matching entity is not produced.

In many previous studies that have used the probe-word procedure to investigate pronoun comprehension, reading times and response times have been slow enough that it is doubtful whether automatic processing could be claimed. Since Chang (1980) first used the test word procedure to investigate pronoun comprehension, others followed (Corbett & Chang, 1983; Gernsbacher, 1989) with a virtually unanimous result: Responses to the intended referent presented as a test word are facilitated relative to responses for other possible referents presented as test words. However, in each case, either reading times or response times, or both, seem slow. For example, Corbett and Chang (1983; Experiment 1) found faster responses for the intended referent than another possible referent, but response times were slow (800–900 ms) and so were reading times (about 380 ms per word controlled by the subjects). Gernsbacher (1989) used reading times of over 500 ms per word (controlled by the experimenter) with response times in the 1,000-ms range. In addition, previous studies may have encouraged strategic processing of pronouns not only by using slow reading rates but also by a specific task demand: asking for the identity of the pronoun immediately after reading. For example, for the text in Table 1, subjects would be asked "Who dropped it on the counter?" immediately after reading the text. The motivation provided by such a specific question in combination with a reading rate slow enough to give time to answer the question during reading may have led subjects to adopt strategies that they might not have under other task conditions.

Our goal for the experiments described in this article was to examine pronoun comprehension as an automatic process. To accomplish this, we changed the experimental procedures used in previous research in two ways. First, both the reading rate and the time for responding to the test word were speeded relative to previous experiments. Second, we eliminated task demands that might encourage special strategic processing of pronouns, such as questions about the referent of a pronoun. Both of these changes were motivated from general notions about automatic processing developed in research areas other than reading (cf. Posner & Snyder, 1975; Ratcliff & McKoon, 1981), and the application of these notions to reading is not straightforward. However, as will be seen, the procedural changes brought about substantial changes in experimental results, lending support to the application of an automatic/strategic distinction to investigations of reading processes.

The procedural changes designed to speed reading and response times were guided by findings from other research domains and by intuition. What times qualify as within the range of automatic processes is fairly clear for recognition responses from both Posner and Snyder's (1975) original studies and a number of other studies with various methodologies (e.g., Neely, 1977; Ratcliff & McKoon, 1981). However, for reading time, deciding what rates qualify as automatic presents a problem; it is not clear how automatic reading processes can be separated empirically from slower, strategic reading processes or even whether there is a clearly separable dichotomy between the two kinds of processing in reading. We decided to speed up the presentation rate of our materials from the rates used by earlier researchers to a rate more nearly approaching what college students have been estimated to use normally. Using texts considerably more difficult than those in the experiments presented here, other researchers (e.g., Just & Carpenter, 1980; Rayner, 1978) found average reading speeds in the range of 200 to 250 ms per word. For texts more similar to those in the following experiments, Ehrlich (1983) found mean eye fixation times consistently below 300 ms, but because only about two thirds of the words of a typical text are actually fixated (Just & Carpenter, 1987), one can calculate the mean effective reading speed to be about 200 ms per word. In fact, Just and Carpenter (1987) considered a reading rate of 240 words per min or 250 ms per word to be "normal" (p. 38). Therefore, in our experiments, we set the reading rate at 250 ms per word. We also instructed subjects to respond quickly with high accuracy with the intention that response times should be in the 700-ms range. On the basis of past experiments (e.g., Dell et al., 1983; McKoon & Ratcliff, 1986, 1989b), we expected that subjects would be able to achieve this level of performance.

The materials in our experiments were modeled on those typical of previous studies of pronouns (Chang, 1980; Corbett & Chang, 1983; Gernsbacher, 1989) except that we used longer texts. Each text began with a sentence that introduced two characters with proper names, continued with a sentence that did not emphasize either character, and concluded with a final sentence made up of two clauses. In the first of these clauses, both characters' names were mentioned (in the same order as in the first sentence), and in the second clause, there was a pronoun intended to refer to the first-mentioned character (the subject of the first clause). The pronominal reference was unambiguous both because the sex of the two characters differed and because the predicate of the second clause described an action that could be performed only by the referent character. An example of one of the texts is shown in Table 1.

In a discourse model of this text, the two characters would be of about the same accessibility. Both were introduced at the beginning of the discourse, and both were rementioned in the first clause of the final sentence. However, the first-mentioned character might enjoy a slight advantage simply because of being mentioned first (Gernsbacher & Hargreaves, 1988; Gernsbacher, Hargreaves, & Beeman, 1989). Also the first-mentioned character was the subject of the first clause of

the final sentence, and the grammatical subject of a sentence is a good candidate for coreference with a subsequent subject-position pronoun (Matthews & Chodorow, 1988; Sidner, 1983a). Therefore, before the reader encounters the pronoun, the first-mentioned subject character may be more accessible than the object character. This initially higher accessibility might lead to a sufficiently higher match between the subject character and the pronoun (assuming also a match in gender and number), so that the subject character is identified as the referent. As a result, the propositions that include the pronoun would be attached to those that include the subject character. The processing involved in attaching the propositions might further increase the referent's accessibility, giving an advantage to the referent when it is presented as a test word.

Alternatively, the grammatical subject might not have an advantage over the grammatical object. The object of a verb in the main clause of a sentence is also often a good candidate for subsequent pronominalization (Clifton & Ferreira, 1987; Sidner, 1983b). Thus, the subject and object might not differ in accessibility; they might both be in the reader's focus of attention. In this case, the only information available that unequivocally distinguishes referent from nonreferent would be gender. It might be that the gender information could be weighted strongly enough by the matching process to give a sufficiently higher degree of match for the intended referent. On the other hand, the gender information might not be sufficient to distinguish between two entities jointly in the focus of attention; then the match between the pronoun and the intended referent might not be sufficiently higher than that between the pronoun and the nonreferent. In this situation, subjects could engage in further, possibly strategic, processing to choose between the possible referents. Alternatively, they could simply attach the new propositions to the discourse entities that are jointly in the focus of attention, failing to identify just one of them as the unique referent for the pronoun because they are both in the focus of attention. In this case, processing of the pronoun would give no advantage in accessibility to either of the two characters over the other.

Experiments 1 and 2 were designed to distinguish between the two hypotheses just described: The subject character might have an advantage in the degree to which it matched the pronoun as cue, because of its higher accessibility and appropriate gender, so that it is identified as the referent of the pronoun and therefore given an increase in accessibility. Alternatively, it might be that neither character has a sufficiently great advantage to be uniquely identified as the referent, and thus neither would gain in relative accessibility. The first hypothesis predicts that processing of the pronoun will facilitate responses to the intended referent relative to responses to the other character name, whereas the second hypothesis predicts that there will be no facilitation of the referent relative to the other character. If the second hypothesis is upheld, it suggests that readers do not always identify a unique referent each time they encounter a pronoun.

The following experiments suggest that readers do not, in fact, always automatically identify referents for pronouns. In Experiments 1 and 2, processing of the pronoun did not facilitate responses to the referent test word relative to the nonreferent test word. Because this is a null result, we conducted a further seven experiments. Experiments 3 and 4 added more subjects and used pronouns for which the intended referent was the object instead of the subject of the first clause of the final sentence. There was still no relative advantage of referent test words over nonreferent test words. Experiments 5, 6, and 7 compared our procedure (relatively fast reading times and relatively fast responses) to a procedure with much slower reading times and response times that has previously been shown to produce facilitation of referents relative to nonreferents (Gernsbacher, 1989). With the slow procedure, we did find facilitation of referents relative to nonreferents but only when the experimental texts were short enough that subjects could predict the occurrence of the pronoun and the test word. This pattern suggests that our finding of no relative facilitation of referents differed from past findings of facilitation because of the difference in procedures and materials. We argue that, with the slow procedure and the predictable materials, subjects invoke strategic processes to resolve the pronoun references. Finally, in Experiments 8 and 9, we used the fast procedure to compare comprehension of the pronouns to comprehension of nominal anaphors. We replicated what has previously been shown (Dell et al., 1983): that processing of a nominal anaphor, such as *the criminal,* facilitates responses for its referent (*burglar*) and also responses for words associated in the text with the referent. Thus, we show that our fast presentation rate is not so fast that it prevents all types of anaphoric processing. In the discussion section, we argue that automatic processing of anaphors does occur with our fast procedure, as evidenced by the results for nominal anaphors, but that automatic processing does not identify a single best referent for the pronouns under investigation. Instead, the propositions that include the pronoun are simply attached to the entities in the focus of attention at that point in the discourse. Because the texts used in these experiments leave both the referent and the nonreferent characters in the focus of attention, neither is given an advantage over the other.

Experiments 1 and 2

An example of the texts used in these experiments appears in Table 1. As previously described here, the first sentence introduced two characters of different gender, the second sentence did not emphasize either character, and the final sentence consisted of two clauses. The first clause of the final sentence had one of the characters as subject and the other as object, and the second clause referred to the subject character with a pronoun. The words of the texts were presented on a cathode ray tube (CRT) screen one at a time at the rate of 250 ms per word. When a test word was presented for recognition, all preceding words of the text were erased from the screen, and subjects were instructed to respond "yes" if the test word had appeared in the text just presented and "no" if it had not.

The aim of the experiments was to determine whether processing of the pronoun gave a relative advantage in accessibility to the referent character. Exactly how to design experiments to address issues like this has been the subject of considerable discussion (cf. Dell et al., 1983; MacDonald &

MacWhinney, 1990). It is first important to distinguish two different questions that might be asked: whether the referent has an advantage relative to the nonreferent and whether the referent has an advantage relative to some neutral baseline. We were mainly concerned with the first question for which the choice of experimental design is straightforward. To find out whether processing of the pronoun gives a relative advantage to the referent test word, we compared responses to the referent and nonreferent test words when the test words were presented before the pronoun to responses when the test words were presented after the pronoun. If processing of the pronoun gives an advantage to the referent, then whatever difference there was in referent and nonreferent responses before the pronoun ought to change in the direction of relative facilitation for the referent. There might, of course, be changes in baseline response time or accuracy as the test point is changed from before the pronoun to after the pronoun, but this would be a simple main effect that should not obscure any change in the relative differences of referent versus nonreferent responses.

We implemented this design in Experiment 1 with two test positions for the referent and nonreferent test words. One test position was immediately before the pronoun in the final clause, and the other was after the word following the pronoun; these are Test Positions 1 and 2 in Table 1. With the text presented at 250 ms per word, the test at Position 2 occurred 500 ms after the pronoun was displayed. Experiment 2 was the same as Experiment 1 except that the two test positions were immediately before the pronoun and at the end of the final clause: Test Positions 1 and 3.

Although we were mainly interested in the relative facilitation given by processing of the pronoun to the referent and nonreferent characters, we also included in the design a test of a hypothesis put forward by Gernsbacher (1989). She proposed that processing of a pronoun gives relative facilitation to the referent test word by means of suppressing the accessibility of nonreferents. As support for this hypothesis, she showed that response times to a nonreferent test word slowed at the end of a sentence containing a pronoun, whereas response times for the referent test word stayed about the same as before the pronoun (Gernsbacher, 1989, Experiment 3). To test her hypothesis, we included a control test word in Experiments 1 and 2. This was a word that had appeared in the text in the first or second sentence (so the correct response for recognition was "yes," the same as for the referent and nonreferent test words). By presenting this word at the same two test points as the referent and nonreferent test words, we could trace changes in response times that should be independent of effects of processing the pronoun. For example, it might be that responses for all test words are slower at the end of a sentence than in the middle of a sentence because the end-of-sentence test word is competing for processing capacity with end-of-sentence comprehension processing. If this were the case, then further research would be needed to support the suppression hypothesis.

It is important to note that the control test word was included only to address the suppression hypothesis. Neither the control word nor any combination of the conditions in the experiment allows the issue of true facilitation relative to a neutral baseline to be addressed. As was pointed out, this issue is not directly relevant to the hypotheses of concern in this article.

Method

Materials. The 60 experimental texts were short three-sentence texts as previously described here. Many of them were based on sentences used by Corbett and Chang (1983). For half of the texts, the first-mentioned character of the first and third sentences was male and for the other half, female. The pronouns in the second clause of the third sentence were always of the same sex as the first-mentioned character. None of the verbs in the first clauses of the final sentences were of the causally biased kind studied by Garvey, Caramazza, and Yates (1976). The test words for the texts were the two character names and a control word from the first or second sentence (usually a noun). The average length of the first and second sentences combined was 18 words, and the average length of the third sentence was 15 words. The number of words between the first character name in the first clause of the third sentence and the pronoun in the second clause averaged 7.9, the number of words between the other character name and the pronoun averaged 3.5, and the number of words between the pronoun and the end of the sentence averaged 6.4.

There were also 60 filler texts used to provide different kinds of test words from the experimental texts. These texts were all three sentences (four lines on the CRT screen) and averaged 34 words in length. Each text had one test word. Forty-five of the test words were negatives (they had not appeared in the text), and 15 were positives. Forty of the test words were tested in the first three lines of their text and 20 in the last line. Twenty-five of the test words were names (7 positive) and 35 were other nouns (8 positive). Each filler had associated with it one true test statement and one false test statement that were written to test a variety of kinds of information from the texts. Some examples of the information tested by the true and false statements include: whether the Cubs game was in the afternoon or evening; whether there were no eggs in the refrigerator or a dozen; whether there were or were not ripe melons at the grocery store; whether a milk shake was chocolate or vanilla.

Procedure. All of the texts and test items were presented on a CRT screen, and responses were collected on the CRT keyboard. Each subject participated in one 50-min session.

The experiment began with 150 lexical decision test items. These items were included to give subjects practice with the response keys on the CRT keyboard. After this practice, there were 20 filler texts, and then the remainder of the texts—60 experimental texts and 40 fillers—were presented in random order. A different random order of presentation of materials was used for every second subject.

Each text began with the instruction to press the space bar on the keyboard to initiate the text. When the space bar was pressed, the text was presented one word at a time. Each word was displayed for 250 ms, then the next word was displayed for 250 ms, and so on until a complete line of the text appeared across the screen. The last word of a line was displayed for 300 ms, and then the whole line was erased, and the next line was displayed in the same manner. When a test word was presented, the current line of text was erased, and the test word appeared where the next text word would have been. The letters of the test word were all in uppercase (unlike the words of the text), and two asterisks were displayed immediately to its right. The test word remained on the screen until a response key was pressed (?/ for "Yes, the word had appeared in the text," and z for "no, the word had not appeared in the text"). After the response and a pause of 100 ms, the text continued unless the response was an error or the response was too slow. If the response was an error, the word *error* was displayed for 1,500 ms before the text continued. If the response

was slower than 1,000 ms, the message *too slow!* was displayed for 500 ms. This response time feedback was included because, in pilot experiments, some subjects had extremely slow response times. In similar experiments reported by Dell et al. (1983), mean response times averaged about 600 ms. The filler texts were followed by a true–false test statement, and incorrect responses to this test statement were followed by the *error* message. Each filler text had a true and a false statement; which one of these was presented was chosen randomly.

Subjects and design. For both experiments, there were two variables: Two test positions were crossed with three test words. The test words were the intended referent of the pronoun in the final clause, the other character name that was not the intended referent, and the control word from earlier in the text. For Experiment 1, the test positions were immediately before the pronoun in the final clause (Test Position 1) and after the word following the pronoun (Test Position 2). For Experiment 2, the test positions were immediately before the pronoun (Test Position 1) and at the end of the sentence (Test Position 3). In each experiment, the two variables were crossed with 6 sets of items (10 per set) and 6 groups of subjects. In each experiment, there were 36 subjects participating to fulfill a requirement in an introductory psychology course.

Results

Means were calculated for each subject and each item in each condition, and means of these means are shown in Table 2. In all of the experiments to be reported, the error rates represent items for which the response was incorrect. Also it should be noted that response times are slower and error rates higher on filler items compared with the name test items of interest. We assume this is because the positive filler test words were from farther back in the text than the name test words and were less memorable words and because negative test words usually have slower response times in experiments of this type. Analyses of variance (ANOVAs) were conducted on both subject (F_1) and item (F_2) means, and $p < .05$ was used unless otherwise noted. Standard errors of the means are given from the subjects' analyses; standard errors from the items' analyses were comparable.

Table 2
Results of Experiments 1 and 2: Response Times (RTs) and Error Rates (ERs) on Test Words

	Test position					
	1		2		3	
Test word	RT (ms)	ER (%)	RT (ms)	ER (%)	RT (ms)	ER (%)
	Experiment 1[a]					
Referent	656	7	669	10		
Nonreferent	633	4	624	3		
Control	729	12	746	15		
	Experiment 2[b]					
Referent	675	7			697	7
Nonreferent	654	5			695	2
Control	705	11			784	20
	Procedure check experiment[c]					
Referent	721	8			731	8
Nonreferent	712	8			718	4
Control	785	15			845	24

[a] Response time and error rate for positive fillers are 779 ms and 11%, respectively, and for negative fillers, 832 ms and 13%, respectively.
[b] Response time and error rate for positive fillers are 711 ms and 26%, respectively, and for negative fillers, 799 ms and 15%, respectively.
[c] Response time and error rate for positive fillers are 820 ms and 22%, respectively, and for negative fillers, 829 ms and 12%, respectively.

In Experiment 1, with Test Positions 1 and 2, there were no significant differences between the test positions (both F_1 and $F_2 < 1.3$) and no interactions between test word and test position (Fs < 1.7). The only significant effect was for test word, $F_1(2, 70) = 61.4$ and $F_2(2, 118) = 64.0$. The response times for the control test word were slower than for the other test words (Fs > 42). The standard error of the response time means was 8 ms. The only significant effect for error rates was the difference among test words, $F_1(2, 70) = 24.7$ and $F_2(2, 118) = 19.1$; the control test words had more errors than the other test words (Fs > 13). The standard error for errors was 1.6%.

True test statements had mean response times of 1,737 ms with 12% errors; false test statements had mean times of 1,603 ms with 20% errors.

The pattern of results was similar for Experiment 2 in that there were no significant differences between the referent and nonreferent test words as a function of test position. The effect of test word was significant, $F_1(2, 70) = 4.0$ and $F_2(2, 118) = 22.0$, as was the effect of test position, $F_1(2, 70) = 47.9$ and $F_2(1, 59) = 28.2$, and the interaction of the two variables, $F_1(2, 70) = 7.6$ and $F_2(2, 118) = 4.2$. The significant interaction is due to the difference between the control test word and the other test words; it does not reflect a difference in the effect of test position on the referent and nonreferent test words. Although the referent does not slow as much from the first to second test positions (22 ms) as the nonreferent (41 ms), suggesting relative facilitation for the referent, the difference was not significant by post hoc tests, $F_1(1, 70) = 2.7$ and $F_2(1, 118) < 1.0$. The control test words had slower response times than the other test words (Fs > 24). Standard error of the response time means was 6 ms. In both experiments, nonreferent response times were somewhat faster than referent response times, suggesting a slight recency effect.

Error rates showed the same effects as response times. Differences among error rates were significant for test words, $F_1(2, 70) = 18.4$ and $F_2(2, 118) = 15.8$, and the interaction of test word and test position was significant, $F_1(2, 70) = 6.5$ and $F_2(2, 118) = 4.6$. The control test words had more errors than the other test words (Fs > 24). The standard error for errors was 1.2%.

For true test statements, the mean response time was 1,937 ms (13% errors) and for false test statements, 1,859 ms (19% errors).

Procedure Check

One question that might arise about the results of Experiments 1 and 2 concerns the extent to which they depend on the cumulative method of presenting the texts, with words appearing across the CRT screen and each word remaining

on the screen as the others were presented. An alternative, noncumulative method is to present all words in the same position on the CRT screen, each word erasing the preceding word. To check for differences between these two procedures, we replicated Experiment 2 with the noncumulative method, each word presented in the same CRT location at 250 ms per word (24 subjects). As can be seen in Table 2, the change in procedure brought about no significant change in results.

Experiment 3

In Experiments 1 and 2, the main result is a null result: Moving from the test position before the pronoun to test positions after the pronoun did not produce any significant facilitation of the referent test word relative to the nonreferent test word. This lack of effect is consistent with the hypothesis that processing of the pronoun does not distinguish between the two characters; we would attribute this to the two characters being equally in the focus of attention. However, before accepting the null result, we tested it further in Experiments 3 and 4.

Method

In Experiment 3, all three of the test positions used in Experiments 1 and 2 were combined in one experiment. The materials and procedure were the same as in Experiments 1 and 2 except that three more experimental texts were added. There were two variables: three test words and three test positions. These nine conditions were crossed with nine sets of texts (seven per set) and nine groups of subjects in a Latin square design. The 45 subjects participated for credit in an introductory psychology course.

Results

The results, presented in Table 3, again show no differences between the referent and nonreferent test words. By ANOVAs, there were main effects of test word, $F_1(2, 88) = 79.6$ and $F_2(2, 118) = 81.8$, and test position, $F_1(2, 88) = 27.8$ and $F_2(2, 118) = 14.5$, but no significant effect of their interaction ($Fs < 1.4$). Response times for the control test words were slower than for the other test words ($Fs > 34$). The standard error of the response time means was 10 ms. For error rates, the only significant effect was for test word, $F_1(2, 88) = 41.4$ and $F_2(2, 118) = 27.1$. The control test words had more errors than the other test words ($Fs > 17$). The standard error for errors was 1.6%. Response times for true test statements averaged 1,748 ms (11% errors) and for false test statements, 1,716 ms (18% errors).

We also analyzed the data by combining the first and third test positions from Experiments 2 and 3, making a total of 81 subjects. The interaction between test word (referent, nonreferent, and control word) and test position was not significant, with a standard error of 6 ms.

Experiment 4

As in Experiments 1 and 2, the results of Experiment 3 showed no significant facilitation of the referent relative to the nonreferent as test position moved from before the pronoun to the test positions after the pronoun. With a total of 117 subjects, this finding seems conclusive.

Table 3
Results of Experiment 3: Response Times (RTs) and Error Rates (ERs) on Test Words

	Test position					
	1		2		3	
Test word	RT (ms)	ER (%)	RT (ms)	ER (%)	RT (ms)	ER (%)
Referent	668	11	679	6	708	8
Nonreferent	643	5	652	4	699	4
Control	761	13	753	18	820	20

Note. Response time and error rate for positive fillers = 775 ms and 26%, respectively, and for negative fillers, 833 ms and 14%, respectively.

The finding is inconsistent with the results of past experiments (Chang, 1980; Corbett & Chang, 1983; Gernsbacher, 1989) in which referent test words were significantly facilitated over nonreferent test words. One possible reason for the difference in results was suggested early in this article: Different kinds of processing may have occurred in our experiments than in the previous experiments. The faster reading times and response times we used may have led to exclusively automatic processing of pronouns, and the slower reading times and response times in the earlier experiments may have led to more strategic processing. The only directly comparable previous research that might have used an equivalently fast presentation rate (MacDonald & MacWhinney, 1990, in which the auditory presentation rate was not specified) did not obtain consistent results across two experiments. In one of their experiments, response times to a referent probe were faster than response times to a nonreferent probe when they were tested immediately after the pronoun, but in a second experiment response times to the two probes did not differ when immediately tested. Also differences between referent and nonreferent response times at later test points were due in one experiment to a relative slowdown of the nonreferent response times from immediate testing to later testing; in the other experiment, they were due to a speedup of the referent. A further difference between past experiments and ours is that we used comprehension questions that tested a variety of kinds of information from the texts. In earlier experiments, the comprehension questions usually required identification of the intended referent for the pronoun by asking subjects to verify which character performed the action of the final clause. Like the slow reading times, these questions may have encouraged strategic kinds of processing during reading.

However, a difference in kind of processing is not the only possible reason for the discrepancy between the results of Experiments 1, 2, and 3 and earlier results. Another possibility might arise from the fact that the pronoun in the final clause in our experiments was always intended to refer to the character that was the subject of the first clause. In other studies, the pronoun sometimes referred to the subject and sometimes to the object. Therefore in Experiment 4 we changed half of our materials to make the object of the first clause the intended referent. It is also possible that there is some other unidentified difference between our materials and those used

previously that is relevant to pronoun comprehension. To check this possibility, we included in Experiment 4 a small set of materials from experiments by Gernsbacher (1989).

Method

For 28 of the texts used in Experiments 1, 2, and 3, the second clause of the final sentence was modified so that the pronoun referred to the character that was the object of the first clause and the action was consistent with having that character as agent. For example, the new version of the final clause for the text in Table 1 was *and he cried out in pain.* In addition, another 28 of the texts from the earlier experiments were used in their original versions, with no changes, so that the referent of the pronoun in the final clause was the character that was the subject of the first clause.

Twelve new texts, each a single sentence, were chosen from the materials used by Gernsbacher (1989). These sentences had the same form as the final sentences of our texts, with two characters mentioned in the first clause and a pronoun in the second clause for which one of the characters was referent. For half of these sentences, the referent was the subject of the first clause, and for half the referent was the object of the first clause.

The filler texts were the same as in the first three experiments, except that eight of them (four with positive and four with negative test words) were reduced to only a single sentence. The procedure was the same as in the first three experiments.

For the original 28 texts and the 28 texts that were modified to make the object of the first clause be the referent of the pronoun, there were four experimental conditions: A test word was presented either before the pronoun of the final sentence or at the end of the final sentence, and the test word was either the referent character name or the nonreferent character name. These four conditions were combined in a Latin square design with four groups of subjects and four sets of items (seven items per set). For the 12 new texts from Gernsbacher's materials, there was only one test point—the end of the sentence—and the test word was either the referent or the nonreferent. The two conditions were crossed with two sets of items (six per set) and two groups of subjects. There were a total of 40 subjects from the same population as the preceding experiments.

Results

The data for the 28 original texts and for the 28 modified texts are shown in Table 4. Just as in the preceding experiments, the data show no significant differences between referent and nonreferent test word responses as a function of test position. All responses slow from the first test point (before the pronoun) to the end of the sentence but not differentially. Analyses confirm the lack of an interaction between test word (referent versus nonreferent) and test position (Fs < 1.2 for response times and error rates for both subject and item analyses).

The effect of test position on response times was significant, $F_1(1, 39) = 31.9$ and $F_2(1, 54) = 24.8$. There was an interaction such that the difference in response times between subject and object test words was not the same for the two sets of sentences according to an ANOVA with subjects as the random factor, $F_1(1, 39) = 4.0$, but this interaction was not significant with items as the random factor, $F_2(1, 54) = 2.4$. *Responses were generally slower for the sentences in which* the intended referent of the pronoun was the subject, but this effect was marginally significant only with subjects as the random variable, $F_1(1, 39) = 3.4$. All other Fs were less than 2.5. The standard error of the response time means was 5.5 ms. There were no significant differences in error rates (all Fs < 2.9) and the standard error was 1.1%.

Table 4
Results of Experiment 4: Response Times (RTs) and Error Rates (ERs) on Test Words

	Test position			
	1		3	
Test word	RT (ms)	ER (%)	RT (ms)	ER (%)
Our materials				
Object (referent)	622	4	649	3
Subject (nonreferent)	638	3	672	4
Subject (referent)	645	5	667	6
Object (nonreferent)	635	3	671	3
Gernsbacher (1979) materials				
Referent			657	8
Nonreferent			645	5

Note. Response time and error rate for positive fillers are 722 ms and 20%, respectively, and for negative fillers, 765 ms and 10%, respectively.

For the Gernsbacher (1989) materials, there were no significant differences in either response time or error rate (Fs < 1.8). The standard error of the mean of the response times was 6.8 ms and of the errors, 1.5%. This result contrasts with Gernsbacher's finding of significant differences between the referent and nonreferent test words when the test words were presented at the ends of their sentences.

Responses to true test statements had a mean response time of 1,590 ms (14% errors), and responses to false test statements had a mean of 1,585 ms (22% errors).

Experiments 5 and 6

The conclusion from Experiments 1 through 4 is clear: For the sentences used in the experiments, referents and nonreferents are not differentially affected by processing of the pronoun. This conclusion holds over 157 subjects, over referents expressed as subjects and referents expressed as objects, over our materials as well as a subset of Gernsbacher's (1989) materials, and over cumulative and noncumulative procedures for presenting texts.

Our interpretation of this result is that subjects were engaging in sentence processing that does not require the referent of the pronoun to be uniquely identified. For the sentences of the experiments, both characters are about equally in the discourse focus of attention, and information in the pronoun's clause is attached to the focus and not to either of the characters individually. Therefore, neither character gains in accessibility relative to the other. From this interpretation, we can make two testable predictions. First, if we can change subjects' processing to the appropriate strategies, the intended referent should be uniquely identified, and we should see a relative advantage of referent over nonreferent test words. This was the aim of Experiments 5, 6, and 7. Second, we should be able to contrast the pronominal anaphors that are

not uniquely identified with other kinds of anaphors for which the referent is identified. We do this in Experiments 8 and 9.

To encourage subjects to adopt a strategy of identifying the referents of the pronouns during reading, we needed to give them motivation to do the appropriate processing; we needed to make it relatively easy for them to do it; and we needed to give them time to do it. To provide motivation, each text was followed by a comprehension question for which the answer required that the actor of an action in the final sentence be identified. For the experimental sentences, this always required that the referent of the pronoun in the final clause be identified. To make the appropriate processing easy, we used texts of only one sentence (for the experimental texts, this was the final sentence) so that subjects would know exactly what information the comprehension question would ask about and when to expect the pronoun in the text. To give subjects time to compute the intended referents of the pronouns, we adopted the procedure used by Gernsbacher (1989) in which the time available for processing each word was 450 ms plus 16⅔ ms multiplied by the number of letters in the word. With this procedure, Gernsbacher (1989) found a large relative advantage of referents over nonreferents at the end of the sentence, and we expected to replicate this effect.

In Experiment 5, the referent and nonreferent character names were tested either immediately before the pronoun or at the end of the sentence. As expected, we found a larger relative advantage for the referent test word over the nonreferent test word at the end of the sentence than before the pronoun, indicating that our efforts to change subjects' processing were successful. The advantage came from an increase in response times for the nonreferent test words, which is consistent with Gernsbacher's (1989) hypothesis that processing of the pronoun gives an advantage to the referent by suppressing the nonreferent. However, as discussed earlier here, this hypothesis can be tested with a control word. If suppression affects only the nonreferents, then the nonreferents should increase in response time at the end of the sentence relative to the referent, but the control word should not. This was tested in Experiment 6.

Method

The materials were the same as in Experiment 2 except that only the final sentence of each text was used, and there was one test word for each sentence. For the fillers, all the test words were negative, and half were tested in the sentence and half at the end of the sentence. For the experimental materials, the test words in Experiment 5 were the referent and nonreferent names tested in Positions 1 or 3. All of the negative test words for the fillers were also names. In Experiment 6, the test words were the referent and a control word tested in Positions 1 or 3. The control word was a word that appeared in the first clause of the final sentence; usually it was a noun. On average, there were 3.4 words between the control word and the pronoun of the second clause. In Experiment 6, only 40 of the experimental items were used in the design; the other 20 experimental items were used as fillers with the test word always the referent of the second clause pronoun tested in the first position half the time and in the third position half the time. For the negative test words, 13 of the 50 tested nouns were not names, and the rest were names. There were 36 subjects from the same population as the other experiments in Experiment 5 and 24 in Experiment 6.

The experiments began with 30 lexical decision test items presented for practice with the response keys. Then there were 10 filler texts, and then the 60 experimental texts and 50 remaining filler texts in random order. The procedure was modeled on the procedure used by Gernsbacher (1989). Each text began with an instruction to press the space bar to begin the text. Then the words of the text were displayed one at a time in the same location of the CRT screen (one on top of another). We used this noncumulative method of presentation to mimic Gernsbacher's procedure as closely as possible and because the procedure check in Experiment 2 showed no differences in results from cumulative versus noncumulative presentation. Each word remained on the screen for 300 ms plus the number of letters in the word multiplied by 16⅔ ms, and there was a 150-ms blank interval between words. A test word was displayed in the same position as the text words, with all letters in uppercase and with two asterisks on each side of it. When a key was pressed in response to the test word, the word was erased, there was a 150-ms pause, and then the text continued. There was no feedback about speed or accuracy.

After each text, a test question was presented. The question asked who did one of the actions in the final sentence of the text. The names of the two characters of the text were displayed with the question, and the subject was instructed to press the key appropriate for the correct choice (the "z" key for the left choice, the "?/" key for the right choice). For the experimental texts, the question always asked who did the action of the second clause of the final sentence, and the correct answer was the referent of the pronoun in that clause. For the filler texts, 24 texts asked about the action of the first clause, and 36 asked about the second clause. If the response to the test questions was incorrect, the word *error* was presented for 1,500 ms.

Results

Experiment 5. Means are shown in Table 5. As predicted, response times for the nonreferent test word increased from Test Position 1 to Test Position 3 more than response times for the referent. This interaction is significant with subjects as the random variable, $F_1(1, 35) = 5.4$, and approached significance with items as the random variable, $F_2(1, 56) = 3.7$, $p = .06$. The main effects of test position and test word were not significant (Fs < 2.7). The standard error of the response time means was 15 ms. Subjects were accurate on the "who did it" questions; error rates were 6% (1,488 ms) for the experimental materials and 11% (1,973 ms) for the filler materials. Conditionalizing response times for the test words on whether the answer to the question was correct did not affect the pattern of the results.

ANOVAs of error rates showed main effects of test word, $F_1(1, 35) = 13.2$ and $F_2(1, 56) = 12.0$, and test positions, $F_1(1, 35) = 7.3$ and $F_2(1, 56) = 6.2$. The Fs for the interaction were less than 1, and the standard error was 1.2%.

Experiment 6. If the increase in response time for the nonreferent test words that was observed in Experiment 5 was due to suppression of the nonreferent, then we should not observe the same increase in response time for the control test word. In fact, however, the increase was actually somewhat larger. Response times for the control word increased from Test Position 1 to Test Position 3 more than did response times for the referent test word, and this interaction was

Table 5
Results of Experiments 5, 6, and 7: Response Times (RTs) and Error Rates (ERs) on Test Words

	Test position			
	1		3	
Test word	RT (ms)	ER (%)	RT (ms)	ER (%)
	Experiment 5[a]			
Referent	1,043	9	1,054	12
Nonreferent	993	4	1,067	8
	Experiment 6[b]			
Referent	1,106	8	1,128	11
Control	1,082	4	1,211	9
	Experiment 7[c]			
Referent	880	5	908	7
Nonreferent	909	5	878	3
Control	999	14	1,073	16

[a] Response time and error rate for negative fillers are 1,239 ms and 8%, respectively.
[b] Response time and error rate for positive fillers are 1,080 ms and 8%, respectively, and for negative fillers, 1,142 ms and 5%, respectively.
[c] Response time and error rate for positive fillers are 1,121 ms and 14%, respectively, and for negative fillers, 1,289 ms and 7%, respectively.

significant, $F_1(1, 23) = 10.0$ and $F_2(1, 39) = 4.1$. There was also a main effect of test position, $F_1(1, 23) = 9.0$ and $F_2(1, 39) = 11.5$. The Fs for the effect of test word were less than 2.9. The standard error of the response time means was 18 ms. There were more errors at the third test position than the first, $F_1(1, 23) = 5.0$ and $F_2(1, 39) = 4.7$. Other Fs in the errors analyses were less than 3.1. The standard error was 1.5%. Subjects were accurate in their responses to the "who did it" questions, with only 3% errors (1,571 ms) on the experimental items and 9% errors (2,097 ms) on the fillers.

Discussion

In contrast to Experiments 1 through 4, the results of Experiment 5 showed a relative advantage for referents over nonreferents. We attribute this advantage to pronominal processing that occurred because subjects were encouraged by the experimental procedure to identify the pronoun's referent during reading. Our interpretation of these results is that, with the same set of materials, processing can be exclusively automatic, leaving the pronoun unresolved (as in Experiments 1–4), or it may also include slower, strategic processes that allow the unique identification of the pronoun's referent (Experiment 5).

The results of Experiment 6 suggest reformulation of the suppression hypothesis proposed by Gernsbacher (1989). Although we replicated the result that nonreferent response times were slower after the pronoun, responses for control words were slowed at least as much. This could be because suppression affects all entities in the discourse model (other than the referent). Alternatively, it could be that all test words are slowed because of end-of-sentence processing, and the underlying mechanism for the referent–nonreferent difference is actually facilitation for the referent. Currently, this issue cannot be resolved, and further research is needed.

Experiment 7

In Experiment 5, strategic processing was encouraged by providing motivation to identify pronominal referents, by providing a sufficiently slow rate of presentation for the text, and by making the task relatively easy with only one pronoun to be identified in a one-sentence text. The result was that referents showed a relative advantage over nonreferents in contrast to Experiments 1 through 4. It might be thought that the only one of the three factors that actually contributed to the difference in findings between the first four experiments and Experiment 5 was the speed of presentation. Automatic processes of identification for the pronominal referents in the experimental texts might require more time than was available at the 250-ms per word rate used in the first four experiments. According to this hypothesis, simply slowing the rate of presentation should lead to an advantage for referents over nonreferents.

In Experiment 7, we tested this hypothesis by replicating Experiment 2 with a slow rate of presentation. The materials were the same multisentence texts used in Experiment 2, but the rate was slowed to 450 ms per word plus 16⅔ ms multiplied by the number of letters in the word, the same rate used in Experiment 5.

Method

The same three-sentence materials were used in this experiment as in Experiment 2. After pilot subjects, we decided not to test the rate of presentation factor alone but to test the rate factor together with the motivational factor. Therefore, for each text, there was a test question that required identification of the referent of a pronoun in the final sentence of the text. These questions asked "who did" one of the actions in the final sentence of the text and were the same questions used in Experiments 5 and 6. With this experiment, both the rate and motivation factors were tested: If the results failed to show facilitation of referent over nonreferent test words, then both factors could be eliminated as being solely responsible for inducing a specific strategy of pronoun identification.

Except for the rate of presentation of the texts, the "who did it" questions, and omission of the "too slow" message for slow responses, the procedure and design were the same as for Experiment 2. Specifically, there were two factors: test position (Position 1 or Position 3) and test word (referent, nonreferent, and control). The "who did it" questions were presented in the same way as in Experiments 5 and 6. There were 24 subjects from the same population as Experiments 1 through 6.

Results and Discussion

The data show clearly that, in this experiment, slowing the rate of presentation did not lead to an advantage for the referent over the nonreferent after reading of the pronoun. There was no advantage even though the rate was extremely slow, and comprehension questions asked for specific knowledge of the pronoun's referent.

The results are shown in Table 5. As a function of test position, the relative referent and nonreferent response times did not change significantly. The interaction between test position and test word was significant with subjects as the random variable because of the increase in response times to the control test word, $F_1(2, 46) = 3.9$, but this interaction was not significant in the items analysis, $F_2(2, 118) = 1.9$. There was a main effect of test word, significant in both analyses, $F_1(2, 46) = 16.0$ and $F_2(2, 118) = 17.2$. The control test words had slower response times than the other test words ($Fs > 15$). The effect of test position was marginally significant in the subjects analysis, $F_1(1, 23) = 3.4$, but not in the items analysis, $F_2(1, 59) = 1.9$. The standard error of the mean was 28 ms. The only significant effect for errors was that of test word, $F_1(2, 46) = 10.3$ and $F_2(2, 118) = 13.2$, with a standard error of 1.7%. The control test words had more errors than the other test words ($Fs > 11$). Correct responses for the comprehension questions on the experimental texts averaged 1,321 ms with 7% errors and on the filler texts, 1,620 ms with 5% errors.

Why did subjects appear to identify the pronominal referent in Experiment 5 but not in Experiment 7? The procedural differences in the two experiments are the number of sentences in the texts—one sentence in Experiment 5 compared with three in Experiment 7—and the inclusion of the control test words in Experiment 7. However, these differences, especially the first, are critical. With only one sentence, a reader can easily anticipate exactly when the pronoun will occur and exactly what the comprehension question must be. Also, in Experiment 5, all the test words were names so that it would make sense for readers to keep track carefully of who did what. In Experiment 7, it would theoretically be possible to anticipate exactly when the critical pronoun would occur and exactly what the comprehension question would be, but to do this readers would have to count the sentences as they read to know which was the third and then anticipate the comprehension question. In short, Experiment 7 reduces the ability of subjects to engage in strategic processing compared with Experiment 5.

Experiments 8 and 9

In Experiment 5, we were able to show that subjects could, under the appropriate conditions, identify the intended referents for the pronouns in the experimental sentences. However, we are still left with a null result for the procedure used in Experiments 1 through 4 for which we claim that fast, automatic processing leaves the pronoun unresolved. In Experiments 8 and 9, we show that this procedure does allow identification of the referent for another type of anaphor. That at least one kind of referent is identified shows that the 250-ms per word reading rate used in our experiments is not so fast that it prevents the comprehension of all kinds of implicit information.

The anaphors we used were the nominals from studies by Dell et al. (1983). An example is shown in Table 6. In the first version of the fourth sentence, the nominal *the criminal* is intended to refer to the burglar mentioned in the first sentence. In the other version, the subject noun phrase is not intended to refer to the burglar. Dell et al., using the same procedure as in Experiments 1 through 4 in this article, showed that when the referent was presented as a test word after the anaphor, response time was facilitated relative to when it was presented after the control noun phrase. From this result (and appropriate control conditions), Dell et al. concluded that comprehension of the anaphor involved identification of its referent. Dell et al. also tested an associate of the referent (e.g., *garage* for the text in Table 6); this test word had occurred in the first sentence of the text with the referent. When this word was presented immediately after the anaphor, it also showed facilitated response time relative to the control condition, indicating that processing of the anaphor increased the accessibility not only of the referent but also of concepts associated with the referent.

Table 6
An Example of the Paragraphs Used in Experiments 8 and 9

Sentence 1: A burglar surveyed the garage set back from the street.
Sentence 2: Several milk bottles were piled at the curb.
Sentence 3: The banker and her husband were on vacation.
Sentence 4 (version 1, anaphor):
The criminal slipped$_2$ away from the streetlamp.$_3$
Sentence 4 (version 2, no anaphor):
A cat slipped$_2$ away from the streetlamp.$_3$
Test words
Referent: burglar
Associate of referent: garage

In Experiment 8, we mixed the texts of the pronominal anaphors from Experiments 1 through 7 with the texts of nominal anaphors used by Dell et al. (1983) and tested the referent of the nominal (e.g., *burglar*). Experiment 9 was similar except that we tested for both the referent and the associated concept from the first sentence (e.g., *garage*). The prediction was that results for both sets of texts would replicate what had been found previously: Relative facilitation would be observed with the nominals (and the concepts associated with them) but not with the pronouns.

Method

Materials and procedure. There were two sets of experimental materials. The first set was 32 of the experimental items from Experiments 1 and 2. The second was 32 of the items used by Dell et al. (1983) shown by example in Table 6. For each of these items, the first sentence introduced a main character, and that character was not referred to again in the second or third sentences. There were two versions of the fourth sentence: In one, the first noun of the sentence was an anaphor that referred to the character, and in the other—the control version—the first noun was some other concept unrelated to the character. Except for the first noun and its determiner, the two versions of the fourth sentence were identical. The texts minus the fourth sentences averaged 26 words in length. The fourth sentences averaged 8.4 words in length. There were two test words for these texts: the noun that referred to the main character introduced in the first sentence (*burglar*) and a word associated with the main character in the first sentence (*garage*).

There were two sets of filler texts. One set was a subset of 44 texts from the fillers used in Experiments 1 to 7. The other was a set of 27 filler texts from the Dell et al. experiment. These averaged 40 words in length and five lines on the CRT screen. Of these texts, 23 had

negative test words and 4 had positive test words. The procedure and comprehension questions were the same as in Experiment 2.

Subjects and design. For both Experiments 8 and 9, there were two variables for the pronoun materials: The test word was either the referent of the pronoun or the nonreferent, and the test word was presented at either Test Position 2 or 3. For Experiment 8, there were also two variables for the nominal anaphor materials: The fourth sentence was presented either in the version that referred to the main character or in the control version, and the referent test word was presented either after the word following the anaphor or at the end of the fourth sentence (Positions 2 and 3 in the table). For Experiment 9, the nominal anaphor materials were also presented in the two versions, but the second variable was different: The test word was either the referent or the word associated with the referent from the first sentence of the text. The test word was always presented after the word following the anaphor (Position 2). For both sets of materials in both experiments, the four conditions were combined in a Latin square with sets of items (8 per set) and groups of subjects. In Experiment 8 there were 16 subjects, and in Experiment 9 there were 44 subjects, all from the same population as in Experiments 1 and 2.

Results and Discussion

For the pronoun materials, once again referent and nonreferent response times were not differentially affected by test position (see Table 7). In Experiment 9, the nonreferent test word responses were faster than the referent test word responses, $F_1(1, 43) = 6.1$ and $F_2(1, 31) = 4.0$, but this difference was not significant in Experiment 8 ($Fs < 1.4$). Responses were slower at Test Position 3 than at Test Position 2 in Experiment 8, $F_1(1, 15) = 7.2$ and $F_2(1, 31) = 5.9$, but not in Experiment 9 ($Fs < 1.1$). The two variables did not interact significantly in either experiment ($Fs < 2.4$). The standard error of the response time means was 16 ms in Experiment 8 and 8 ms in Experiment 9. There were no significant differences in error rates in Experiment 8 (the standard error was 1.6%), but in Experiment 9, there were significantly more errors on the referent than the nonreferent, $F_1(1, 43) = 11.4$ and $F_2(1, 31) = 15.1$; the standard error was 1.0%.

In contrast, the nominal anaphors showed significant facilitation for their referents and for concepts associated with their referents (see Table 7). In general, the pattern of data for the nominal anaphors closely replicates the pattern obtained by Dell et al. (1983).

In Experiment 8, when the final sentence mentioned the anaphor, the responses to the referent test word were faster than when the final sentence mentioned the control word, $F_1(1, 15) = 17.2$ and $F_2(1, 28) = 4.5$. This facilitation did not interact significantly with test position ($Fs < 1.6$). The effect of test position was significant, $F_1(1, 15) = 11.7$ and $F_2(1, 28) = 4.6$. The standard error of the response time means was 14 ms. There were no significant effects on error rates ($Fs < 2.4$), and the standard error was 3.1%.

In Experiment 9, when the final sentence mentioned the anaphor, then responses to both the referent test word and the associate test word were faster than when the final sentence mentioned the control word, $F_1(1, 43) = 15.5$ and $F_2(1, 31) = 10.2$. Referent response times were faster than associate response times, $F_1(1, 43) = 10.4$ and $F_2(1, 31) = 8.9$. The interaction of the two variables was not significant ($Fs < 1.2$). The standard error of the means was 11 ms. By planned test, response times for the associate test words were faster when the final sentence mentioned the anaphor than when it did not, $F_1(1, 43) = 4.6$ and $F_2(1, 31) = 4.3$. There were more errors on the associate test words than the referent test words, $F_1(1, 43) = 38.7$ and $F_2(1, 31) = 18.1$. No other effects of error rates were significant, with a standard error of 2.3%.

Table 7
Results of Experiments 8 and 9: Response Times (RTs) and Error Rates (ERs) on Test Words

Variable	Test position 2 RT (ms)	Test position 2 ER (%)	Test position 3 RT (ms)	Test position 3 ER (%)
Experiment 8: pronoun materials				
Test word				
Referent	682	9	707	7
Nonreferent	658	3	707	5
Experiment 8: anaphor materials[a]				
Fourth sentence				
Anaphor version	748	13	786	15
Control version	770	21	850	15
Experiment 9: pronoun materials				
Test word				
Referent	707	9	711	11
Nonreferent	683	4	708	5
Experiment 9: anaphor materials[b]				
	Referent test word		Associate test word	
Fourth sentence				
Anaphor version	726	18	774	31
Control version	786	19	811	34

[a] Response time and error rate for positive fillers are 866 ms and 21%, respectively, and for negative fillers, 850 ms and 6% respectively. [b] Response time and error rate for positive fillers are 804 ms and 24%, respectively, and for negative fillers, 813 ms and 12%, respectively.

In Experiment 8, for true test statements, the mean response time was 1,788 ms (11% errors) and for false test statements, 1,681 ms (18% errors). In Experiment 9, true test statements averaged 2,199 ms (11% errors), and false statements averaged 2,074 ms (20% errors).

The results of these experiments were exactly as predicted: At a relatively fast presentation rate, in the absence of comprehension questions designed to motivate identification of anaphoric referents during reading, recognition responses for referents were facilitated for the nominal anaphors but not for the pronominal anaphors in the experimental materials. Our interpretation of these results is that the referent of a nominal anaphor was uniquely identified during reading but that the referent of a pronoun was not. We interpret the results for the nominal anaphors as showing referent identification in light of several converging pieces of data. First, the relative facilitation for the referent test word (*burglar*) might be due solely to the semantic relation with the anaphor (*criminal*), but this cannot be the case because the associated

test word (*garage*) also shows facilitation. Second, the relative inhibition in the control condition might be due to the introduction of a new concept (*cat*), but such inhibition would also be expected to appear on responses to test words other than the referent and the associate, and it did not (Dell et al., 1983).

There are several reasons why the referent of a nominal might have been identified under the same conditions in which the referent of a pronoun was not. One possibility is that the nominal was a word semantically related to its referent, and the pronoun was not (except with respect to gender). It has been suggested that semantic relatedness is a general aid to inference processes because semantic information is easily and quickly available during processing (McKoon & Ratcliff, 1989a, 1989b). Another possibility, suggested by Gernsbacher (1989), is that the nominal is more specific than the pronoun. The nominal might contain such specific information that, in the relevant discourse, no discourse entity other than the intended referent matches the nominal to any degree at all. For example, the nominal *criminal* may contain information specific enough that only *burglar* and no other entities in the discourse (such as *banker*) match the nominal to any degree. Finally, it could be that the nominal provides a second repetition of its referent entity in a way that a pronoun does not (i.e., the nominal may add information about the entity to its discourse representation). Obviously, more research is needed to distinguish among these possibilities. However, the contrast between processing of the nominal and pronominal anaphors does make clear one point: It makes little sense to ask whether a reader understands a discourse overall and in general; under the same contextual conditions, a reader may identify a unique referent for one kind of anaphor but not for another. Empirical investigations of discourse comprehension can only be made up of tests of the many individual processes that may or may not, depending on experimental and contextual conditions, constitute comprehension.

General Discussion

Our conclusion that people do not always identify a unique referent for a pronoun, although consistent with current discourse models, stands in contrast with previous work. Hence, we should consider the reasons we have come to a different conclusion than have previous researchers. In empirical terms, our conclusion was different because our procedures for testing pronoun resolution were different. More important, our procedures were motivated by a different theoretical view than has previously guided psycholinguistic research on pronoun resolution. Representing a text as a discourse model entails consideration of the relative accessibilities of the entities in the model. In this context, a pronoun is viewed as a cue to one or more of the entities. This "pronoun as cue" notion naturally suggests the parallel access matching process assumed by current memory models. These models distinguish automatic processes from strategic processes, and our experiments were designed to examine the identification of referents as an automatic process.

To move readers away from special strategies brought about by task demands that might have occurred in previous studies, we introduced three major methodological modifications. First, our texts were presented at a rate of 250 ms per word compared with an average of about 500 ms per word in some previous work (e.g., Gernsbacher, 1989). Second, our texts contained three sentences (compared with the single sentence used by other researchers) and multiple test points throughout the texts. Third, comprehension questions presented after the texts tested a variety of kinds of information in our experiments, whereas previous experiments often asked specifically for information about the intended referents of pronouns. These three changes were introduced to discourage subjects from engaging in strategic processes during reading to identify the pronouns. Avoiding strategic processing is important because of the nature of the question we are studying. We are not asking whether people can uniquely identify referents for pronouns but whether they automatically do so during comprehension and whether they always do so. It is clear that readers are capable of uniquely identifying pronominal referents; what is less clear is whether it is always a part of the processes of comprehension.

In our efforts to eliminate strategic processing of pronouns, we might have used reading times so fast that readers engaged in no processing at all. However, the reading rates that we used were appropriate for our subject population. As Experiments 8 and 9 demonstrate, the same subjects reading at the same speed did appear to resolve other types of anaphors. Furthermore, a slower reading rate by itself was not sufficient to guarantee resolution of the pronominal anaphors in our experiments. We found facilitation of pronominal referents over nonreferents only when the slow rate was combined with motivation to identify uniquely the referents and with procedures that made the identification task relatively easy.

Throughout the experiments described in this article, the distinction between automatic and strategic processes was used to guide choices of experimental variables. The application of the automatic–strategic distinction to reading processes is not straightforward. However, in some sense, the distinction must apply; in reading, as in other cognitive tasks, there are processes that are slow and invoked to meet specific contextual demands, and there are processes that are faster and less constrained by a particular context (McKoon & Ratcliff, in press). In addition, the distinction can usefully be applied even though there are many open questions, such as whether the distinction represents a dichotomy or a continuum and how the particular variables and results found for automatic processes in other domains can be applied to reading.

The usefulness of the distinction is demonstrated by the outcomes of the experiments. The distinction suggests experiments designed to move processing away from strategies adopted for a particular experimental task. Such strategies are generally assumed to be slower and more influenced by specific task demands than automatic processes, and so, to eliminate them, reading and response rates were speeded and task demands specific to anaphoric identification were eliminated. Clearly, if there is a distinction (or a continuum) between automatic and strategic processes in reading, these procedural changes should represent a move toward the automatic. That

these procedural changes brought about substantial changes in the results of the experiments gives support to the utility of the automatic–strategic distinction in investigations of reading. The support for the automatic–strategic distinction is particularly impressive because it is only this notion, and not other current views, that would have guided us to address these questions in these ways. Previous views would have labeled anaphor resolution a necessary part of reading and would not have suggested that anaphor resolution would depend on manipulations of task demands and rate of processing except as part of a general failure in processing. Thus, the automatic–strategic distinction led to experiments that would otherwise not have been conducted and yet demonstrate important and unexpected boundary conditions on a fundamental aspect of reading.

By adopting the procedural manipulations suggested by an automatic–strategic distinction, we showed that the advantage in testing for the referent of a pronoun over a nonreferent could be eliminated. We interpret this result as indicating that the referent did not enjoy a processing advantage during reading over the nonreferent and as providing support for the discourse model framework proposed early in this article. According to this framework, the referent has no advantage because it was not uniquely identified as the referent of the pronoun.

An alternative interpretation of the experimental data is that the referent of the pronoun was, in fact, identified but that this identification process did not lead to an advantage on the recognition test. One obvious possible reason for this would be that responses on the recognition test were at ceiling, but responses in Experiment 7 were relatively slow and yet still showed no facilitation for the referent. Other reasons that recognition might fail to show the consequences of identification would be less plausible. For identification, the comprehension system must by some mechanism choose between two possible referents (e.g., *John* and *Mary*) on the basis of gender. Then, after making a choice, the system must either create a new token of the referent to which to attach the information given with the pronoun or attach the new information to the referent directly. Either way, new information about the referent would be encoded in memory. Thus, resolving the pronoun would entail both choosing the referent and encoding additional information about it, and this processing would have to be assumed to leave no consequences detectable in the recognition test.

Furthermore, assuming that identification leaves no traces detectable by recognition probes runs counter to all current accounts of on-line recognition testing (Chang, 1980; Corbett & Chang, 1983; van Dijk & Kintsch, 1983; Gernsbacher, 1989; MacDonald & MacWhinney, 1990; McKoon & Ratcliff, 1986, 1990). The effects of a variety of similar on-line processes are frequently observed on recognition tests. Experiments 8 and 9 present one example in which the effects of processing a noun anaphor are observed. Other examples include the processing of explicitly mentioned entities (Caplan, 1972; Jarvella, 1971), the processing of pronouns in object case (*him, her*; Cloitre & Bever, 1989), the processing of empty syntactic traces (Bever & McElree, 1988), the processing of pronouns that refer to entities introduced in previous sentences (McKoon et al., 1991), and the processing of verbs that take implicit instruments (McKoon & Ratcliff, 1981). Collectively, these examples overlap with the experiments in this article in many ways. The distance, in terms of number of words, between pronoun and antecedent is about the same in the current experiments as in the experiments of McKoon and Ratcliff (1980; two sentence texts), McKoon et al. (1991), and Bever and McElree (1988). The type of pronoun (subject of its clause) is the same as in McKoon et al. (1991). The use of the referent as test word is the same as in McKoon and Ratcliff (1980, 1981) and McKoon et al. (1991). In all of these cases, processing facilitates recognition responses for the referenced entity. The only apparent difference in the experiments reported here is the presence of two possible referents for the pronoun.

We believe that the more plausible interpretation of the data is that the referent of the pronoun is not uniquely identified; instead, information given with the pronoun is attached to the current focus of attention, which includes both potential referents. One way that this could come about is suggested by current discourse models.

Discourse models have been proposed to describe the information that is used to establish coreference among discourse entities. For a discourse model, the important variables that distinguish entities are their relative accessibilities and their semantic (and possibly pragmatic) content. Variables such as recency of mention in the text and syntactic category are relevant only in their indirect effects on accessibility. More directly relevant are variables such as the relation between an entity and the discourse topic (Kintsch, 1974; McKoon et al., 1991), and variables that affect the semantic overlap among the entities. For example, reference processes can be affected by the degree of semantic association between an anaphor and its possible referents (Corbett, 1984).

A model of discourse processing in which pronouns are matched against all entities in memory suggests that there may be some contexts in which no single discourse entity matches sufficiently better than all others to be selected as the referent. In the experiments presented here, it appears that we have found one set of contextual factors in which that happens. However, we would be ill-advised to conclude that this situation is the general one or even a common one. We have only studied texts with two relatively indistinguishable characters, one of whom is referred to by a pronoun. In fact, much of the research on pronoun comprehension consists of studies using materials that fit the same general description (Chang, 1980; Corbett & Chang, 1983; Ehrlich, 1980; Garnham & Oakhill, 1985; Gernsbacher, 1989; MacDonald & MacWhinney, 1990). However, this is far from the situation in which we would expect pronouns to occur most often in natural discourses. Normally, when a pronoun is used, one discourse entity is already in the focus of attention (Brennan, 1989; Chafe, 1974; Fletcher, 1984). It seems that we have been studying pronouns outside their natural habitat.

Moreover, it may be that pronouns have been studied for the wrong reasons. In past studies, the problem has been to find out how the processing system uses a pronoun to find its referent. Phrasing the question this way puts the burden on processes driven by the pronoun. However, the appropriate

question may be to ask not what the pronoun does for the discourse but what the discourse does for the pronoun. When the discourse has only one entity in the focus of attention at the time the pronoun is encountered, then it may be that essentially no processing is required for the pronoun. It may be that information predicated of the pronoun is attached to the focused entity by means of an attachment process that is simple, automatic, and demanding of little processing capacity. If this is the case, then pronouns are interesting not because of the effort they require but precisely because of the effort they do not require.

We suggest that pronouns are most frequently dealt with by an automatic process of attaching their propositions to the current discourse focus and the propositions relevant to it. It follows that the referent of a pronoun will be completely and correctly identified only if the discourse focus contains the uniquely correct referent. If the focus contains more than one possible referent, as in our experiments, then the propositions of the pronoun are attached equally to all the focused entities. In effect, the automatic processes of comprehension treat the new information simply as predicated of the entity or entities in focus. This processing may not always result in the correct representation of a text in some ultimate sense for some particular set of experimental materials; instead, the processing system is designed to operate under stringent time constraints to provide a useful understanding of natural discourse. Of course, if comprehenders have special motivation and enough time to resolve a pronoun reference more completely, they can engage in further strategic processing to do so.

Viewing pronouns as cues to discourse entities is consistent with three phenomena previously pointed out by other researchers: pronouns that refer using demonstration, "unheralded pronouns" (see Gerrig, 1986), and "conceptual anaphors" (see Gernsbacher, 1986). First, if a discourse is about some unique but linguistically unspecified referent, then the lack of linguistic specification does not necessarily impede comprehension. This has been documented by Clark, Schreuder, and Buttrick (1983), who noted that linguistically underdetermined noun phrases can be used to refer to unstated entities that are nevertheless in common ground. For example, the assertion, "They publish gossip," uttered while pointing to a newspaper, refers successfully to the newspaper's publishers. Theories of pronoun resolution that conceive of pronouns as triggering a search for a linguistic referent cannot explain this example. In contrast, such examples fit naturally into a theory such as ours that views a pronoun as a cue relevant to some entity in the comprehender's discourse model. Reference by demonstration may not be understood by entirely automatic processes, yet whatever the processing the result is resolution of an anaphor as referring to a focused entity.

Unheralded pronouns (Gerrig, 1986) are also consistent with the pronoun-as-cue framework. An unheralded pronoun refers to an entity not previously referred to either linguistically or deictically. Consider the following conversation between two popular music buffs:

Penny: Do you have a CD of "Abbey Road?"

Cindy: Oh, sure. I have CDs of all their stuff.

For these speakers (and perhaps for some readers of this article), the pronoun *their* refers successfully to the Beatles. The pronoun-as-cue framework can account for this example by assuming that the album title brings the concept of the Beatles into the comprehender's discourse model, making it sufficiently accessible for the pronoun to be uttered felicitously.

The third phenomenon that can be understood from the pronoun-as-cue framework is what Gernsbacher (1986) referred to as conceptual anaphora. Normally, pronouns in English agree in number with their referents. However, Gernsbacher noted exceptions such as the following:

I need a plate. Where do you keep them?

For examples such as this, in which the speaker is referring to an unspecified member of a set of items that all will serve equally well, the plural pronoun is rated as being more natural and is comprehended more quickly than the singular pronoun. Again, a traditional view of pronoun resolution would have difficulty explaining this phenomenon. However, the pronoun-as-cue framework simply assumes that the speaker's use of the word *plate* focuses the comprehender's attention on all of his or her plates. In this context, it is natural to refer to the entire set of plates using a pronoun.

As illustrated by these examples, the pronoun-as-cue framework encourages us to examine the larger discourse context to understand how pronouns are used felicitously. Pronouns are viewed as doing little more than signaling the comprehender that the speaker (or author) is referring to whatever entity is in the current focus of attention within the constraints imposed by syntax. In this view, the interesting questions for research concern how various discourse elements are deployed to help the speaker (or author) and comprehender share the same focus of attention. To answer these questions, it is necessary to look beyond the literal text of a discourse.

References

Bever, T., & McElree, B. (1988). Empty categories access their antecedents during comprehension. *Linguistic Inquiry, 19,* 35–43.

Brennan, S. E. (1989). *Centering attention in discourse.* Unpublished manuscript, Stanford University.

Caplan, D. (1972). Clause boundaries and recognition latencies for words in sentences. *Perception and Psychophysics, 12,* 73–76.

Chafe, W. L. (1974). Language and consciousness. *Language, 50,* 111–133.

Chafe, W. L. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. N. Li (Ed.), *Subject and topic* (pp. 25–55). New York: Academic Press.

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

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 10–63). Cambridge, England: Cambridge University Press.

Clark, H. H., Schreuder, R., & Buttrick, S. (1983). Common ground and the understanding of demonstrative reference. *Journal of Verbal Learning and Verbal Behavior, 22,* 245–258.

Clark, H. H., & Sengul, C. J. (1979). In search of referents for nouns and pronouns. *Memory & Cognition, 7,* 35–41.

Clifton, C., Jr., & Ferreira, F. (1987). Discourse structure and anaphora: Some experimental results. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 635–654). Hillsdale, NJ: Erlbaum.

Cloitre, M., & Bever, T. G. (1989). Linguistic anaphors, levels of representation, and discourse. *Language and Cognitive Processes, 3,* 293–322.

Corbett, A. T. (1984). Prenominal adjectives and the disambiguation of anaphoric nouns. *Journal of Verbal Learning and Verbal Behavior, 23,* 683–695.

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

Dell, G. S., McKoon, G., & Ratcliff, R. (1983). The activation of antecedent information during the processing of anaphoric reference in reading. *Journal of Verbal Learning and Verbal Behavior, 22,* 121–132.

Dijk, T. A. van, & Kintsch, W. (1983). *Strategies of discourse comprehension.* New York: Academic Press.

Ehrlich, K. (1980). Comprehension of pronouns. *Quarterly Journal of Experimental Psychology, 32,* 247–255.

Ehrlich, K. (1983). Eye movements in pronoun assignment: A study of sentence integration. In K. Rayner (Ed.), *Eye movements in reading: Perceptual and language processes* (pp. 253–272). New York: Academic Press.

Ehrlich, S. F., & Rayner, K. (1983). Pronoun assignment and semantic integration during reading: Eye movements and immediacy of processing. *Journal of Verbal Learning and Verbal Behavior, 22,* 75–87.

Fletcher, C. (1984). Markedness and topic continuity in discourse processing. *Journal of Verbal Learning and Verbal Behavior, 23,* 487–493.

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

Garvey, C., Caramazza, A., & Yates, J. (1976). Factors influencing assignment of pronoun antecedents. *Cognition, 3,* 227–243.

Gernsbacher, M. (1986). The comprehension of conceptual anaphora in discourse. *Proceedings of the eighth annual conference of the Cognitive Science Society* (pp. 110–125).

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

Gernsbacher, M., & Hargreaves, D. (1988). Accessing sentence participants: The advantage of first mention. *Journal of Memory and Language, 27,* 699–717.

Gernsbacher, M., Hargreaves, D., & Beeman, M. (1989). Building and accessing clausal representations: The advantage of first mention versus the advantage of clause recency. *Journal of Memory and Language, 28,* 735–755.

Gerrig, R. J. (1986). Process models and pragmatics. In N. E. Sharkey (Ed.), *Advances in cognitive science* (pp. 23–42). Chichester, England: Ellis Horwood.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1–67.

Givon, T. (1976). Topic, pronoun, and grammatical agreement. In C. N. Li (Ed.), *Subject and topic* (pp. 149–188). New York: Academic Press.

Grosz, B. (1981). Focusing and description in natural language dialogues. In A. K. Joshi, B. L. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 84–105). Cambridge, England: Cambridge University Press.

Grosz, B., Joshi, A. K., & Weinstein, S. (1983). Providing a unified account of definite noun phrases in discourse. In *Proceedings of the 21st annual meeting of the Association of Computational Linguistics* (pp. 44–50).

Grosz, B., & Sidner, C. (1986). Attention, intentions and the structure of discourse. *Computational Linguistics, 12,* 175–204.

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

Hawkins, J. A. (1977). The pragmatics of definiteness I and II. *Linguistiche Berichte, 47,* 1–27.

Hintzman, D. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95,* 528–551.

Hobbs, J. (1978). Resolving pronoun references. *Lingua, 44,* 311–338.

Hudson, S. B., Tanenhaus, M. K., & Dell, G. S. (1986). The effect of the discourse center on the local coherence of a discourse. In *Proceedings of the eighth annual conference of the Cognitive Science Society* (pp. 96–101).

Jarvella, R. J. (1971). Syntactic processing of connected speech. *Journal of Verbal Learning and Verbal Behavior, 10,* 409–416.

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

Just, M. A., & Carpenter, P. A. (1987). *The psychology of reading and language comprehension.* Boston, MA: Allyn & Bacon.

Kintsch, W. (1974). *The representation of meaning in memory.* Hillsdale, NJ: Erlbaum.

MacDonald, M. C., & MacWhinney, B. (1990). Measuring inhibition and facilitation from pronouns. *Journal of Memory and Language, 29,* 469–492.

Matthews, A., & Chodorow, M. (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. (1980). The comprehension processes and memory structures involved in anaphoric reference. *Journal of Verbal Learning and Verbal Behavior, 19,* 668–682.

McKoon, G., & Ratcliff, R. (1981). The comprehension processes and memory structures involved in instrumental inference. *Journal of Verbal Learning and Verbal Behavior, 20,* 671–682.

McKoon, G., & Ratcliff, R. (1984). Priming and on-line text comprehension. In D. E. Kieras & M. A. Just (Eds.), *New methods in reading comprehension research* (pp. 119–128). Hillsdale, NJ: Erlbaum.

McKoon, G., & Ratcliff, R. (1986). Inferences about predictable events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 82–91.

McKoon, G., & Ratcliff, R. (1989a). Inferences about contextually-defined categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 1134–1146.

McKoon, G., & Ratcliff, R. (1989b). Semantic association and elaborative inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 326–338.

McKoon, G., & Ratcliff, R. (1990). Textual inferences: Models and measures. In D. A. Balota, G. B. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 403–421). Hillsdale, NJ: Erlbaum.

McKoon, G., & Ratcliff, R. (in press). Inference during reading. *Psychological Review.*

McKoon, G., Ward, G., Ratcliff, R., & Sproat, R. (1991). *Morpho-syntactic and pragmatic factors affecting the accessibility of discourse entities.* Manuscript submitted for publication.

Murdock, B. B. (1974). *Human memory: Theory and data.* Potomac, MD: Erlbaum.

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review, 89,* 609–626.

Neely, J. H. (1977). Semantic priming and retrieval from lexical

memory: Roles of inhibitionless spreading activation and limited capacity intention. *Journal of Experimental Psychology: General, 106,* 226–254.

Nicol, J., & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research, 18,* 5–20.

Posner, M. I., & Snyder, C. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition* (pp. 55–85). Hillsdale, NJ: Erlbaum.

Prince, E. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical pragmatics* (pp. 223–255). New York: Academic Press.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85,* 59–108.

Ratcliff, R., & McKoon, G. (1981). Automatic and strategic priming in recognition. *Journal of Verbal Learning and Verbal Behavior, 20,* 204–215.

Rayner, K. (1978). Eye movements in reading and information processing. *Psychological Bulletin, 85,* 618–660.

Reinhart, T. (1982). *Pragmatics and linguistics: An analysis of sentence topics.* Bloomington, IN: Indiana University Linguistics Club.

Rothkopf, E., Koether, M., & Billington, M. (1988). *Why are certain sentence constructions mnemonically robust for modifiers?* (Technical memorandum). AT&T Bell Laboratories, Murray Hill, NJ.

Sidner, C. (1983a). Focusing in the comprehension of definite anaphora. In M. Brady & R. Berwick (Eds.), *Computational models of discourse* (pp. 267–330). Cambridge, MA: MIT Press.

Sidner, C. (1983b). Focusing and discourse. *Discourse Processes, 6,* 107–130.

Tanenhaus, M. K., Carlson, G. N., & Trueswell, J. C. (1989). The role of thematic structures in interpretation and parsing. *Language and Cognitive Processes, 4,* 211–234.

Webber, B. (1983). So what can we talk about now? In M. Brady & R. Berwick (Eds.), *Computational models of discourse* (pp. 331–371). Cambridge, MA: MIT Press.

Yule, G. (1982). Interpreting anaphora without identifying reference. *Journal of Semantics, 1,* 315–322.

Received May 1, 1991
Revision received August 2, 1991
Accepted August 22, 1991 ■