Accessibility of Potential Referents Following Categorical Anaphors

Jennifer Wiley
University of Illinois at Chicago

Robert A. Mason
Carnegie Mellon University

Jerome L. Myers
University of Massachusetts at Amherst

In 4 experiments, the authors investigated the accessibility of potential antecedents that were either referents or nonreferents of categorical anaphors using several methodologies. In 2 preliminary experiments, using an immediate probe recognition task, the authors replicated previous findings of inhibition for nonreferents following categorical anaphors. In Experiments 1–3, naming times to nonreferent probes were inhibited, but only at a delay and at the end of the sentence. Eye movements were monitored in Experiment 4, and reading times were found to be slower in end-of-sentence regions following the remention of the nonreferent but not on the nonreferent itself. These results suggest that decreases in nonreferent accessibility are due to postanaphoric processes, such as integrating the nonreferent into the discourse structure, and are not due to more immediate changes in activation occurring as part of the resolution process. Further, some inhibition effects may result from the use of probe tasks that require decisions.

Comprehending a discourse requires readers to integrate the current linguistic input and associated elements of their world knowledge into the evolving discourse representation. This representation has generally been viewed as an associative network, a collection of nodes representing concepts and propositions, and interconnected by strength-valued links (e.g., Gernsbacher, 1990; Kintsch, 1988; van Dijk & Kintsch, 1983). Within such models, any linguistic input is viewed as an initiating signal, serving to alter the distribution of activation on the nodes in the discourse representation. Furthermore, at any given time, the reader’s comprehension of the text corresponds to the pattern of activation on the nodes. A major theoretical question is how the linguistic signal alters the pattern of activation. One answer to this question has been provided by Gernsbacher (1990, 1997a, 1997b). According to her structure building model, linguistic input activates memory nodes that transmit processing signals to enhance the activation of other relevant nodes and to suppress the activation of other irrelevant nodes, with suppression assumed to be “even more crucial to the goal of comprehension” (Gernsbacher, 1997a, p. 86) than enhancement. Thus, although all models of discourse representation use a concept of activation due to input, the key characteristic that distinguishes the structure building model from others (e.g., Kintsch, 1988, 1998) is that comprehension is largely driven by the suppression of irrelevant information.

This emphasis on suppression has been supported by several findings of inhibition effects in studies of discourse processing. For example, in studies of lexical access, Gernsbacher and Faust (1995; Faust & Gernsbacher, 1996) demonstrated that times to judge the appropriateness of a probe word were slowed when that word was related to the irrelevant meaning of a homonym that appeared in the preceding sentence. In a study of cataphoric processing, Gernsbacher and Jescheniak (1995) demonstrated that stressing one noun in a sentence resulted in slowed recognition of a second noun that had appeared in that sentence. Inhibition effects have also been obtained in several studies of anaphoric processing in which the discourse contained two potential referents. In these experiments (e.g., Gernsbacher, 1989, 1990; Lucas, Tanenhaus, & Carlson, 1990; MacDonald & MacWhinney, 1990), the candidate referents were introduced before an anaphoric word or phrase. They were then presented as probes following the anaphor. Response times were slower to probes of the nonreferent following the anaphoric reference than in a nonanaphor control condition. In these experiments (e.g., Gernsbacher, 1989, 1990) has hypothesized that this inhibition is due to the suppression of the nonreferent during the process of antecedent selection.
The present experiments further investigate the nature of such inhibition in the context of anaphoric reference. Of specific interest is whether the inhibition that is observed is the result of an immediate and automatic process where irrelevant information is suppressed during the process of forming an initial discourse representation or whether the inhibition effects may reflect more delayed, strategic, or controlled processes. If suppression is the mechanism by which readers are able to arrive at the correct resolution of an anaphoric reference, as suggested by early versions of the structure building model, then this would be consistent with an immediate and automatic suppression mechanism. Many previous results, such as those mentioned above, appear to provide strong support for this view that comprehension is driven by an immediate and automatic suppression mechanism.

However, all of the previous experiments that found inhibition for nonreferents involved binary decision tasks, either recognition (e.g., Gernsbacher, 1989; MacDonald & MacWhinney, 1990) or lexical decision (Lucas et al., 1990). This limitation may be of some importance because there is evidence in several areas of research that the nature of the probe task affects the results. In experiments on inferential processing, Potts, Keenan, and Golding (1988) found evidence of predictive inferences when a lexical-decision task was used but not when a naming task was used. Presumably, this did not reflect a lack of sensitivity of the naming task because these investigators found naming-time effects when other types of inferences were tested. In experiments on semantic priming, Seidenberg, Waters, Sanders, and Langer (1984) also found that priming effects depended on the nature of the probe task; for example, fly facilitated a lexical decision to butter but did not facilitate a naming response. Further, in a recent study on the activation of multiple meanings of homographs, Long, Seely, and Oppy (1999) found that readers had difficulty suppressing the irrelevant meaning of a word only when the task was susceptible to context checking. No difficulties were seen in naming tasks. These findings suggest the decision probe tasks may be causing the inhibition effects and not simply measuring the results of basic processes of comprehension.

Parallel concerns about the influence of probe tasks on the observation of inhibition effects have been noted in the selective attention literature. Similar to tasks in the discourse processing area in which an ambiguous word’s meaning or anaphor’s referent needs to be selected from multiple candidates, tasks in the selective attention area require participants to attend to only one of two presented stimuli. In both literatures, responses to probes corresponding to the unselected entity are found to be inhibited. As in the account of inhibition offered by the early structure building model, the dominant explanation for negative priming effects in the selective attention literature has been that the distractor has its activation suppressed (inhibited below baseline) during the selection of the target.

However, in recent studies of selective attention, inhibition of responding to the previously unattended stimulus (negative priming) has been shown to depend on the probe task (Milliken, Joordens, Merikle, & Seiffert, 1998). As Milliken et al. have noted, activation "can be qualitatively modulated by processing that occurs during the computation of a response to the probe" (p. 203). There are two possible explanations for why decision probe tasks in and of themselves may cause inhibition effects. One possible explanation is that during the decision task, the participant experiences a response conflict or interference due to a memory trace related to prior experience with the distractor. Consistent with this position, Neill, Valdes, Terry, and Gorfein (1992) have proposed an episodic-retrieval account of negative priming that may also apply to inhibition effects in language processing. The basic idea is that response information may be encoded as part of the prime episode. For example, an ignore-this-stimulus tag may be associated with the distractor in the selective-attention task. When the probe is the distractor, there is a conflict between the tag based on prior experience and the current need to make a positive response. Similarly, when a binary decision is made in an anaphor-resolution experiment with several candidate referents, in the preceding process of selecting the antecedent, the nonreferent may receive consideration and a "no-tag" may become associated with it. In this account, the slow responses to binary probes of the nonreferent reflect a conflict between the yes response required by the probe and the no response that became associated with the concept during the search for the referent. Whereas the automatic-suppression account assumes a lowering of activation on the nonreferent during anaphor resolution and therefore predicts slowed responses independent of the probe task, the no-tag hypothesis predicts a slowed response only in decision tasks and has nothing to say about response times to nonreferents in tasks that do not involve a choice.

A second explanation for why decision probe tasks may cause inhibition effects is that they may prompt a context-checking process, where slower responses result from the incompatibility of the probe with the preceding context (Potts et al., 1988; Seidenberg et al., 1984; West & Stanovich, 1982). This explanation of inhibition in discourse-processing experiments would suggest that it reflects the failure of a comparison process between the probe and the immediately preceding information. For example, in the anaphor-resolution experiments of Lucas et al. (1990), a sentence such as I was followed by either 1a, the referent, 1b, the nonreferent, or 1c, the control:

1. At the zoo, the children’s favorite animals are the lions and the camels.
   1a. Sandra’s favorite animal is the jungle animal.
   1b. Sandra’s favorite animal is the desert animal.
   1c. We go to the zoo about once a year.

Following the second member of the pair, readers were required to make a lexical decision in response to the probe word lion. Lexical-decision times were slower in Condition 1b than in Condition 1c. The context-checking account would suggest that the incompatibility of the nonreferent probe lion following the reference, the desert animal, caused the slowing of the appropriate yes response. Milliken et al. (1998) have proposed a similar comparison process to account for negative priming in a selective-attention task; the comparison in those experiments is assumed to be between the probe and the representation of the attended-to stimulus.

Either the context-checking or the no-tag/response-conflict account provides an alternative explanation in which inhibition effects may occur as a function of the probe task, as opposed to as a function of automatic or immediate changes in activation for the nonreferent. These alternative explanations for inhibition effects suggest that the suppression of irrelevant information that has been observed in probe tasks may be the result of strategic processes.
that are prompted by the decision tasks themselves. Thus, this hypothesis would seem to support a model of suppression as a more delayed, controlled, or strategic process, as opposed to a more immediate and automatic one. Further, the results of a recent study by Gernsbacher and St. John (1994) also lean toward this conceptualization of the suppression mechanism. In discussing their investigation of the time course of inhibition for nonselected meanings of ambiguous words, Gernsbacher and St. John suggested that suppression of irrelevant information may occur only after a sentence-level Gestalt has been computed. Once a coherent model of the sentence has been constructed, it is then used to direct the reduction in activation at that level and at lower levels of representation. This is consistent with a model of suppression as a more delayed, controlled, and strategic process. By the same reasoning, for anaphor resolution, we might expect that nonreferents may be inhibited or suppressed only at later stages of discourse processing and not during the resolution of the anaphor or during the process of forming an initial discourse representation.

In the present experiments, we used measures other than binary-decision response times to assess changes in activation of nonreferents following anaphoric reference, in order to specifically evaluate the hypothesis that activation decreases for nonreferents during the process of antecedent selection. In two preliminary experiments, we replicated previous inhibition effects using our materials in a recognition probe paradigm. In Experiments 1–3, we used naming time rather than recognition probes, because as several individuals have noted (e.g., Potts et al., 1988; Seidenberg et al., 1984; West & Stanovich, 1982), the time to pronounce a word appears to be less affected by decision processes following the onset of the probe than are lexical-decision or recognition–response times. Naming times have frequently been used as a contrast to decision probes. In a fourth experiment, we used a previously untapped source of evidence to bear on the question of inhibition effects in discourse processing; we examined the amount of time the eyes rest on a word as a measure of activation (e.g., Clifton & Ferreira, 1987; Duffy & Rayner, 1990; Ehrlich & Rayner, 1983). Because the tasks used in these experiments, naming and reading, do not involve making a choice, such measures should more directly reflect the activation levels of the referent and nonreferent.

In summary, there is substantial evidence in the discourse processing literature of inhibition in response to probes inappropriate to the discourse segment currently being processed. These results have been taken as support for a model in which the current linguistic input causes immediate and automatic enhancement of relevant information and suppression of irrelevant information. An assumption critical to this model is that response times to probes directly reflect the activation state of the represented concept prior to the onset of the probe and independent of the nature of the probe. However, the evidence rests largely on results obtained with recognition and lexical-decision probes. Because there is evidence from several areas of cognitive research that results are qualitatively influenced by the probe task, and because alternatives to an immediate, automatic suppression mechanism have been suggested to account for results in these areas, it is important to reexamine the conditions under which inhibition is obtained. Therefore, the present experiments were designed to provide converging evidence about the nature of processes underlying resolution of anaphora using several different measures. The critical issue is whether inhibition effects are observed in all tasks. If the model of suppression as an immediate and automatic process is correct, then the decrease in activation associated with the nonreferent should be reflected in slower processing for the nonreferent in all tasks, including naming and reading. However, if inhibition is only obtained with recognition tasks, it would call into question the hypothesis that antecedent selection involves an automatic and immediate lowering of activation of nonreferent nodes in the discourse representation and would suggest that inhibition effects are a consequence of more strategic, delayed, or controlled processing.

**Experiment 1**

Using materials similar to those of Lucas et al. (1990), we evaluated the hypothesis that activation decreases for nonreferents during antecedent selection in the present experiments. In two preliminary experiments, we replicated findings of inhibition of nonreferents with categorical anaphora in a recognition probe task, using two alternative control procedures. In one experiment, we probed before and after the anaphor. In a second experiment, we compared categorical anaphor sentences with two control sentences that contained nonreferential categorical nouns and we probed at the end of the categorical phrase. In both experiments, response times were significantly slower in the nonreferent probe than in the control condition. A summary of salient aspects of the materials, method, and results is presented in Appendix A.

The results of both preliminary recognition probe experiments were consistent with those reported by Lucas et al. (1990); there was no evidence of facilitation for the referent and there was clear evidence of inhibition of the nonreferent. Thus, the results are replicable despite differences in procedure and materials, and most importantly, with two different types of controls. Despite the fact that Lucas et al.'s study and the first preliminary experiment probed at the end of a sentence and the second preliminary experiment probed in mid-sentence, all studies revealed inhibition effects with immediate probes. Further, responses to the nonreferent condition were inhibited, but responses to a specific control condition were not. The specific control condition introduced a categorical noun phrase that excluded both possible referents. If the increase in response time to nonreferent probes was due to an incompatibility of the probe to the categorical noun phrase, then both the nonreferent and the specific control condition should show inhibition. The failure to find inhibition in the specific control condition suggests that the nonreferent inhibition result was not simply due to an incompatibility between the probe and the reference in the preceding sentence.

Both preliminary experiments provide evidence that we can demonstrate inhibition in a recognition probe task with our materials. Further, the use of alternative control conditions suggests that this result cannot be attributed to context checking or differences in probe locations. However, there are still two distinct explanations that can account for these results: inhibition in responses due to the nature of the probe task versus inhibition due to changes in activation of potential referents during the course of anaphoric processing. In order to distinguish between these accounts, it is important to determine whether this inhibition effect will generalize to other tasks. Thus far, inhibition of nonreferents has been obtained only with binary decision tasks. As several individuals
have noted (e.g., Potts et al., 1988; Seidenberg et al., 1984; West & Stanovich, 1982), the time to pronounce a word appears to be less affected by strategic processes that follow the probe onset. If activation of a nonreferent concept is suppressed during anaphor resolution, naming times should reflect this. The present experiment addresses this hypothesis.

The observation of inhibition effects in naming times would strengthen the argument that anaphor resolution involves the suppression of nonreferents. There is evidence that inhibition effects can be obtained with naming probes, although the stimulus materials did not involve anaphor resolution. MacDonald and Just (1989) presented sentences such as

(2a) Almost every weekend, Elizabeth bakes no bread but only cookies for the children.
(2b) Almost every weekend, Elizabeth bakes some bread but no cookies for the children.
(2c) Almost every weekend, Elizabeth bakes some bread and some cookies for the children.

The negated word (bread or cookies) was named more slowly following Sentences 2a and 2b than following Sentence 2c. Although negation, not anaphor resolution, is the source of the inhibition, the results encourage us to consider naming time in an anaphor-resolution experiment. However, in two previous experiments (Dopkins & Nordlie, 1997; Lucas et al., 1990), times to name nonreferents were not slower in anaphor than in control conditions. Aside from the fact that the sentences did not involve anaphoric processing, the MacDonald and Just experiments differed in one other respect from the Lucas et al. (1990) and Dopkins and Nordlie (1997) experiments; the probe appeared several words after the targeted word (e.g., cookies) and at the end of the sentence. It may be that the inhibition observed in the MacDonald and Just experiments is due to a process that takes some time to complete, such as updating a discourse representation. If so, that would explain why naming-time effects were not observed with probes immediately after the anaphor (Dopkins & Nordlie, 1997), even when the anaphoric expression occurred at the end of the sentence (Lucas et al., 1990).

Because naming effects have not been obtained using anaphoric sentences, in Experiment 1, we attempted to stay as close as possible to a condition in which a similar effect had been obtained, that of the MacDonald and Just (1989) study. To this end, we designed our materials so that a few words (a prepositional phrase) intervened between the anaphor and the sentence end, and we presented target words to be named following the end of the sentence. The passages were based on those used in the Lucas et al. (1990) study. Each version of each passage began with the same two or three introductory sentences, the last of which mentioned two possible referents. This was followed by one of three sentences:

Billy’s parents wanted him to take music lessons.
They wanted him to learn the violin or the drum.
Referent. (3a) He finally chose the stringed instrument for his lessons.
Nonreferent. (3b) He finally chose the percussion instrument for his lessons.
Control. (3c) He finally chose a wind instrument for his lessons.

The probe word was violin. Experiment 2 tested whether the final prepositional phrase was necessary in order to obtain an inhibition effect. Experiment 3 combined both kinds of materials in a single experiment.

Method

Participants. The participants were 48 University of Massachusetts at Amherst undergraduates enrolled in psychology courses who received extra credit. Each participant was randomly assigned to one of two material sets, with the restriction that each set be read by an equal number of participants. Seven participants were replaced because of errors on more than 12 questions, 2 participants were replaced because of very long naming times (mean response time greater than 1100 ms) and 2 were replaced because of equipment malfunction.

Materials. The materials were based on the 24 experimental texts used by Lucas et al. (1990). Each experimental text introduced two potential antecedents followed by a categorical noun phrase. For these materials, the categorical noun phrase was included as the direct object phrase of the final sentence. After the categorical noun phrase, a final prepositional phrase was included (following MacDonald & Just, 1989), and probes were presented at the end of the sentence. There were three conditions for each text: the referent, nonreferent, and specific control. In the referent condition, the anaphoric phrase referred to the antecedent that was used as a probe. In the nonreferent condition, the anaphoric phrase referred to the potential antecedent that was not used as a probe. In the specific control condition, the categorical noun phrase excluded both of the potential antecedents and referred to a subset of the category. In addition, in the specific control condition, the categorical noun phrase was preceded by the indefinite pronoun a, whereas in the referent and nonreferent condition, the categorical noun phrase was introduced by the article the. We opted for the indefinite article because the use of the definite article in the control condition would have falsely implied that the categorical noun was a reference to a prior entity, which may have inflated reading times in the control condition (cf. Gernsbacher & Jescheniak, 1995; Haviland & Clark, 1974; Murphy, 1984).

There were two sets of experimental materials. Within a set, in half of the items the referent (e.g., “violin”) was mentioned before the nonreferent (“... to learn the violin or the drum”); the reverse order was used in the other half of the items in that set. The second set counterbalanced the order of mention (“... to learn the drum or the violin”). Thus there were two groups of participants, differing only in the set to which they were assigned.

In addition to the 24 experimental texts, there were an additional 24 filler texts of similar form (containing two possible referents and categorical anaphors), which were followed by probe words from previous texts. Half of these probes were possible referents, whereas the other half were other content words. There were also an additional 24 filler passages with no categorical anaphor in which the probe word either appeared very early (i.e., after 4 or 5 words) or very late (i.e., at the end of the fourth sentence). Half of the early probe and late probe filler had correct recognition probes and the other half were incorrect (words that did not appear in the text). This yielded an equal number of yes and no recognition responses. Examples of the similar, nonsimilar/early probe and nonsimilar/later probe filler passages are presented in Appendix B.

Gordon, Henrick, and Foster (2000) convincingly demonstrated that the use of fillers is very important in studies with recognition probes, especially to the extent that they limit the unpredictability of probe terms, and thus limit strategic responding. To limit confounds because of such concerns, participants were presented with twice as many filler passages as target passages. Further, one quarter of probes in the filler conditions were content words other than the referent and nonreferent terms that appeared in the passage (requiring a yes response); one quarter were referent and nonreferent terms from previous passages (requiring a no response); and
one quarter were content words that appeared in previous passages (requiring a no response). The remaining filler condition were words that did not appear in any texts (requiring a no response). In this way, attention to the two possible referent candidates alone would not allow for good performance on the recognition probe task, limiting the possibility of strategic responding in the probe task.

Two practice passages, similar in form to the experimental passages, preceded the task. A comprehension question followed each passage to ensure that participants were reading carefully. The questions did not concern resolution of the potential anaphor or either of the potential antecedents.

Procedure. Participants were tested individually in an experimental session that lasted approximately 30 min. All materials were presented on a video monitor controlled by a 386 computer. Before the experimental task, participants were given a 50-item practice task to familiarize them with a speeded-naming procedure. When responses were longer than 1200 ms, the words "TOO SLOW" appeared on the screen for 750 ms as feedback. For the experimental task, the phrase "Press the advance key to continue" appeared at the top of the screen before each trial. When participants were ready to begin a passage, they pressed the advance key. This initiated presentation of the passage. Participants were presented one word at a time at the center of the screen at a rate of 450 ms per word, plus an additional 16.67 ms per character, as in Gernsbacher (1989). Each word was erased and replaced by the onset of the next word. At the end of the passage, the screen was blanked and a word appeared in all capital letters two lines above the center of the screen. Participants were instructed to read this word aloud as quickly as possible. Then, participants were presented a question about the passage on the computer screen. Participants answered the question by pressing the yes or no key. On those trials in which the question was answered incorrectly, the word "ERROR" was presented for 750 ms as feedback.

Results and Discussion

Because there were some trials in which participants had to repeat the probe word in order to activate the voice key, all response times greater than 1200 ms were deleted. This removed 5% of the scores from the analysis. The means for the remaining scores are presented in the first row of data in Table 1. There was a significant main effect for condition, $F_1(2, 92) = 4.103, MSE = 1.942, p < .02; F_2(2, 46) = 4.109, MSE = 861, p < .02$. Both referent and nonreferent naming times were slower than those for the control. The contrast of the referent mean with the control mean was not significant; $F_1$ and $F_2$ were both less than 1. However, the nonreferent mean was significantly slower than the control; $F_1(1, 46) = 7.595, MSE = 4.016, p < .01; F_2(1, 23) = 8.302, MSE = 1.659, p < .01$.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Mean Naming Times for Experiments 1–3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probe location</td>
<td>Condition</td>
</tr>
<tr>
<td>Experiment 1</td>
<td>Delayed</td>
</tr>
<tr>
<td>Experiment 2</td>
<td>Immediate</td>
</tr>
<tr>
<td>Experiment 3</td>
<td>Delayed</td>
</tr>
<tr>
<td></td>
<td>Immediate</td>
</tr>
</tbody>
</table>

Note. Mean number of voice-key errors are shown in parentheses.

This result parallels the recognition probe results and provides important support for the suppression account of inhibition effects. The slower naming times for just the nonreferent probes provides converging evidence that nonreferents are less accessible after the anaphor has been resolved.

However, the present experiment did not produce the same results as the naming time experiment in Lucas et al. (1990). In their experiment, Lucas et al. found facilitation of responses to both the referent and nonreferent probes in their naming-time study. Presumably, this reflects the lack of a control for semantic priming. Such priming may have more than offset the inhibition found in the present experiment. In addition, there is a further difference between the Lucas et al. experiment and ours. We inserted several words between the anaphor and the probe following the end of the sentence. In Experiment 2, we tested whether this additional prepositional phrase before the probe is necessary.

Experiment 2

Experiment 2 is a replication of Experiment 1 except for one change; the words between the anaphor and the end of the sentence were removed from all versions of the passages. Sentences 3a–3c were truncated to end with the word instrument and the probe was then presented. If the activation level of the nonreferent has been lowered during the anaphor-resolution process, we should again find inhibition.

Method

Participants. The participants were 20 University of Massachusetts at Amherst undergraduates enrolled in psychology courses who received extra credit. Each participant was randomly assigned to one of two material sets, with the restriction that each set be read by an equal number of participants. One participant was replaced because of long response times, and one was replaced because of equipment malfunction.

Materials and procedure. The materials were the same 24 experimental texts used in Experiment 1, presented without the final prepositional phrase. The rest of the materials and the procedure were exactly the same as in Experiment 1.

Results and Discussion

We again deleted the highest 5% of the scores. The mean naming times of the remaining scores are presented in the second row of data in Table 1. The main effect for condition was not significant; both $F$s < 1. The mean for the referent condition is somewhat lower than that for the control, but the effect is far from significant; both $F$s < 1. There is no indication of inhibition of the nonreferent mean; again, both $F$s < 1. We terminated this experiment with only 20 participants, because it was clear that there was not even a trend for inhibition of responses to the nonreferent probe.

Experiment 3

The strong inhibition effect obtained in Experiment 1, coupled with the lack of any indication of inhibition in Experiment 2, suggests that the inhibition of naming-time responses is not immediate. If true, this is an especially interesting finding because it is in conflict with previous findings of immediate inhibition of nonreferents with decision tasks. In order to see if we could
replicate the naming-time findings with a different design we ran Experiment 3, a replication of Experiments 1 and 2 in which the probe delay was a within-participant manipulation. Experiment 3 provides a particularly strong test of the replicability of the previous findings, not only because of the change in design, but because a different participant population was sampled and the experimenter recorded those occasions when a response failed to activate the voice key. In this way, there was no need to delete response times on the basis of some arbitrary criterion for outliers.

Method

Participants. The participants were 36 undergraduates at Washington State University–Vancouver enrolled in psychology courses who received extra credit. Each participant was randomly assigned to one of six material sets, with the restriction that each set be read by an equal number of participants. Participants were replaced because of long response times, and one was replaced because of equipment malfunction.

Materials. The experimental materials were the same 24 texts used in Experiments 1 and 2. In each set, half the passages were presented in an immediate condition in which the probe was presented immediately after the anaphoric phrase as in Experiment 2. The other half of the materials were presented in a delayed condition with the probe following a final prepositional phrase as in Experiment 1. In both the immediate and delayed conditions, one-third of the probes were referents, one-third were nonreferents, and the final third were presented after the specific control sentences. Each passage appeared in each condition an equal number of times, and conditions were counterbalanced across the six sets of materials. The probe word was the first mentioned of the possible referents for one half of the passages for all participants. The filler materials were the same as those used in previous experiments.

Procedure. Participants were tested individually in an experimental session that lasted approximately 30 min. The experimenter was present in the room for this experiment. All materials were presented on a video monitor controlled by a 486 computer. The program was written using Micro-Experimental Laboratory software (Schneider, 1988) and a serial-response box from Psychology Software Tools. Participants were given a 50-item practice task. Participants were presented with feedback about the length of their response time after each trial. When responses were longer than 1200 ms, the experimenter told them to try to respond faster. For the experimental task, the phrase “Press the plus key for more text” appeared at the top of the screen before each trial. When participants were ready to begin a passage, they pressed the plus key on the right keypad. Passages were presented one word at a time at the center of the screen at a rate of 450 ms per word. Each word was erased and replaced by the onset of the next word. At the end of the passage, the screen was blanked and a word appeared in all capital letters two lines above the center of the screen. Participants were instructed to read this word aloud as quickly as possible. After each naming response, the experimenter coded whether the response was accurate. Then, participants were presented with a question about the passage on the computer screen. Participants answered the question by pressing the 1 key on the right keypad for yes or the 2 key for no.

Results and Discussion

The means for the naming times on accurate trials are presented in Table 1; the patterns are similar to those observed in Experiments 1 and 2. Whereas the main effect for delay was not significant, $F_{(1, 30)} = 3.02, MSE = 2,541, p = .09$; $F_{(2, 36)} = 3.89, MSE = 2,368, p = .06$. The difference between the nonreferent and control naming was very small and clearly not significant, $F < 1$. However, in the delayed condition, the nonreferent mean was significantly slower than the control, $F_{(1, 30)} = 20.5, MSE = 2,674, p < .01$; $F_{(2, 36)} = 20.01, MSE = 1,660, p < .01$; the referent mean again was not significantly faster than the control, $F < 1$. The mean number of voice-key errors (trials in which the response box failed to pick up the naming response accurately, or trials in which the word was named incorrectly) are presented in parentheses in the third and fourth rows of Table 1. It should be noted that these were not errors in the usual sense but were primarily failures of vocal responses to trigger the voice-activated relay.

The naming-time results replicated the findings of Experiments 1 and 2, with the delay manipulation run within participants. Without delay, there was no inhibition for the nonreferent compared with the control condition. The inhibition was present in the nonreferent condition only when the probe was delayed until after a prepositional phrase at the end of the sentence.

The strong inhibition effect obtained in both Experiments 1 and 3 after a sentence-ending prepositional phrase, coupled with the lack of any indication of inhibition immediately following the anaphoric phrase in Experiments 2 and 3, demonstrates that the inhibition of naming-time responses was not immediate. One possible explanation is that the anaphoric inference required time to be resolved and the nonreferent was not suppressed until resolution was completed. However, in our first preliminary experiment, as well as in other such experiments (e.g., Gernsbacher, 1989, 1990; Lucas et al., 1990; MacDonald & MacWhinney, 1990), recognition was probed immediately after the anaphor, and inhibition was observed. This appears to argue against the hypothesis that time is required for anaphor resolution. One explanation for the difference in time courses for naming and recognition measures is that the inhibition effects in the two tasks may have different causes. Because naming times have been shown to be less sensitive than binary-decision times to processes following the probe onset, it is reasonable to conclude that naming tasks provide a more veridical picture of the changes in activation resulting from an anaphoric inference. Thus, the results of Experiments 1 and 3 indicate that naming responses to nonreferent probes can be inhibited; however, contrary to the implication of the recognition data, this is the result of a process that requires time to complete after the anaphor is presented. Thus, only the later inhibition effects are likely due to changes in activation following anaphor resolution. Without converging evidence from naming tasks, the immediate inhibition

\[\text{ACCESSIBILITY OF POTENTIAL REFERENTS}\]
effects for nonreferents in decision probe tasks are more likely due to difficulty in the decision phase of the task.

As we noted in the introduction, one possible reason for that difficulty may be context checking; the slow recognition time in the nonreferent case would be because the probe was incompatible with the categorical anaphor; for example, a violin is not a percussion instrument. However, a violin is also not a wind instrument, which is the control used in the second of our two preliminary recognition experiments. Nevertheless, responses were significantly slower in the nonreferent than in the control condition. A second possible cause of the inhibition effect in recognition responding seems more likely. Following Neill et al.'s (1992) episodic-retrieval account of negative priming in selective attention experiments, we assume that in the process of selecting an antecedent, the nonreferent was marked as "not the referent." The presence of this no-tag could interfere with a positive recognition response when the referent was later presented as a probe. Further, it should not interfere with a naming response. This is the pattern we found for immediate probes.

In summary, the inhibition observed in recognition tasks seems likely to be the result of the decision process, whereas the inhibition observed in the naming-time data of Experiments 1 and 3 appear more likely to be independent of the probe task. Before further considering the implications of these findings, we examine one other source of information about how and when nonreferents may become less accessible. In Experiment 4, we investigated whether an inhibitory effect might be seen in eye-movement data as well, and if so, when the effect might show up in the course of normal reading.

Experiment 4

In our preliminary experiments, participants were required to decide whether a probe word had been in the preceding text. In Experiments 1–3, the participant had to say the probe word aloud. The response, whether recognition or naming, effectively terminated the process of reading the passage. However, reading requires that the words are integrated into the developing discourse representation. In Experiment 4, we monitored eye movements in order to see the effect of an anaphoric reference on reading times. Instead of using probes to investigate changes in activation, the amount of time that a reader spent on a second mention of a referent or nonreferent following an anaphoric reference was measured. Previous experiments have shown that variables that are generally assumed to reflect activation levels have affected gaze durations (e.g., Clifton & Ferreira, 1987; Duffy & Rayner, 1990; Ehrlich & Rayner, 1983).

In particular, results of previous eye-tracking investigations on the "distance effect," in which the distance between an anaphor and its antecedent was varied, have shown that the availability of the antecedent noun in memory influences the speed of anaphor assignment. When an antecedent noun is distant, it leads to longer viewing times on an anaphoric noun (Duffy & Rayner, 1990) and longer fixation times on the fixation following an anaphoric pronoun (Ehrlich & Rayner, 1983) than when the antecedent is near. Further, it is not distance per se that contributes to difficulty of resolving an anaphor. Other factors, such as whether the antecedent is still the topic of the discourse (Clifton & Ferreira, 1987), the typicality of the antecedent, and the presence of distractors (Crottell, 1984) also affect the amount of difficulty readers will encounter in resolving an anaphor. Whether the availability of the anaphor is varied as a function of distance or other factors, we predict that decreased availability of an antecedent results in difficulty resolving the anaphor, and that difficulty is observed as longer reading times on the anaphoric expression or on the words that immediately follow it.

In the present study, we modified materials from previous experiments to include a second mention of one of the possible referents. This second mention followed the anaphoric reference to the violin, or reference to another object already mentioned (e.g., the drum), or reference to a third, new object, as shown in the following sample passage from the referent condition:

Billy's parents wanted him to take music lessons. They wanted him to learn the violin or the drum.

After some consideration of the stringed instrument, he decided that the violin would be very nice to learn.

For the nonreferent and control conditions, the final sentence was replaced by:

Nonreferent. After some consideration of the percussion instrument, he decided that the violin would be very nice to learn.

Control. After some consideration of a wind instrument, he decided that the violin would be very nice to learn.

The time that the eye rests on the second mention of "violin," and on the words following it, provides a measure of the activation of referents and nonreferents, and of how reading proceeds following the remention of such terms. Thus, if readers suppress the concept of "violin" when they resolve the anaphoric expression the percussion instrument, then the concept "violin" should be less available at its mention later in that sentence. On the basis of previous work in pronouns and noun anaphors, lessened availability should result in longer fixations on the anaphoric noun ("violin") and/or on the fixations that follow it in the nonreferent condition. However, if no differences are seen on the anaphor itself ("violin") or immediately following it ("would be"), it would suggest that the availability of the nonreferent has not been altered by the categorical anaphor that was read earlier in the sentence. Further, the delayed/end-of-sentence inhibition results we obtained with naming-time data would predict that no effects should be seen in availability when the nonreferent is reintroduced into the text before the end of the sentence containing the anaphoric expression. The fact that inhibition effects on naming time were observed only when the probe was presented at the end of the sentence suggests the possibility that inhibition effects on eye movements may be limited to the end-of-sentence region ("very nice to learn") following the second mention of "violin."

It is possible that fixations on the second mention of "violin" may also be affected by distance or topic shift between the explicit anaphor and an antecedent phrase. In that case, we would expect that reading times on "violin" and/or on the words immediately following it ("very nice") should be longer in both the nonreferent and the control condition, because in both these conditions, there was no reference to the violin in the immediately preceding clause, and a different instrument ("the percussion instrument" or "a wind instrument") was introduced. Either distance or topic shift could affect the availability of the violin concept, but would do so in both conditions.
Method

Participants. Thirty-six students from the University of Massachusetts at Amherst participated in this experiment for class credit in a psychology course or for $8. Participants had normal uncorrected vision, or vision corrected with soft contact lenses, and were native English speakers.

Apparatus. Eye movements were recorded from the right eye using a Dual Purkinje Eyetracker (Fourward Technologies, Buena Vista, VA), which has a resolution of 10 min of arc. The eye tracker was interfaced with an IBM compatible computer, which controlled all phases of the experiment. The computer was programmed to sample eye position every millisecond. The passages were displayed in standard uppercase and lowercase format on an NEC monitor, using white letters on a black background. Participants’ eyes were 62 cm from the monitor, and viewing was binocular. Monitor brightness was adjusted to a comfortable level for each participant. The texts occupied between three and six lines, which were presented starting on the eighth line of screen and were double spaced. Most lines did not exceed 60 characters.

Materials. The materials were based on the 24 experimental passages used in Experiments 1-3. As in previous materials, each passage set up a choice between two category members, and included a categorical noun phrase that either referred to the first mentioned category member, the second mentioned category member, or neither of the two potential referents. However, instead of presenting possible referents as probe words, the probes were presented by rementioning one of the possible referents in the last sentence of the text. A number of other changes were made from previous materials. For more natural presentation, all of the texts were completed. Special attention was paid to finishing the control and nonreferent sentences in such a way so that they would not seem like obvious contradictions. Further, because shorter words are often not fixated on during normal reading, shorter referents were replaced so that all target words were at least five letters long. Finally, passages were written so that the target words appeared in the middle of a line, where track losses are least likely, and so that the next three or four words following the target region were the same in each condition. Half of the rementions in the last sentence were the first mentioned possible referent and half were the second mentioned possible referent.

Because reading time on the nonreferent passages was of central interest in this experiment, a norming study was performed on the naturalness of the passages to ensure that longer reading times might not result from infelicity. The ratings showed no indication that the nonreferent passages were less natural than the other conditions of the experimental passages. Twenty-four students at Washington State University–Vancouver rated the naturalness of the passages using a scale rated from 1 (meaning not at all natural) to 7 (meaning quite natural). Each participant rated eight passages in each condition, with the passages assigned to each condition counterbalanced across participants. Mean ratings for the three conditions were 4.66 (nonreferent), 4.64 (referent), and 4.31 (control).

In addition to the 24 experimental passages, there were 24 filler passages that were two or three sentences in length and were not similar to the experimental passages (they did not include categorical anaphors). Passages were presented in a random order for each participant.

Procedure. Participants were tested individually in an experimental session that lasted approximately 45 min. When the participant arrived, he or she was given a brief description of the procedure, and a bite bar was prepared to prevent head movements during reading. The participant was told that the purpose of the experiment was to study how people read. The participant was instructed to read the passages in a normal manner, at whatever rate was comfortable, so that they could answer questions about the passages. After the instructions were understood, the eye tracker was calibrated. The calibration procedure never took longer than 5 min.

All materials were presented on a video monitor controlled by a 486 computer. Before each passage, five fixation boxes were displayed where the first line of the text would appear. The participant was instructed to look at the center box and then at each box to the left until reaching the left-most box. If the calibration was accurate, the experimenter then initiated the trial and the text replaced the fixation boxes. After reading the passage, the participant pressed a response key, which caused the screen to clear and the fixation boxes to appear. Simple yes–no comprehension questions were asked at random intervals when the fixation boxes appeared after reading a text.

Results and Discussion

Reading time data from three regions were examined: the target word, which was the second mention of either the referent or nonreferent (“violin”); the posttarget region, which consisted of the two (or three) words that immediately followed the second mention of the possible referent (“would be”); and the end-of-sentence region, which consisted of the remaining words in the sentence (“very nice to learn”). First-pass and total reading times for the target word, posttarget, and end-of-sentence regions are presented in Table 2. The first pass measure is the sum of all the fixations a reader makes from first entering a given region until leaving the region. The total time measure represents the total amount of time a reader spends fixating in a region, including regressions. Single fixations less than 120 ms and greater than 700 ms were excluded from all analyses.

There were no significant differences between either the referent or nonreferent and the control condition for either of the measures recorded for either the target or posttarget regions; all Fs < 1.16. Thus, there was no indication in the eye-tracking data of inhibition of access of nonreferents. However, longer reading times in the end-of-sentence region indicated that readers had difficulty in integrating nonreferents. First-pass times in that region were slower in the nonreferent than in the control condition; the effect was significant when tested against participant variability, $F_1(1, 35) = 6.06, MSE = 19,748, p < .02$, and marginally significant against item variability, $F_2(1, 20) = 3.74, MSE = 21,992, p < .07$. First-pass times also were slower in the referent than in the control condition, but they were not significant; $F_1(1, 35) = 3.50, MSE = 17,443, p < .07$, $F_2 = 1.35$. The same pattern of significantly longer reading times in the nonreferent condition was found for total times in the end-of-sentence region, $F_1(1, 35) = 15.6, MSE = 18,076, p < .0001$, $F_2(1, 20) = 6.19, MSE = 26,248, p < .02$. Similarly, total times were slower in the

<table>
<thead>
<tr>
<th>Measure</th>
<th>Referent</th>
<th>Nonreferent</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First pass</td>
<td>286</td>
<td>279</td>
<td>291</td>
</tr>
<tr>
<td>Total time</td>
<td>305</td>
<td>296</td>
<td>306</td>
</tr>
<tr>
<td>Posttarget</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First pass</td>
<td>404</td>
<td>416</td>
<td>398</td>
</tr>
<tr>
<td>Total time</td>
<td>475</td>
<td>477</td>
<td>460</td>
</tr>
<tr>
<td>End of sentence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First pass</td>
<td>473</td>
<td>489</td>
<td>431</td>
</tr>
<tr>
<td>Total time</td>
<td>533</td>
<td>564</td>
<td>476</td>
</tr>
</tbody>
</table>
referent than in the control condition, but this effect was only significant against participant variability, $F_{1}(1, 35) = 4.12$, $MSE = 28,806, p < .05, F_{2} = 1.44$. Further, the total times in the end-of-sentence region were longest for the nonreferent condition, suggesting that integration into the discourse representation for the nonreferent condition may have been even slower than that for the referent, although the difference between referent and nonreferent processing times was not significant.

In attempting to integrate the second mention of the violin into the representation, the reader presumably mapped it onto the earlier mention; this would be true in all three conditions. In both the referent and nonreferent conditions, the reader may have also attempted to construct a bridging inference to link the second mention of the target noun with the preceding discourse. One could argue that the longer integration times in both the referent and nonreferent conditions simply reflect the need to resolve the anaphor, which may take extra time and effort as compared with the nonanaphoric control condition. However, rereading patterns suggest that the longer reading times in nonreferent and referent conditions may arise for different reasons.

As shown in Table 3, people made more regressions out of end-of-sentence regions ("very nice to learn") in both the referent, $F_{1}(1, 35) = 7.0, MSE = 357, p < .01; F_{2}(1, 23) = 5.02, MSE = 332, p < .04$, and nonreferent conditions, $F_{1}(1, 35) = 16.5, MSE = 236, p < .0001; F_{2}(1, 23) = 13.4, MSE = 195, p < .001$, than in the control condition. In the referent condition, readers tended to spend more time rereading the categorical anaphor ("the stringed instrument") than in the nonreferent condition ("the percussion instrument"), $F_{1}(1, 35) = 3.11, MSE = 61,069, p < .08; F_{2}(1, 23) = 2.65, MSE = 47,790, p < .11$. For example, after reading through the sentence "After some consideration of the stringed instrument, he decided that the violin would be very nice to learn," the reader might return to the phrase "the stringed instrument." Because the violin is still in focus following the anaphoric reference "the stringed instrument," the use of the full noun phrase "violin" to refer to the object that is still the current topic of the discourse may seem odd to the reader. The use of a pronoun would be the more natural continuation (Fletcher, 1981, Marslen-Wilson, Levy, & Tyler, 1982). As a result, the reader may regress to the anaphoric expression to check that the reference was interpreted correctly. In other words, reading the full name of an entity that is already in focus may prompt the reader to go back and check that the preceding topic was indeed the same. This slowdown on the noun phrase is reminiscent of the repeated name penalty, where readers take longer to read repeated noun phrases than pronouns when the discourse topic has not shifted (Gordon, Grosz, & Gilliam, 1993; Sanford, Moar, & Garrod, 1988). Repeating the full noun phrase when that concept is still in the foreground results in a very marked discourse status. Thus, the explicit naming of the full noun phrase "violin," which is already in focus following the anaphor "stringed instrument," may have contributed to the slower processing observed in the end-of-sentence region in the referent condition.

However, the rereading patterns in the nonreferent condition indicated that readers returned to the beginning of the passage and reread the section that listed the two possible referents ("the violin" or "the drum") more often than in the control condition, $F_{1}(1, 35) = 5.36, MSE = 28,182, p < .03; F_{2}(1, 23) = 5.04, MSE = 19,983, p < .04$. As opposed to the referent condition, where both the anaphoric expression ("the stringed instrument") and the remention of the referent ("violin") referred to the same entity, readers had to deal with two separate references in the nonreferent condition. The anaphoric phrase ("the percussion instrument") refers to one candidate referent ("drum"), but the nonreferent ("violin") was then mentioned again, all in the same sentence. Returning to the earlier mention of the referent and nonreferent could have helped readers as they attempted to sort out the two references and tried to build bridging inferences in order to integrate the nonreferent into the current discourse representation. Especially because of the shift in topic to the drum concept following the anaphoric reference ("the percussion instrument"), the reader may need to reactivate the earlier mention of "violin" in order to process the second mention of the concept. Regressing to the earlier mention of "violin" in the passage could help the reader to reactivate and integrate this concept.

In short, these different rereading patterns suggest that the longer reading times in the nonreferent condition may reflect difficulty in building bridging inferences to earlier occurrences, whereas the longer times following the referent condition may reflect a boggle caused by repeating the full name of an entity that is currently in focus following an anaphoric reference. The slowdown in the referent condition could possibly be eliminated by the use of a pronoun, suggesting that the slowdown in the end of sentence region in the referent condition may not reflect difficulty in the integration of the referent into the ongoing discourse representation. However, longer reading times in the end-of-sentence region in the nonreferent condition do seem to reflect a difficulty with integrating the nonreferent, and the regressions suggest the need to build bridging inferences to the same entity mentioned earlier in a text in order to integrate the second mention of the nonreferent. Further, this difficulty may be what is reflected in the inhibition effects found with end-of-sentence probe tasks.

The eye-movement data offer several insights into the anaphor-resolution process. First, the nonreferent did not result in longer reading times immediately following the anaphoric phrase. Second, there is evidence that sentences containing nonreferents are processed more slowly than control sentences, but the difficulty during reading is not in accessing the nonreferent but in integrating it. Third, it appears that both a delay and the prompt for integration provided by the period at the end of a sentence may be necessary for the inhibition of nonreferents following an anaphor (Just &
Carpenter, 1980). If only a delay following the anaphor was required, there should have been slower reading times in the posttarget region. However, such effects occurred only in the end-of-sentence region.

General Discussion

Gernsbacher (1989, 1990, 1997a, 1997b) theorized that the processing of a linguistic input involves both the enhancement of activation on nodes representing relevant discourse information and the suppression of activation on irrelevant nodes. In the particular instance of anaphor processing, it has been suggested that the competing referent candidates are automatically and immediately suppressed, allowing for the establishment of the correct referent (Gernsbacher, 1989). This assumption of nonreferent suppression during antecedent selection was supported by immediate inhibition results in several studies. However, because previous experiments used binary decision tasks, and because inhibition in response times to the probes used in such tasks are open to several interpretations, we attempted to replicate inhibition effects in a recognition paradigm in preliminary experiments, and then sought converging evidence for the immediate suppression hypothesis in naming time and eye movement. An inhibition effect was obtained immediately following the anaphor in the recognition paradigm of our preliminary experiments, whereas there was no inhibition of responses immediately following the anaphor in the naming paradigm of Experiments 2 and 3. However, when several words intervened between the anaphor and an end-of-sentence probe in Experiments 1 and 3, the nonreferent was named significantly slower than the control. Finally, eye-movement data revealed increased reading times in end-of-sentence regions following the mention of the nonreferent but not on the second mention of the nonreferent itself. The discrepant findings across methodologies suggest that immediate inhibition effects of nonreferents are not due to immediate suppression of the concepts but instead are due to the nature of the probe task.

Context checking (e.g., Potts et al., 1988; Seidenberg et al., 1984; West & Stanovich, 1982) provides one alternative account of the inhibition of recognition responses. However, if the slow response in the nonreferent condition reflects the incompatibility of the probe, *violon*, with the immediately preceding phrase, *the percussion instrument*, it should be equally incompatible following *a wind instrument*, and therefore equally slow. A more promising explanation of the inhibition of nonreferent recognition responses is that the reader retrieves the episodic trace associated with the probe (Neill et al., 1992). When the probe is the nonreferent, that trace includes a no-tag. Then, slow responses reflect a conflict between the correct yes response that is required for this decision task, and the no response that has been associated with the concept.

Thus, the results of the present experiment support the hypothesis that some inhibition effects that have been observed in anaphoric processing may have reflected response conflict as a result of the probe tasks that were used. However, although this response-conflict explanation accounts for the slowed performance associated with nonreferents in a binary decision task, it fails to account for the inhibition effects observed with naming times in response to delayed, end-of-sentence nonreferent probes. This result, as well as the observation of slowed reading following nonreferents in end-of-sentence regions, is consistent with a model in which inhibition of nonreferents is the result of more delayed, strategic, or controlled processing.

According to such an account, the nonreferent may become less accessible only after a delay, perhaps because resolution of the anaphor takes time. This would be consistent with the naming-time data of Experiments 1 and 3, as well as with Corbett's (1984) work on categorical anaphors and distractors. However, the fact that inhibition effects in Experiment 4 required not only a delay, but the termination of the sentence as well, indicates that inhibition requires an integration process prompted by sentence wrap-up. It appears that nonreferents may be inhibited only after an initial discourse representation has been constructed. As mentioned previously, Gernsbacher and St. John (1994) proposed a similar account of suppression as a delayed effect in a recent discussion of lexical ambiguity effects. They hypothesized that inhibition for nonselected meanings of ambiguous words occurs only after a sentence-level Gestalt has been computed.

The present results clearly indicate that when inhibition effects are observed for nonreferents, they reflect inhibition as a function of delayed, controlled, or strategic processing, prompted by discourse integration or decision-based probe tasks, and are not the direct result of immediate, automatic suppression of activation of nonreferent concepts during antecedent selection. In addition, the present findings have important implications for all researchers interested in inhibition effects. Although it is important to note that the present results do not rule out the possibility of a suppression mechanism, either as an explanation for anaphor resolution in particular or as a more general cognitive process, the present experiments point out that demonstrations of inhibition are not necessarily the result of suppression. Like researchers in the selective attention paradigm, we have found that some inhibition effects may result from the use of probe tasks that require decisions. Response conflict and its effect on decision times can be mistaken for more basic changes in availability of a stimulus. The present results as well as those of Long et al. (1999) show that this has happened in discourse processing studies, and reinforce the need to provide direct evidence when changes in activation are postulated by theories, whether of selective attention, semantic priming, or discourse processing.

References


Appendix A
Preliminary Experiments

In adapting the Lucas et al. (1990) materials, we changed the categorical anaphora that were used. Lucas et al. used widely varying modifiers, some comparative (e.g., larger or smaller) and some that were arbitrary (e.g., red book or blue book). We created a more homogeneous set of adjectives that were more semantically related to one referent or the other as in Sentences la and lb:

Billy's parents wanted him to take music lessons.
The girl wanted him to learn the violin or the drum.
(la) He finally chose the stringed instrument.
(lb) He finally chose the percussion instrument.

In the first experiment we followed Gernsbacher's (1989) methodology, and measured response times to recognition probes, in this case violin, presented before and after the anaphoric phrase (indicated by ^). Recognition responses were significantly slower for nonreferent probes after the anaphoric phrase than before it, $F_1(1, 31) = 33.5, MSE = 18,208, p < .01; F_2(1, 23) = 40.9, MSE = 2,578, p < .01$, indicating that responses to nonreferent probes were inhibited.

In the second preliminary experiment, there were four versions of each passage, all introduced by the same lines; for example:

Billy's parents wanted him to take music lessons.
They wanted him to learn the violin or the drum,
(2a) Referent. and the stringed instrument
(2b) Nonreferent. and the percussion instrument
(2c) Specific Control. and a wind instrument
(2d) General Control. and any instrument

In this example, the probe word presented for recognition immediately after "instrument" was violin. We labeled sentence 2c a specific control because it specified a subset of the category. If the inhibition effect observed by Lucas et al. (1990) in sentences like 2b reflected an incompatibility between the referent and probe word, we reasoned that the effect should also be present when the probe followed Sentence 2c. Sentence 2d provides an additional control for semantic priming and compatibility checks. In this experiment, we probed after the categorical noun ("instrument") and prior to the end of a sentence. The mean recognition time for the nonreferent condition was slower than either the mean for specific control, $F_1(1, 31) = 7.3, MSE = 3,576, p < .01; F_2(1, 23) = 4.17, MSE = 7,530, p < .05$, or the general control, $F_1(1, 31) = 16.7, MSE = 3,595, p < .01; F_2(1, 23) = 3.09, MSE = 10.355, p < .09$, indicating that responses to nonreferent probes were inhibited.

Appendix B
Example Filler Passages for Experiment 1

**Similar filler passage (Two referents, categorical noun anaphor, "no" response to probe):**
Bob needed a way to travel the mile to school. He considered a motorcycle or an automobile and the larger vehicle, the smaller vehicle

**MUSIC**

**Nonsimilar filler passage (Early probe):**
The manager of the Wildcat baseball team wanted.

**EGGPLANT**

**Nonsimilar filler passage (Late probe):**
The old rancher was dying when he made a new will leaving everything to his son and nothing to his daughter. The girl was furious and decided to contest the will. She hired a lawyer who had an excellent record in civil suits of this sort. He immediately.

**DYING**

Received June 14, 2000
Revision received January 3, 2001
Accepted January 22, 2001