

Levels of representation in the interpretation of anaphoric reference and instrument inference

MARGERY M. LUCAS

Wellesley College, Wellesley, Massachusetts

and

MICHAEL K. TANENHAUS and GREG N. CARLSON

University of Rochester, Rochester, New York

The level of representation accessed when inferences are made during sentence comprehension was examined. The inferences investigated included antecedent assignment for both definite noun phrase anaphors and pronouns and also instrument inferences. In making these inferences, a listener must access the inferred element, whether an antecedent or an instrument, in either a linguistic form representation or a discourse model. The level of representation involved in these inferences was determined by exploiting differences in the lexical decision and naming tasks, which were argued to exhibit differential sensitivity to representational levels. In three experiments, the priming of antecedent and instrument targets in the lexical decision task was compared with priming of the same targets in the naming task. Differences in the patterns of activation across the two tasks indicated that all three types of inferences required accessing elements in a discourse model. Three control experiments ruled out simple context or congruity checking as an explanation for our results. The following conclusions were also supported by these studies: (1) Antecedent assignment occurs immediately after processing an anaphor; (2) antecedent assignment involves inhibition for the inappropriate antecedent rather than facilitation for the appropriate antecedent; (3) although subjects do not make instrument inferences when they hear isolated sentences containing verbs that strongly imply certain instruments, the inferences are made when sentences are preceded by a context that mentions the instrument.

Understanding sentences in discourse involves integrating information in the sentence currently processed with information in a memory representation of the previous discourse. Sometimes the need for integration is explicitly marked, as when a sentence contains an anaphoric expression. Often it is less explicit (and necessary for maintaining coherence), as when actions mentioned in the sentence imply agents or instruments that were mentioned earlier in the discourse. A great deal of psycholinguistic research has been devoted to investigating how often, at what point in the sentence, and under what circumstances integrative processes occur. A basic question that has received relatively little attention, though, is that of the form of the representation involved in integrative processes. Although a common assumption is that integrating an element in the currently processed sentence requires changing the activation of an element in a memory representa-

tion of prior discourse, little experimental work has been done to investigate the form of the representation that is accessed (but cf. Cloitre & Bever, 1989; Murphy, 1985). In this article, we examine the types of representations used in accessing antecedents of definite noun phrases and pronouns as well as implied instruments. We also explore the conditions under which these inferences are made, and their time course.

The two most natural candidates for levels of representation in which prior information might be found are a representation of the surface form of a sentence and a more abstract discourse representation. One of the most robust results in the psycholinguistic literature is that memory for linguistic form is relatively transient, in contrast with memory for content or meaning (Bartlett, 1932; Johnson-Laird & Stevenson, 1970; Sachs, 1967; Wanner, 1974). Accordingly, most models of discourse processing distinguish between recovering the linguistic form of a sentence and developing a discourse model. The linguistic form representation includes the results of processing the words and the syntactic structures relating them, and it allows information about lexical form and syntax to be recalled for a short time. The discourse model includes the set of entities and relationships among them that are evoked by the discourse.¹ Discourse models preserve little or no information about linguistic form.

The linguistic literature on anaphora suggests that both of these types of representations might be used in dis-

The research presented here was supported by National Science Foundation Grants BNS-8217378 and BNS-8617738 to Michael Tanenhaus and Greg Carlson and by a Wellesley College Faculty Award to Margery Lucas. We are grateful to Robert Peterson and several anonymous reviewers for their many helpful comments on an earlier version of this paper. We also thank Cindy Nichols and Marlon Housman for their help in conducting the research and Sharon Utakis for her help in analyzing the data. Special thanks to Robyn Vallejos for her yeoman efforts in running the control experiments. Send reprint requests to Margery M. Lucas, Wellesley College, Department of Psychology, Wellesley, MA 02181.

course processing. Hankamer and Sag (1976) present evidence that anaphors can be divided into two classes, *deep* and *surface* anaphors. Surface anaphors, which include many types of verb phrase anaphora, are sensitive to the linguistic form used to introduce the referent of the anaphor. Thus there are syntactic constraints on the form of the antecedent for a surface anaphor and the antecedent must be introduced linguistically (for details see Garnham, 1986; Hankamer & Sag, 1976; Sag & Hankamer, 1984; and Tanenhaus & Carlson, in press). In contrast, deep anaphors, which include definite pronouns and definite noun phrase anaphora, are not similarly sensitive to the linguistic form of their antecedents. Deep anaphors can be used deictically to refer to salient objects or events in the environment. Deep anaphors also do not place constraints on the form of the antecedent when the antecedent is introduced linguistically. These differences between surface and deep anaphors can be explained by assuming that interpreting a surface anaphor requires the listener or reader to consult the linguistic form of the antecedent of the anaphor, whereas interpreting a deep anaphor requires consulting only discourse representations (Sag & Hankamer, 1984).

The view that definite pronouns require access to a discourse model but not to a linguistic level of representation appears open to immediate empirical challenges, for there appear to be syntactic effects on antecedent assignment. One type of constraint is most definitely linguistic in nature, and only affects pronoun-antecedent relations in the same sentence. For instance, in a sentence like "She thinks that Mary should be invited," the pronoun "*she*" cannot be interpreted as referring to Mary, since the syntactic relationship between the noun phrases "*she*" and "*Mary*" precludes making this assignment. However, the most common linguistic view is that although the syntactic relationship serves to rule out certain potential antecedents, the actual selection of the antecedent must be accomplished at a nonlinguistic level (Lasnik, 1976; but also see Reinhart, 1983, and Kuno, 1987, who argue that these constraints may be pragmatic rather than syntactic).

Another type of constraint that might appear to require the accessing of the form of the antecedent is that pronouns agree with their antecedents in gender and number. However, syntactic number typically corresponds to conceptual or notional number. In cases in which there is a mismatch between syntactic and notional number, agreement is typically with notional number. Gernsbacher (1986) contrasted three discourse situations in which the true antecedent was a multiple, collective, or generic item and found that in those situations subjects rated sentences with a mismatched plural pronoun more natural than those with a matched, singular pronoun. An example is the short discourse, "I think I'll order a margarita. I just love them," where the true antecedent is not the literal antecedent, which is singular, but a multiple item—margaritas in general.

A number of studies also suggest that memory for syntactic form is not necessary for interpreting pronouns and definite noun phrase anaphors. Memory for syntactic form

decays rapidly (Sachs, 1967; Wanner, 1974). If syntactic form were necessary for interpreting an anaphor, then comprehension should become increasingly difficult as the distance between the antecedent and the anaphor increases. However, discourse factors such as focus and referential continuity affect the ease of interpreting these anaphors more than distance (Anderson, Garrod, & Sanford, 1983; Ehrlich & Johnson-Laird, 1982; Lesgold, Roth, & Curtis, 1979). In fact, distance effects are eliminated when focus is controlled for both definite noun phrase anaphors (Lesgold et al., 1979) and definite pronouns (Anderson et al., 1983).

Although this evidence suggests that deep anaphors access a discourse model, in most studies processing-load measures such as reading time have been used, which are global measures of ease of comprehension and are not sensitive to the level of representation. An on-line measure of antecedent assignment would provide more precise and accurate information, especially if it could also distinguish among the levels of representation.

Two on-line processing measures reported in the word recognition literature as having been used successfully are lexical decision and naming. At one time, it was generally assumed that these two tasks provided similar information about word recognition (Meyer & Schvaneveldt, 1976). For example, in both tasks there are faster responses to words preceded by semantically or associatively related words than to words preceded by unrelated words, indicating that both tasks are sensitive to the activation of items in semantic memory. But recently, a number of interesting differences between the tasks have been discovered. Lexical decisions, but not naming responses, are facilitated by syntactic relationships between words, backward associations (i.e., target-to-prime associations), and changes in proportion of related to unrelated items (these differences are described in more detail in Seidenberg, Waters, Sanders, & Langer, 1984). Lexical decisions are also affected by a number of factors related to the meaning of the word, such as the likelihood that the target will be given as a response to a category name and the number of dictionary meanings it has; in contrast, naming seems to be relatively insensitive to the same factors (Balota & Chumbley, 1984; Chumbley & Balota, 1984). Also, inhibition for contextually incongruent targets has been observed primarily with the lexical decision task, and not with naming (West & Stanovich, 1982; but cf. West & Stanovich, 1986, on inhibition in naming in syntactically inappropriate contexts).

A reasonable explanation for these differences appears to be that naming is primarily sensitive to lexical processes and lexical decision to both lexical and postlexical processes. This may be because lexical decisions require one to make a conscious decision as quickly and as accurately as possible. Forster (1976, 1979) suggested that under such circumstances, a simple strategy of matching the orthographic characteristics of the visually presented word with an internal representation may be time-consuming and inaccurate, causing the lexical decision targets to undergo postaccess checks for error. Forster

argues that an effective error-checking mechanism for a discriminatory task would involve a different format from the one involved in the initial match. In a typical lexical decision task, this means using message-level processes to integrate the target with its context. Successful integrations support *yes* decisions, failures support *no* decisions. Naming, in contrast to lexical decision, does not require a conscious *yes/no* decision. Instead, it requires the accessing of procedural knowledge directly tied to the lexical representation of the word. Knowing whether or not a word makes sense in context is unlikely to facilitate this process, which may explain why naming has been shown to be less sensitive to contextual integration effects. Although there is little disagreement that lexical decision is more sensitive to such effects than naming, the differences in the two tasks are not absolute. Keefe and Neely (in press) and Norris (1987), for example, have found proportionality effects with naming under certain circumstances. Nonetheless, the generalization that lexical decision is more sensitive than naming to postlexical processes does seem to hold across most studies that have been focused on differences.

Lexical decisions, then, seem to be the result of cognitive processes that cut across levels of representation, whereas naming seems primarily sensitive to representations of lexical form. This analysis suggests a possible way of distinguishing between the levels of representation associated with anaphoric processing, through the use of differences in patterns of responses in the lexical decision and naming tasks as a diagnostic test for determining representational level. The logic behind the use of the two tasks follows from two possibilities for the mechanism involved in antecedent assignment or any other integrative process: Integration requires changing the activation level of either the lexical form associated with the relevant discourse element in a linguistic level of representation or the activation level of the conceptual form of the relevant element in a discourse model. If the conceptual form is accessed, this will be reflected in lexical decisions; if the lexical form is accessed, this will be reflected in both naming time and lexical decision (given that pure associative or intralexical priming is eliminated as a possibility). Direct experimental evidence for the representational level or levels accessed in integrative processes would provide an important constraint on models of integration.

In the experiments presented here, the lexical decision and naming tasks were used to examine the processing of definite noun phrase anaphors, pronouns, and verbs with implicit instruments. In Experiments 1A-1C, definite noun phrase anaphora were examined; in Experiments 2A-3D, pronouns and instruments were examined.

EXPERIMENT 1 Definite Noun Phrase Anaphors

In this set of experiments, we examined two hypotheses. The first involved the time course of antecedent assign-

ment. The assumption behind the use of priming tasks to study antecedent assignment is that once antecedents are selected, they will be more activated than other discourse entities, including entities that are potential antecedents. This may be because the actual antecedent will be more highly activated or because the potential antecedents will be somewhat inhibited. Previous research has demonstrated the validity of using priming tasks in the study of anaphoric processing, but these have mostly been probe recognition studies (Cloitre & Bever, 1989; Corbett & Chang, 1983; Dell, McKoon, & Ratcliff, 1983; Gernsbacher, 1989; McKoon & Ratcliff, 1980; Walker & Yekovich, 1987; but cf. Nicol & Swinney, 1989, for a review of a few studies using the cross-modal priming task). These have shown that the actual antecedent is activated more than other potential antecedents at some point after the anaphor (usually a pronoun) has been processed. However, there is no difference in activation among potential antecedents either before or immediately after the pronoun has been read (Gernsbacher, 1989). In the first two experiments here, two different priming tasks were used to measure the activation of antecedents of definite noun phrase anaphors. Subjects listened to sentence pairs like the following:

1. Sarah could not decide whether to buy steak or hamburger.
She finally chose the more expensive meat.

Subjects then made either lexical decisions or naming responses to visual targets that were appropriate or inappropriate antecedents. On the basis of the probe recognition studies, we expected to find that both antecedents would be activated immediately after processing the anaphor, but that only the actual antecedent would be activated sometime later.

The second hypothesis involved representational issues. We expected, on linguistic grounds, that the definite noun phrase in the second sentence—"the more expensive meat"—would be linked to an entity in a discourse model rather than a lexical representation for "steak." As discussed earlier, deep anaphors, including definite noun phrases, are not sensitive to the linguistic form of the antecedent. Also, there need not be a linguistic expression to act as antecedent in order for a definite noun phrase to be understood. For example, if a speaker who is looking at two pieces of meat says, "I'd like the bigger piece," it will be apparent from following the direction of the speaker's gaze (and the size of the meat) which piece the speaker is referring to.

We would expect, then, to find an appropriateness effect with the lexical decision task, because lexical decision is sensitive to both discourse level and lexical form. But we would expect to see an effect in naming only if lexical form is involved. Thus, by comparing the two tasks, we should be able to establish the level or levels of representation involved in anaphoric processing.

Experiment 1A

Method

Subjects. Forty-eight Wayne State University undergraduates participated for course credit.

Materials and Design. Three sentence lists contained 24 experimental sentence pairs. Examples of sentence pairs consist of 2, followed by a, b, or c:

2. Mary couldn't decide whether to buy steak or hamburger.
 - a. She finally decided to buy the more expensive meat.
 - b. She finally decided to buy the less expensive meat.
 - c. The store had a wide selection of foods.

The last noun phrases in Sentences 1a and 1b are definite noun phrase anaphors that refer to one of two possible antecedents in Sentence 2—in this example, “steak” or “hamburger.” Sentences 2a and 2b differ only in one word that would make the antecedent target appropriate or inappropriate. Knowledge-level processing establishes which is the correct antecedent—in the present case “steak,” because it is usually more expensive than hamburger. Sentence 2c is an example of a nonanaphor control. Lexical decision targets were always just one of the possible antecedents; in this case, “steak” makes Sentence 2a the appropriate condition, and Sentence 2b the inappropriate condition.² The three conditions for each sentence pair (appropriate, inappropriate, and control) were counter-balanced across lists, and each subject saw just one list. The full list of experimental sentences may be found in Appendix A. There were also 47 filler sentence pairs, for a total of 71 sentence pairs. Half of these were nonwords that provided “no” responses for the lexical decision task. They were formed by altering one letter of an actual word to make a pronounceable pseudoword.

Procedure. Subjects were randomly assigned to one of the three sentence lists and one of three target-delay conditions: 0, 600, and 900 msec. Three target delays were chosen, because, in this first experiment, the point at which the effects of antecedent assignment would be noticeable was unknown. The subjects listened to binaurally presented sentences over headphones and made lexical decisions to target words presented at the end of the second sentence. The visual targets were displayed on an Apple IIe computer screen. The computer was signaled by a 1000-Hz tone to present the target. The tone appeared on the second channel of the sentence tape at a point that corresponded to the end of the last word in the sentence. The subjects heard only the first channel of the tape, which contained the sentences. The channel containing the 1000-Hz tone was output to a relay device, which simultaneously started a Digitry millisecond clock inside the computer and presented the visual target onscreen immediately, 600 msec, or 900 msec later. Targets remained onscreen until the subject pushed one of two response buttons. The subjects were told to listen to the sentences carefully, and this was encouraged by following one third of the sentences with a yes/no question, to which the subject was instructed to respond.

Results and Discussion

For each subject and condition, outlier scores greater or less than 2.0 standard deviations from the subject mean were replaced by the 2.0 standard deviation cutoff score. Mean reaction times for all conditions are reported in Table 1. Table 1 also includes the percentage of errors for each condition. There were 26 errors out of a possible 1,152 responses (2.3%). Errors occurred either when a subject made an incorrect lexical decision or when there was a failure to respond in less than 2 sec.

Analyses of variance with subjects and items as random factors were performed on the reaction time data and are reported here as F_1 and F_2 , respectively. There was a significant main effect of prime type (appropriate, inappropriate, and control) in both subject and item analyses [$F_1(2,90) = 14.46$, $MS_e = 7,484$, $p < .0001$, and $F_2(2,46) = 6.36$, $MS_e = 25,488$, $p < .005$]. The main effect of delay (0, 600, and 900 msec) was significant only in the item analysis [$F_1(2,45) = 1.68$, $MS_e = 116,373$, and $F_2(2,46) = 19.95$, $MS_e = 14,646$, $p < .0001$], and the interaction between delay and prime type was not significant in either analysis [$F_1(4,90) < 1$, $MS_e = 7,484$, and $F_2(4,92) < 1$, $MS_e = 28,503$].

An examination of the means shows that the main effect of prime type was due largely to inhibition for inappropriate targets. This appears even in the 0-msec delay, indicating that listeners begin to make the anaphoric inference as soon as the anaphor is encountered. The inhibition is still strong at 600 and 900 msec. This interpretation is supported by paired comparisons which show that reaction times in the inappropriate conditions differed significantly from those in the control conditions and in the appropriate conditions, but that appropriate conditions did not differ from the controls.

Before the results of the comparisons are reported, however, it should be noted that item comparisons were performed twice. In the first set of item comparisons, item means were collapsed across the three delay conditions, yielding 24 items in each comparison (reported here as F_2). This is commensurate with the treatment of delay as a within-item factor, but it is possible that this test was contaminated by differences in variance across the delay conditions, which were introduced by the use of different groups of subjects at these delays. So, in a second set of comparison tests, the items in different delays were treated

Table 1
Mean Lexical Decision Times to Targets (in Milliseconds) and Percentages of Error for Experiment 1A

Condition	0 msec			600 msec			900 msec		
	<i>M</i>	FS	% Error	<i>M</i>	FS	% Error	<i>M</i>	FS	% Error
Appropriate	859	-36	0.0	715	27	.1	859	0	.1
Inappropriate	917	-94	.9	834	-92	.6	924	-65	.4
Control	823		.1	742		.1	859		.1

Note—“Condition” refers to the relationship between the definite noun phrase anaphor and the lexical decision target. FS = facilitation scores. These were calculated by subtracting reaction times in the appropriate or inappropriate conditions from reaction times in the nonanaphor control conditions. Positive numbers indicate facilitation. Negative numbers indicate inhibition.

as separate items, yielding 72 items in each comparison (reported here as F_3). The results of the two comparisons are essentially the same; nonetheless both are reported here.

The inappropriate condition differed significantly from the control condition [$F_1(1,47) = 23.08$, $MS_e = 7,333$, $p < .0001$, $F_2(1,23) = 8.52$, $MS_e = 11,113$, $p < .01$, and $F_3(1,71) = 10.12$, $MS_e = 25,050$, $p < .005$] and from the appropriate condition [$F_1(1, 47) = 15.77$, $MS_e = 9,831$, $p < .0001$, $F_2(1,23) = 4.62$, $MS_e = 13,932$, $p < .05$, and $F_3(1,71) = 7.91$, $MS_e = 29,414$, $p < .01$]. The appropriate condition did not differ significantly from the control condition (all $F_s < 1$).³ The pattern of errors is congruent with the reaction time analysis. There were errors on only 2.3% of the trials for a total of 26 out of 1,152 responses. Most of these (21) were in the inappropriate condition, with only 2 in the appropriate and 3 in the control conditions, indicating that the inappropriate conditions were indeed more difficult for subjects. This pattern of inhibition for only the inappropriate targets could not have been due solely to inhibition from the final word of the sentence to the target, because in both appropriate and inappropriate conditions the final word and target are the same (in our example, "meat-steak").

One somewhat surprising finding was the evidence, in the 0-msec delay condition, that the process of antecedent assignment begins immediately upon encountering the anaphor. Gernsbacher (1989) found no differences in activation of potential antecedents either before or immediately after the anaphor, indicating that antecedent assignment had not begun at that point. But Gernsbacher used pronouns with ambiguous referents. To use an example from her study, in the sentence "Ann predicted that Pam would lose the track race, but she came in first very easily," "she," at the point at which it is first encountered, could refer equally plausibly to either Pam or Ann. Under these conditions, there is no activation or inhibition of either potential antecedent. Our stimuli, in contrast, are definite noun phrases that express an explicit comparison. At the end of the anaphor, only one of the potential antecedents is a plausible referent. Under these conditions, antecedent assignment is made immediately. The results are consistent with Gernsbacher's, in the finding that antecedent assignment takes the form of inhibiting or suppressing the inappropriate antecedent rather than reactivating or enhancing the appropriate antecedent.

The results provide some evidence for an antecedent assignment effect in lexical decision. To provide more support for this interpretation of the data, however, some concerns about the materials need to be addressed. In some of the stimulus sentence sets, the target is consistent with both the appropriate and the control sentences but not the inappropriate sentence. In the example sentence, "steak" is compatible with "more expensive meat" and "wide selection of foods," but not with "less expensive meat." This suggests an alternative explanation of the lexical decision results. What looks like inhibition in the inappropriate condition may be due to the outcome of a postaccess

check in which the congruity of the target is compared with the immediately preceding words or phrases. That is, subjects might attempt to integrate the lexical decision target with the preceding information, in effect drawing the inference only after the target was presented (a kind of backward priming effect that does not reflect the state of the discourse model). The outcome of this postaccess congruity check would facilitate decisions in both the appropriate and control conditions, accounting for the comparatively slow responses in the inappropriate condition. In fact, the inhibition-dominant pattern of responses found here is typical of contextual congruity judgments found elsewhere with the lexical decision task (Fischler & Bloom, 1979; West & Stanovich, 1982). It is unlikely that a congruity effect is involved in this experiment, however, given the nature of the inference required to determine that the target and the words or phrases preceding it could be integrated. In all test sentences, the definite noun phrases involved a comparative term (as in "the *more* expensive meat"). A simple integration strategy under these circumstances would not result in selective inhibition for inappropriate antecedents unless anaphora resolution had already occurred and the results of that process became part of the input to the lexical decision. Nonetheless, to eliminate the possibility of a postaccess congruity check, in the following experiments, the second sentences were presented alone, without the context sentence. If the results of the lexical decision experiment were due to simple congruity checking, the same pattern of results should obtain.

Experiment 1B

Method

Subjects. Twenty-four University of Rochester undergraduates were paid \$6 for their participation.

Materials and Procedure. These were the same as in Experiment 1A, except that context sentences were eliminated in each list. The subjects heard only the second sentence of each sentence pair. Because the inhibition effect was found at all three delays, only the 0-msec delay was used.

Results and Discussion

Mean reaction times and percentage errors for all conditions are reported in Table 2. Although it looks as if there is slight facilitation for both appropriate and inappropriate targets, one-way analyses of variance indicated that there was no appropriateness effect [$F_1(2,46) < 1$, $MS_e = 2,893$, and $F_2(2,46) < 1$, $MS_e = 3,827$]. Therefore, the inhibition for inappropriate targets found in Experiment 1A is not likely to have been due to a simple

Table 2
Mean Lexical Decision Times to Targets (in Milliseconds) and Percentages of Error for Experiment 1B

Condition	Reaction Time	% Error
Appropriate	995	.2
Inappropriate	1006	.7
Control	1017	.7

congruity check that resulted from the requirements of the lexical decision task. Not only was inhibition not found with the target sentence alone, the pattern of the means, if anything, was somewhat facilitatory.

The last two experiments indicate that the lexical decision task is sensitive to antecedent assignment. In the next experiment, the same materials were used, but with the naming technique. If antecedent assignment involves the lexical form of the antecedent, an appropriateness effect should also be obtained in the naming task. If, however, anaphoric processing requires accessing antecedents in a discourse representation only, there should be no appropriateness effect with naming.

Experiment 1C

Method

Subjects. Thirty-two University of Rochester undergraduates participated for course credit.

Materials and Procedure. These were the same as those in Experiment 1A, except that subjects were asked to say the visual target word out loud, and the time to initiate the naming response was recorded. Because the 900-msec delay condition in Experiment 1A yielded results very similar to the 600-msec condition, it was dropped from Experiment 1C. Nonword stimuli were replaced by word stimuli in the filler sentences.

Results and Discussion

Mean reaction times and percentages of error for all conditions are reported in Table 3. The error rate was 2% and all errors were due to failures of the subject to activate the voice relay.

Analyses of variance revealed a main effect of prime type (appropriate, inappropriate, and control) in the subject analysis [$F_1(2,60) = 6.54$, $MS_e = 1,182$, $p < .005$, and $F_2(2,46) = 1.65$, $MS_e = 7,136$, n.s.]. The main effect of delay (0 and 600 msec) was significant in the item analysis [$F_1(1,30) = 2.01$, $MS_e = 25,325$, n.s., and $F_2(1,23) = 21.67$, $MS_e = 3,502$, $p < .0001$], the result of a slower overall response rate in the 0-msec delay condition. There was no interaction between prime type and delay in either analysis [$F_1(2,60) < 1$, $MS_e = 1,182$, and $F_2(2,46) < 1$, $MS_e = 3,973$].

The pattern of means indicates that there was some facilitation for both appropriate and inappropriate antecedents at both delays. Planned comparisons provide some support for this interpretation. Item analyses were again performed twice for the reasons given in Experiment 1A. Appropriate conditions differed significantly from con-

trol conditions in two out of the three comparisons [$F_1(1,31) = 22.33$, $MS_e = 678$, $p < .0001$, $F_2(1,23) = 2.52$, $MS_e = 4,067$, n.s., and $F_3(1,47) = 3.92$, $MS_e = 5,875$, $p = .05$]. The inappropriate and control comparison was marginally significant in the subject analysis [$F_1(1,31) = 3.81$, $MS_e = 1,562$, $p = .06$, $F_2(1,23) < 1$, and $F_3(1,47) = 1.59$, $MS_e = 5,648$, n.s.]. Appropriate conditions did not differ significantly from inappropriate conditions [$F_1(1,31) = 1.71$, $MS_e = 1,229$, n.s., $F_2(1,23) = 1.76$, $MS_e = 3,062$, n.s., and $F_3(1,47) < 1$, $MS_e = 3,243$].⁴

Discussion: Experiments 1A, 1B, and 1C

At this point, it is useful to compare the results of Experiments 1A, 1B, and 1C. The lexical decision reaction times exhibit inhibition for inappropriate items across all delays, the naming times facilitation for both appropriate and inappropriate items. The results of the lexical decision task alone indicate that the subjects were accessing either a linguistic form representation or a discourse model (or both) to make anaphoric inferences. The results of the naming task should clarify the issue. An appropriateness effect in the naming task would indicate that a lexical representation is involved in anaphoric inference. If that were the case, two possibilities would still remain. One possibility is that antecedent assignment involves accessing a lexical item in the linguistic form representation exclusively, which seems unlikely, given the linguistic evidence cited earlier. A more likely possibility is that it involves accessing both a linguistic form and a discourse representation, the linguistic form representation perhaps being used as an aid in locating the relevant entities in the discourse representation.

The results of Experiment 1C indicate that there was an appropriateness effect in the naming task, but not of the kind to support the interpretation that a lexical form is involved in antecedent assignment. This is because there appeared to be facilitation of *both* appropriate and inappropriate antecedents at both delays. Note that this pattern of facilitation for both targets is similar to the pattern found in Experiment 1B with lexical decisions. In that experiment, targets were presented after the second sentence alone. Without the discourse context, the results of the lexical decision task changed from one showing inhibition of the inappropriate antecedent to one showing no significant change in activation for potential or actual antecedents (except some slight numerical facilitation). The fact that naming results in a discourse context look like lexical decision results without context suggests that the priming in Experiment 1C was due to semantic or associative priming from the noun at the end of the sentence to the naming target (e.g., *meat-steak*), a lexical effect that is tangential to the kinds of discourse processes that are the concerns of this paper. The results of Experiments 1A and 1B combined, then, indicate that a discourse model but probably not a linguistic form representation is involved in anaphoric inference for definite noun phrase anaphors.⁵

Table 3
Mean Naming Latencies to Targets (in Milliseconds) and Percentages of Error for Experiment 1C

Condition	0 msec			600 msec		
	M	FS	% Error	M	FS	% Error
Appropriate	566	33	0.0	526	29	.7
Inappropriate	585	14	.1	531	24	.4
Control	599		.4	555		.4

Note—"Condition" refers to the relationship between the definite noun phrase anaphor and the naming target. FS = facilitation scores, calculated as described in Table 1.

The results of Experiments 1A-1C support two points. First, antecedent assignment for definite noun phrases occurs rapidly. Although there is a slight possibility that antecedent assignment is occurring even before the end of the phrase, it is clear from the results that the process is completed by the end of the phrase. This process takes the form of inhibiting the inappropriate antecedent rather than reactivating or providing additional activation to the appropriate antecedent. The results also indicate that lexical representations are not involved in antecedent assignment. This was shown by using the lexical-decision-naming-task difference to distinguish between levels of representation. Once lexical form is eliminated as a possibility, a reasonable inference is that antecedents are accessed in a discourse model.

The remaining experiments reported here were carried out to explore the time course and constraints on other forms of sentential inference as well as to extend the methodology developed in the first set of experiments to the problem of determining the level of representation involved in those inferences.

EXPERIMENT 2

Instrument Inferences

Another interesting class of inferences has been identified in both schema theory (Rumelhart, 1975; Schank, 1975; Schank & Abelson, 1977) and case theory (Fillmore, 1968, 1971) as involving knowledge of the tools or instruments typically used to accomplish some action. For example, on hearing the sentence, "Jack cut into a thick juicy steak," one is likely to infer that Jack used a knife. But, unlike in the case of anaphoric inference, it is an open question whether or not such inferences are drawn routinely in the course of normal sentence comprehension. Before the nature of the representation involved in such inferences can be investigated, it will be necessary first to determine whether or not they are an obligatory part of the comprehension process.

An assumption in both schema theory and case theory is that instrument inferences, as well as any other inferences involving the cause and effect of actions, are necessary for comprehension and will be part of the final discourse representation. Empirical support for this assumption has been mixed, however. Although several studies have provided evidence that instrument inferences are encoded and stored as part of the memory representation of a sentence (Johnson, Bransford, & Solomon, 1973; McKoon & Ratcliff, 1981; Paris & Lindauer, 1976; Singer, 1980), in others this has not been found to be the case (Corbett & Doshier, 1978; Doshier & Corbett, 1982; Singer, 1979, 1980). One reason for these discrepancies may lie in the difference in methodologies and design across studies. Paris and Lindauer (1976) and Corbett and Doshier (1978) both used cued recall as a measure of encoding, but Corbett and Doshier showed that retrieval cue effectiveness is a questionable measure of memory en-

coding. They found that high-probability instrument cues were effective not only for sentences that explicitly mentioned the high-probability instrument but also for sentences that explicitly mentioned a low-probability instrument. They concluded that the effect was mediated by preexperimentally learned relationships.

Johnson et al. (1973) and Singer (1979, 1980) used recognition rate or verification latency. Johnson et al. (1973) and Singer (1980) found that instrument inferences were encoded, Singer (1979) that they were not. This inconsistency may be due to differences in the discourses used. Singer (1980) noted that, although inferences are often necessary for comprehension, the number of inferences that could be drawn for a particular sentence is potentially quite large. Presumably, limits on processing capacity will restrict the number of actual inferences to some small subset of those possible, and the inferences that would contribute to the coherence of the discourse would have priority. In fact, Singer (1980) showed that when an inference is unnecessary for preserving coherence, as were the inferences in Singer (1979), sentence verification times for those inferences were slower than those for directly expressed ideas. When the inferences were necessary for preserving coherence, however, results were comparable to those for directly expressed ideas, indicating that necessary inferences do become a part of the memory representation of a sentence. McKoon and Ratcliff (1981) came to a similar conclusion. Using a probe recognition task, they found that instrument inferences were activated after subjects read sentences that contained actions to which the instruments were highly related. These actions and inferences continued to be related in the long-term memory representation of the text. Actions and instruments that were not highly related did not show evidence of activation. The authors concluded that instrument inferences are not always necessary for comprehension. Activation of instruments occurs only when subjects find it necessary to make the inference or explicitly set themselves the goal of making the inference.

The studies just cited reveal the form of the final memory representation, but it is also important to consider processes as they occur during sentence comprehension. Doshier and Corbett (1982) also point out that even if instrument inferences are not always encoded in memory, it is possible that they are, nonetheless, briefly activated as part of the schema accessed by the verb. This activation may be insufficient to make the inference part of a permanent memory representation, unless, as Singer (1980) and McKoon and Ratcliff (1981) found, it is necessary for the coherence of the discourse. Also, the methodologies used, whether cued recall, verification, or probe recognition, force the subject explicitly to consider the instrument in relation to the discourse. This may not reflect, then, what normally happens during comprehension. An on-line measure of inference encoding that does not require an explicit recall or recognition decision and that would provide information about comprehension processes before the final product is stored in memory

would seem to be a necessary complement to the previous work.

Applying this logic, Doshier and Corbett (1982) used Stroop interference to measure the activation of instruments. They presented target instruments either at the end of simple sentences containing predicates that implied the use of certain instruments or at the end of control sentences that did not have such predicates. They found no evidence that either the most probable or next most probable instrument was activated, although they did find some facilitation when subjects were explicitly instructed to generate the most likely instrument for each sentence. This last task is so far removed from normal sentence processing, however, that they argued that encoding implicit instruments is a redundant process that is not necessary for discourse processing.

Doshier and Corbett's (1982) conclusions need to be qualified in light of the research discussed above, which provides evidence that instrument inferences may occur when there is a context (McKoon & Ratcliff, 1981; Singer, 1980) and, particularly, when the instrument has already been explicitly mentioned in that context (McKoon & Ratcliff, 1981). This suggests a modification of the Doshier and Corbett (1982) study that involves measuring priming with and without context to determine if comprehension strategies do indeed differ in the two situations. The purpose in the next two studies, then, was to replicate the Doshier and Corbett findings with both lexical decision and naming in a cross-modal paradigm. This is a prelude to examining comprehension strategies when the same stimuli are preceded by a context that explicitly mentions the instrument, making it an element in the discourse representation. The use of contexts will occur in Experiments 3A and 3B, where the nature of the representational level used in such inferences will also be considered. But first, the conditions under which the inferences occur must be established.

Experiment 2A

Method

Subjects. Sixteen University of Rochester undergraduates participated in order to satisfy a course requirement.

Materials and Design. The experimental sentences were generated from 28 verb-instrument pairs in which the verb strongly implied an instrument used to carry out the action described by the verb. Half of these pairs were derived from Doshier and Corbett's (1982) stimuli. The other half were generated by the experimenters. The instruments were not explicitly mentioned in the sentences. An

example of an experimental sentence that has the implied instrument of "broom" is given below. The full set of materials is listed in Appendix B.

3. He swept the floor every week on Saturday.

Each sentence was paired with two targets—its corresponding instrument, or a control target that was a highly implausible instrument for the action specified by the verb. For example, targets for Sentence 3 were "broom," the appropriate instrument, and "closet," the inappropriate control. The controls were chosen to accommodate the next series of studies, in which both targets would be incorporated into a context sentence. For that purpose, both targets had to be potential antecedents, so both were nouns. The mean length for instruments was 5.1 letters, and for controls it was 5.4. The mean frequencies from the Kučera and Francis (1967) word frequency norms were 61.0 for instruments and 67.8 for controls. Two presentation conditions were produced by combining the sentence list with two target lists that counterbalanced type of target (instrument or control) across lists. We did not have a third, neutral control condition, because the results of Experiment 1A indicated that the effects of primary interest involve differences between the appropriate and inappropriate conditions. Eliminating the neutral control condition simplified the experiment without the loss of essential information for the purposes of this series of studies. The experimental sentences were intermixed with 92 filler sentences. Sixty of the filler trials were paired with nonword targets, and the remaining 32 with word targets.

Procedure. Eight subjects were assigned to each of the two presentation conditions. Each presentation version was preceded by 10 practice trials. The subjects heard the sentences binaurally over stereo headphones and were asked to make a lexical decision to a word appearing at an unspecified point in the sentence. In the experimental sentences, this point was at the end of the second phrase following the verb (but never at the end of a sentence), so that the subjects would have plenty of time to make the inferences before target presentation. A timing tone inaudible to the subjects was placed at the target point to initiate presentation of targets midscreen on an Apple IIe computer monitor. The subjects pressed a button to indicate their responses. If there was no response within 2 sec, the target was removed. Decision times were recorded by a Digistry millisecond timer from the onset of the target to the subject's buttonpress. In order to encourage the subjects to attend to the sentences, comprehension questions were asked following one third of the trials.

Results

Condition means and percentages of error are presented in Table 4. One-way analyses of variance with both subjects and items as random factors and appropriateness (plausible instrument or implausible control) as a fixed factor were run. The results of these analyses confirm what is also clear from an inspection of the means: there was no hint of an appropriateness effect [$F_1(1,15) = .01$,

Table 4
Mean Reaction Times (in Milliseconds) and Percentages of Error for
Experiments 2A (Lexical Decision) and 2B (Naming)

Type of Task	Type of Target			
	Appropriate (Plausible Instrument)	% Error	Inappropriate (Implausible Control)	% Error
Lexical decision	929	5.1	926	3.8
Naming	604	2.9	630	3.6

$MS_e = 8,223$, and $F_2(1,27) = .001$, $MS_e = 18,607$]. The results, then, replicate Doshier and Corbett's (1982) in that no evidence was found for instrument priming, this time using a cross-modal lexical decision task. In the next experiment, an attempt was made to replicate this finding with the naming task.

Experiment 2B

Method

Subjects. Sixteen University of Rochester undergraduates participated as subjects in order to satisfy a course requirement.

Materials and Procedure. These were identical to those in Experiment 2A, except that nonword targets were replaced by word targets. All words on filler trials were unrelated to the sentences with which they were matched. Instead of pressing a button for a lexical decision, the subjects said the target word out loud as soon as it appeared. Responses were spoken into a microphone, and naming times were measured from the onset of the target to the onset of the subject's spoken response.

Results

Mean reaction times and percentage of errors are reported in Table 4. All errors were due to the failure of the voice-activated relay to pick up the onset of the subject's responses. The analyses of variance found no main effect of appropriateness [$F_1(1,15) = 1.91$, $MS_e = 2,860$, n.s., and $F_2(1,27) = 1.64$, $MS_e = 4,985$, n.s.]. Together, Experiments 2A and 2B failed to reveal instrument priming for isolated sentences.

Discussion: Experiments 2A and 2B

The fact that there was no appropriateness effect in either experiment indicates that the implied instrument was no more accessible than an inappropriate control. The results, therefore, replicate those of Doshier and Corbett (1982) in that no evidence was found for instrument priming in sentences without context, using lexical decision and naming tasks instead of Stroop interference. In the next two experiments, a context in which the instrument was explicitly mentioned was provided for the sentences used in Experiments 2A and 2B. The hypothesis was that this would cause subjects to infer the use of the instrument.

EXPERIMENT 3

Instrument Inferences and Pronouns

One problem in designing the next two experiments was the possibility, given the results of Doshier and Corbett (1982) and Experiments 2A and 2B, of another null result. Doshier and Corbett dealt with this by performing a series of systematic replications. The present studies, with somewhat different materials and techniques, themselves constitute further replications of Doshier and Corbett. Still, it would be useful, as an additional check, to directly compare instrument inferences with another kind of inference to ensure that the measures and materials used were sensitive to complex discourse inferences. Antecedent assignment for pronouns seemed a likely candidate for this control. Although instrument inferences may not be necessary

for comprehension, antecedents must be assigned to pronouns if sentences are to be understood.

In addition to investigating whether instrument inferences are drawn when a context makes the instrument available, the experiments were designed to use the lexical-decision-naming difference to determine the levels of representation involved in both instrument inferences and antecedent assignment for pronouns. Earlier, it was proposed that instrument inferences could require access to either a linguistic form representation or a discourse model. The same is true of pronominal reference. Although pronominal reference, like the definite noun phrases that were the topic of Experiments 1A, 1B, and 1C, is classed linguistically as a deep anaphor, it is not obvious that assigning antecedents to pronouns involves immediately accessing a discourse model. Drawing inferences based on world knowledge is not as essential to initial antecedent assignment for pronouns as it is for definite noun phrases. This is because pronominal reference might first require choosing from a pool of candidates on the basis of some kind of linguistic agreement between the anaphor and previous noun phrases (e.g., gender and number agreement), information that would be preserved in a linguistic form representation. Later, a choice could be made between these candidates on the basis of semantic and pragmatic factors.

Corbett and Chang (1983) found support for this in a probe recognition study of antecedent assignment for pronouns. They showed that pronouns reinstated both the true antecedent and other noun phrases in the sentence that were acceptable syntactically. This was true even if the potential antecedents were implausible candidates, although gender-mismatched proper names do not appear to be reinstated (Chang, 1980). As noted earlier, however, the probe recognition task still had a decision component that explicitly requires the subject to note whether or not the probe target was in the discourse. These results, then, may reflect these decision processes rather than normal comprehension.⁶

To summarize, in Experiments 3A and 3B there were two goals. One goal was to discover whether instrument inferences would be made when sentences with implied instruments were presented following contexts that explicitly mentioned the instrument. Sentences that referred to the instrument by means of pronouns would serve as a check on the materials and design of the experiment should there be another null result for the instruments. The second goal was to use the lexical-decision-naming-task difference to reveal whether representations for implied instruments and pronoun antecedents were found in a linguistic form representation or in a discourse model. In Experiment 3A, the lexical decision task was used, in Experiment 3B, the naming task.

Experiment 3A

Method

Subjects. Twenty-four undergraduates from the University of Rochester participated to fulfill a course requirement.

Materials and Design. The sentences from Experiments 2A and 2B were used with two modifications. First, a context sentence was constructed for each of the experimental sentences. This context sentence contained both the instrument and control words used as targets in the previous studies. For example, the sentence "He swept the floor every week on Saturday" was given a proper name and a context, thus creating the following sentence pair 4:

4. There was a *broom* in the *closet*.
John swept the floor every week on Saturday.

The italicized words are instrument and control targets, respectively. These sentences constituted the instrument version of the study (see Appendix B).

A second modification involved the introduction of an additional set of sentences, based on the sentence pairs used in the instrument version, in which the second phrase following the verb was replaced by a phrase that contained a pronoun. The pronoun always referred to the instrument explicitly mentioned in the preceding sentence. For example, the second sentence of the instrument version given above would be changed in the pronoun version to "John swept the floor *with it* on Saturday."

Procedure. Four presentation conditions were formed. Two sentence lists, which counterbalanced pronoun and instrument versions of the second sentences in the 28 sentence pairs, were combined with the two target lists from Experiments 2A and 2B, so that pronoun or instrument version was a within-subject variable. Context sentences were always the same. The lists differed only in whether the second sentence did or did not include the pronoun. Six subjects were randomly assigned to each of the four presentation conditions. The procedure for this experiment was identical to that for Experiment 2A, except that in sentences containing pronouns, targets were presented at the end of the pronoun. Again, this gave the subjects enough time after having heard the verb to make the inference or antecedent assignment. Targets were never presented at the end of a sentence. Yes/no questions about the sentences were asked after one third of the sentence pairs, in order to ensure that the subjects were attending to the stimuli, but these responses were not analyzed or used in any way to eliminate subjects from the experiment.

Results

Mean lexical decision reaction times and percentages of error are reported in Table 5. Errors were 7.6% of the responses and were evenly distributed across conditions. Analyses of variance with subjects and items as random factors and appropriateness (plausible or implausible antecedent/instrument) and version (pronoun or instrument) as fixed factors were run on the lexical decision times. There was a main effect of appropriateness [$F_1(1,23) = 10.15$, $MS_e = 6,331$, $p < .005$, and $F_2(1,26)$

$= 6.31$, $MS_e = 15,343$, $p < .05$], but no main effect of version and no interaction of version \times appropriateness (all $F_s < 1$). Paired comparisons confirmed what is indicated by the pattern of the means, that there was an appropriateness effect for both instrument and pronoun versions of the sentences. The comparison of appropriate and inappropriate items was significantly different in the instrument version [$F_1(1,23) = 4.19$, $MS_e = 9,528$, $p = .05$, and $F_2(1,27) = 5.24$, $MS_e = 10,421$, $p < .05$], although the same comparison only reached marginal significance in the pronoun version [$F_1(1,23) = 3.52$, $MS_e = 7,168$, $p < .10$, and $F_2(1,27) = 3.20$, $MS_e = 13,288$, $p < .10$].⁷

The analyses showed that there was an appropriateness effect for both relevant instruments and pronoun antecedents. As in Experiment 1A, the question arises as to whether the appropriateness effect here involves a congruity effect that is the result of using general conceptual knowledge to integrate the target with its immediate context as a postlexical check, or whether it is the result of accessing a discourse model. If the former account were correct, however, we would expect to see a congruity effect in Experiment 2A with the isolated sentences, but we did not. Subjects need to access the information that goes beyond the immediate context, information that is only available from a mental model of the previous discourse.

What is most interesting about the instrument data is the evidence for greater activation of appropriate instruments in sentences where the instrument is merely implied, but where filling that case role is not necessary for understanding. In combination with the results of Experiments 2A and 2B, which failed to show evidence of an appropriateness effect for instruments in which the same sentences were presented in isolation, these results support the hypothesis that under certain conditions instrument inferences will be made. In particular, although these inferences will not be made for sentences presented out of context, they will be drawn when sentences are presented in a context that explicitly mentions the instrument. Because there was no noninference control condition in this experiment, however, there is no way to say whether the difference in reaction times across appropriateness conditions was due to facilitation for appropriate items or to inhibition for inappropriate items.

The main interest in including the pronoun condition in this experiment was to establish a baseline against which instrument activation could be assessed. Because pronouns require accessing their antecedents, if another null result had been found for instrument activation, the absence of activation for the pronoun condition would have called our materials or design into question. As it is, although there was priming in the pronoun condition, the conclusion that pronoun antecedent assignment begins immediately after processing the pronoun cannot be made definitively. This is because the results of the instrument conditions indicate that the instrument is already available at that point in the sentence. Therefore, the greater activation for appropriate antecedents (compared with in-

Table 5
Mean Reaction Times (in Milliseconds) and Percentages of Error for Experiments 3A (Lexical Decision) and 3B (Naming)

Type of Task	Type of Target			
	Appropriate	% Error	Inappropriate	% Error
Pronoun Version				
Lexical decision	987	1.6	1032	1.9
Naming	735	1.5	739	1.0
Instrument Version				
Lexical decision	982	2.2	1040	1.9
Naming	718	.7	738	1.0

appropriate antecedents) could just be the greater activation for appropriate instruments, especially given that there is no more priming in the pronoun conditions over and above that in the instrument conditions. This would be the case if it is too early at the end of the pronoun to see the effects of antecedent assignment. This seems unlikely, however, given the results of Experiment 1, which showed that antecedent assignment occurs rapidly.

A more likely possibility, which is consistent with the results of Experiment 1A, is that antecedent assignment occurs immediately but results in no additional priming. This interpretation is supported by the results of Experiment 1A, which indicated that antecedent assignment does not involve reactivating or providing additional activation to the appropriate antecedent. If the appropriate antecedent is already salient in the discourse representation, there is no reason to expect additional facilitation over that found in the instrument condition.

The results just discussed suggest that certain kinds of inferences, instrument inferences and antecedent assignment for pronouns, are made when the right context conditions are in place. In order to evaluate the form of representation accessed in both types of inference, it is necessary to compare performance in the lexical decision task with performance on the naming task. This was our purpose in the next experiment.

Experiment 3B

Method

Subjects. Twenty-four subjects from the University of Rochester participated in this experiment for course credit.

Materials and Procedure. These were the same as for Experiment 3A, except that nonwords were replaced by word targets unrelated to the sentences and subjects were asked to say the target word out loud as quickly as possible. Naming responses were measured from the onset of the visual target to the onset of the spoken response. Failures to activate the microphone were counted as errors.

Results

Mean reaction times and percentages of error are reported in Table 5. Errors constituted only 4.2% of responses, evenly distributed across conditions. Analyses of variance performed on the data showed no evidence of any appropriateness effect with the naming responses. There were no main effects of either appropriateness or version and no interaction of version \times appropriateness (all $F_s < 1$).

Although there was no statistically significant appropriateness effect in this experiment, numerically there is a hint of an appropriateness effect in the instrument conditions (718 vs. 738 msec). This was also true of naming Experiment 2B (604 vs. 630 msec). It is possible that a combined analysis of these experiments might show a statistically significant appropriateness effect where the individual analyses showed none. Accordingly, a combined analysis of Experiments 2B and 3B (instrument conditions only) was done. There was still no main effect of appropriateness, however [$F_1(1,38) = 2.77$, $MS_e = 10,028$, and $F_2(1,54) = 1.16$, $MS_e = 7,856$], and no in-

teraction between experiment and appropriateness [$F_1(1,38) < 1$, $MS_e = 205$, and $F_2(1,54) < 1$, $MS_e = 1,545$]. The only significant effect was that of experiment [$F_1(1,38) = 6.29$, $MS_e = 237,630$, $p < .02$, and $F_2(1,54) = 27.57$, $MS_e = 304,097$, $p < .0001$].

Discussion: Experiments 3A and 3B

Because there was an appropriateness effect only in the lexical decision responses and not in the naming responses, we conclude that only the discourse model and not the linguistic form representation is accessed in both instrument inferences and antecedent assignment for pronouns.

There is, however, an alternative interpretation of both the instrument and the pronoun data, which does not require the assumption that the appropriateness effect is the result of an inference. A more mundane possibility is that the appropriate target, an instrument, is simply more salient than the inappropriate target, a location. This may be because the latter is more likely to be perceived as the focus of the sentence, but it may be due simply to the perceived greater importance of instruments over locations. If instruments are more salient, the priming observed in Experiment 3A could be the result of a postaccess context check wherein the instrument is found to be more consistent with the discourse than is the location. The instrument target would be primed solely because of its discourse function in the context sentence, not because of an inference made in the second sentence. Two control experiments were performed to investigate this possibility. In the first, the second sentence was changed so that no anaphoric or other inferential reference was made to the instrument. If context-checking was the cause of the priming for the instrument in Experiment 3A, then the instrument targets should be primed even when no inferential reference is made to the instrument. In the second control experiment, only context sentences were presented. If instruments are more salient than locations, then instrument targets should elicit faster responses than the location targets after the context sentence alone.

Experiment 3C

Method

Subjects. Twenty-two University of Rochester students were paid \$6 for their participation.

Materials and Procedures. The same context sentences were used in this experiment as were used in Experiment 3A. The second sentence of each sentence pair was replaced by a sentence that made no inferential reference to the instrument but which was still consistent with the context sentence. For example, the first sentence pair (5) below was replaced by the next one (6). The slash marks indicate where targets were presented for lexical decision. Targets always followed the main verb and were never presented at the end of the sentence (see Appendix B).

5. There was a broom in the closet next to the kitchen.
John swept the floor every week/ on Saturday.
6. There was a broom in the closet next to the kitchen.
John is the most organized person/ I know.

The procedure for this experiment was identical to that for Experiment 3A.

Results

Mean reaction times and percentages of error are reported in Table 6. *t* tests indicate no significant difference between instrument and control targets [$t(21) = .72$, $SE = 12.31$, by subjects and $t(27) = .59$, $SE = 26.68$ by items; two-tailed tests].

Experiment 3D

Method

Subjects. Twenty-two University of Rochester students were paid \$6 for their participation.

Materials and Procedure. These were the same as in Experiment 3A, except that only the first (context) sentences of each sentence pair were presented. Visual targets were presented for lexical decision immediately following the last word of the context sentence.

Results

Mean reaction times and percentages of error are presented in Table 6. There was no significant difference between instrument and location targets [$t(21) = 1.07$, $SE = 18.04$, by subjects and $t(27) = .56$, $SE = 38.13$ by items; two-tailed tests].

Discussion: Experiments 3C and 3D

Postaccess context checking to facilitate the lexical decision task seems an unlikely explanation for the results of Experiment 3A in light of these two experiments. There is no evidence that instrument targets were more salient or consistent with the context. Instrument targets were not responded to faster than location targets when the second sentence did not contain an inferential verb but was still consistent with the context sentence, nor was there an advantage following the context sentence alone.

GENERAL DISCUSSION

The studies reported here support a number of conclusions: (1) Antecedent assignment occurs immediately after an anaphor is processed. This seems to be inconsistent with other results from a probe recognition study by Gernsbacher (1989), but her materials had pronouns with ambiguous referents. Our materials, in contrast, had only one plausible referent by the end of the pronoun or definite noun phrase. Also, probe recognition, by requiring the subject to decide consciously if the target had been encountered in the context, may have encouraged subjects

to develop strategies focused more on explicitly matching the target and context, obscuring the antecedent assignment process. (2) Antecedent assignment involves inhibition for the inappropriate potential antecedent rather than facilitation for the appropriate antecedent, suggesting that the mechanism for antecedent assignment involves backgrounding the inappropriate element rather than providing additional activation to the appropriate antecedent. (3) Inferences about the instrument accompanying an activity will be made if previous discourse has explicitly mentioned the instrument, but not if sentences with verbs implying the use of an instrument are presented out of context. This is similar to what McKoon and Ratcliff (1981) found, using a probe recognition task.

These three conclusions rest on the results of the lexical decision experiments. Although it can be argued that lexical decisions involve a certain amount of congruity or context matching, the control experiments (1B, 3C, 3D) rule this out as an explanation for our results. There is still the possibility, however, that a more sophisticated form of context checking was taking place. Rather than simply checking the target against the context to see if it "made sense" in some simple or superficial way, the subjects in Experiments 1A and 3A might have been using the representation of the *integrated* context and target sentence to make their context-target match. In this integrated representation—the discourse model—the instrument or antecedent from the first sentence and the verb or anaphor from the second sentence might be connected in such a way that, although one is not explicitly indicated to be the instrument or antecedent for the other, they are positioned to facilitate that inference under certain conditions. One such condition might be the presentation of a target word for lexical decision. Experiments 1B, 3C, and 3D did not rule out this possibility, but it is not clear just how different this kind of sophisticated context checking is from the inferential processes that were the focus of our experiments (or how it would be possible to tease out the differences experimentally). Whether the relationship between anaphor and the antecedent or the action and the instrument used to perform the action is explicit in the discourse model or just easily formed and facilitated by the representation, the fact remains that under conditions of context there is a special relationship between the action and the instrument and the anaphor and its antecedent that is not present when the action or anaphor is presented alone. Whether one chooses to call this special relationship inference or not would seem to be a matter of terminology more than substance.

The view that antecedent assignment and instrument inference require integration of sentences in a discourse is consistent with Doshier and Corbett's (1982) conclusion that some sentential inferences are not made automatically. It is also consistent with the conclusions of McKoon and Ratcliff (1981) and Singer (1980), who showed that inferences can become part of the memory representation for a sentence if the inference is necessary for preserving discourse coherence and if context is available to provide

Table 6
Mean Reaction Times (in Milliseconds) and
Percentages of Error for Experiments 3C and 3D

Experiment	Type of Target			
	Instrument	% Error	Location	% Error
3C	1020	.9	1029	.8
3D	880	.7	900	.7

the elements that will establish that coherence. The fact that inferences as disparate as filling instrument case roles and assigning antecedents to anaphors were made when the right conditions were in place suggests that it is not a particular type of inference but rather the necessity to integrate information that causes the implicit inference to be made. This also seems to be implied in recent work by Till, Mross, and Kintsch (1988). Using rapid serial visual presentation of sentence contexts and lexical decisions to targets as a measure of availability of discourse elements, they found evidence that words related to the topic of a discourse were facilitated relative to unrelated control words, but only at the end of the sentence, not before. It is at that point, they argued, that the coherence of the sentence in relation to the overall theme of the discourse can be established. It also raises the possibility that establishing discourse coherence is not only a necessary but also a sufficient condition for drawing implicit inferences. That is, it may not be a requirement for an instrument inference that there be an explicit mention of an instrument in the previous discourse but only that something in the context constrain the situation such that an appropriate discourse model would include the inference. These tentative conclusions about inferences and discourse coherence, of course, require further research into other types of sentential inference and types of contexts.

The last and perhaps most interesting conclusion from these studies concerns the representational issue. The results support the thesis that different types of sentential inference involve accessing elements in a discourse model. In particular, this is true for two types of anaphoric inference—antecedent assignment for definite noun phrases and for pronouns—and instrument inferences. This was shown by finding differences in the pattern of responses for the lexical decision and the naming tasks, a useful diagnostic test for the levels of representation involved in sentence comprehension. For all three types of inferences, lexical decisions, which have been shown to be sensitive to both lexical and discourse level representations, were sensitive to the differences between appropriate and inappropriate antecedents and instruments. Naming responses, which have been shown to be primarily sensitive to lexical representations, did not show antecedent assignment or instrument inference effects. We conclude, therefore, that making instrument inferences and assigning antecedents to anaphors involves accessing entities in a discourse model.

Even though there was no evidence for changing activation in lexical forms, the possibility remains that lexical forms are used in some limited way not detected by the naming task, particularly in antecedent assignment. Nonetheless, we can rule out the possibility that this involves either activation of the lexical form of the appropriate antecedent or inhibition of the lexical form of the inappropriate antecedent.

There are background assumptions in this work, however, that are critical to drawing these conclusions about levels of representation. These assumptions need

to be addressed again in light of recent evidence that seems to call these assumptions into question. The validity of the lexical-decision-naming-task difference as a diagnostic test for levels of representation is based on work that shows that the naming task is not as sensitive to discourse level representations as it is to lexical representations, although the lexical decision task is sensitive to both. This difference in sensitivity to different levels, it should be emphasized, is one of degree. Naming is not a pure measure of access only to lexical representations. As mentioned in the introduction, there is evidence for an effect in naming of expectancy based on the proportion of related prime-target pairs in a list (Keefe & Neely, in press; Norris, 1987), a type of strategic, message-level effect normally associated with lexical decisions.

Other recent work suggests that naming also benefits from higher order contextual integration. Potts, Keenan, and Golding (1988) have found that naming will show congruity effects when a backward inference is necessary for the coherence of a discourse (an example of a backward inference of this kind would be to infer that a vase broke after hearing that it was thrown against a wall and later had to be replaced). They found that subjects were faster to name a word that is related to the backward inference than one that is not. Also, in a study of anaphoric inference, O'Brien, Duffy, and Myers (1986) found priming effects in naming for reinstated antecedents of anaphoric noun phrases. The time to respond to a previously mentioned antecedent that had been taken out of focus was shorter after subjects had read an anaphoric phrase that reinstated the antecedent than it was after they had read an anaphoric phrase that reinstated a different antecedent. O'Brien et al. determined that the effect was not due merely to semantic priming from the noun to the antecedent (example: *the stalled vehicle-bus*) by including a control condition that used the head noun of the phrase without referring back to the antecedent (example: *the police vehicle-bus*). Under these circumstances, naming in the reinstatement condition was still faster than naming in the semantic priming control condition, indicating that naming was indeed sensitive to anaphoric inference in this experiment. Finally, Osterhout and Swinney (1989) recently reported finding instrument inference effects with naming. As in the present set of experiments, they found no evidence for instrument inferences, using either lexical decision or naming, when sentences that implied the instrument were presented out of context. When sentences were preceded by a rich context, however, Osterhout and Swinney found facilitation for implied instruments with both the lexical decision and the naming tasks. The contexts did not specifically mention the instrument but instantiated a schema in which instruments typically played a prominent role.

Two points need to be made here. First, it is possible that there is a floor effect in the naming data. The reaction times and error rates for the naming data are very low in comparison to the lexical decision data. It suggests that subjects found the naming task so easy that there was

too narrow a margin of difference across conditions to reach statistical significance. Without doing the experiments again under different instructions, we cannot say for sure, however, whether or not this was the case. Floor effects are unlikely, however, in light of finding associative priming effects with naming in Experiment 1C.

Second, the conditions under which the studies mentioned above showed contextual integration effects for naming were different from those in the current experiments. What is common to all three studies mentioned above but not to the present experiments is that the relevant entity either had not been mentioned at all in the discourse (Keenan et al., 1988; Osterhout & Swinney, 1989) or had been mentioned earlier but taken out of focus (O'Brien et al., 1986). In all cases, the item relevant to the inference was not currently available in the focus of the constructed representation of the sentence. In contrast, in our experiments, the relevant discourse entities were introduced recently and were still part of the ongoing discourse focus. This difference may explain why the lexical form is not accessed in our experiments but was in the others. In Keenan et al. (1988), the item needed to be created to preserve coherence; in O'Brien et al. (1986), it needed to be re-created or brought back into the focus of the discourse model. In Osterhout and Swinney (1989), the item needed to be brought into focus as a part of instantiating a schema. When antecedents are not in focus, it may be easier to access the long-term memory representation via a lexical representation. This may provide the quickest path to an appropriate element (or help create a new element) in the constructed representation of the discourse. The lexical representation would then be briefly primed for pronunciation.

We have shown that task comparisons can be used to tap into different levels of representation for different sentence comprehension processes. This has the potential to provide valuable information about the structure of information flow in the language processor, especially in cases where the relative contributions of surface form representations and discourse models are in question (e.g., which aspects of a word's meaning when instantiated in a sentence are available at a lexical level of processing and which in a mental model of the discourse?). Although it is possible that the bases for difference in context sensitivity that support this diagnostic test will turn out to be more complex than was originally thought, our experiments present encouraging evidence that comparisons among lexical decisions and naming can be useful in answering questions about the nature of representations involved in discourse processing.

REFERENCES

- ANDERSON, A., GARROD, S. C., & SANFORD, A. J. (1983). The accessibility of pronominal antecedents as a function of episode shifts in narrative text. *Quarterly Journal of Experimental Psychology*, *35A*, 427-440.
- BALOTA, D. A., & CHUMBLEY, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception & Performance*, *10*, 340-357.
- BARTLETT, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge, U.K.: Cambridge University Press.
- CHANG, F. R. (1980). Active memory processes in visual sentence comprehension: Clause effects and pronominal reference. *Memory & Cognition*, *8*, 58-64.
- CHUMBLEY, J. I., & BALOTA, D. A. (1984). A word's meaning affects the decision in lexical decision. *Memory & Cognition*, *12*, 590-606.
- CLOITRE, M., & BEVER, T. G. (1989). Linguistic anaphors, levels of representation, and discourse. *Language & Cognitive Processes*, *3*, 293-322.
- CORBETT, A. T., & CHANG, F. R. (1983). Pronoun disambiguation: Accessing potential antecedents. *Memory & Cognition*, *11*, 283-294.
- CORBETT, A. T., & DOSHER, B. A. (1978). Instrument inferences in sentence encoding. *Journal of Verbal Learning & Verbal Behavior*, *17*, 479-491.
- 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 & Verbal Behavior*, *22*, 121-132.
- DOSHER, B. A., & CORBETT, A. T. (1982). Instrument inferences and verb schemata. *Memory & Cognition*, *10*, 531-589.
- EHRlich, K. (1980). Comprehension of pronouns. *Quarterly Journal of Experimental Psychology*, *32*, 247-255.
- EHRlich, K., & JOHNSON-LAIRD, P. N. (1982). Spatial descriptions and referential continuity. *Journal of Verbal Learning & Verbal Behavior*, *21*, 296-306.
- FILLMORE, C. J. (1968). The case for case. In E. Bach & R. T. Harms (Eds.), *Universals in linguistic theory* (pp. 1-90). New York: Holt, Rinehart & Winston.
- FILLMORE, C. J. (1971). Types of lexical information. In D. D. Steinberg & L. A. Jakobovits (Eds.), *Semantics: An interdisciplinary reader in philosophy, linguistics and psychology* (pp. 370-392). Cambridge, U.K.: Cambridge University Press.
- FISCHLER, I., & BLOOM, P. A. (1979). Automatic and attentional processes in the effects of sentence contexts on word recognition. *Journal of Verbal Learning & Verbal Behavior*, *18*, 1-20.
- FORSTER, K. I. (1976). Accessing the mental lexicon. In K. J. Wales & E. Walker (Eds.), *New approaches to language mechanisms* (pp. 257-287). Amsterdam: North-Holland.
- FORSTER, K. I. (1979). Levels of processing and the structure of the language processor. In W. E. Cooper & E. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 27-85). Hillsdale, NJ: Erlbaum.
- GARNHAM, A. (1986). Understanding anaphora. In A. W. Ellis (Ed.), *Progress in the psychology of language* (Vol. 3, pp. 253-300). London: Erlbaum.
- GERNSBACHER, M. A. (1986). The comprehension of conceptual anaphora in discourse. *Proceedings of the Eighth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.
- GERNSBACHER, M. A. (1989). Mechanisms that improve referential access. *Cognition*, *32*, 99-156.
- HANKAMER, J., & SAG, I. (1976). Deep and surface anaphors. *Linguistic Inquiry*, *7*, 391-426.
- HEIM, I. (1983). File change semantics and the familiarity theory of definiteness. In R. Bauerle, C. Schwarze, & A. von Stechow (Eds.), *Meaning, use, and interpretation of language* (pp. 164-190). New York: Walter de Gruyter.
- JOHNSON, M. K., BRANSFORD, J. D., & SOLOMON, S. K. (1973). Memory for tacit implications of sentences. *Journal of Experimental Psychology*, *98*, 203-205.
- JOHNSON-LAIRD, P. N., & STEVENSON, R. (1970). Memory for syntax. *Nature*, *227*, 412.
- KAMP, H. (1981). A theory of truth and semantic representation. In J. A. G. Groenendijk, T. M. U. Janssen, & M. B. J. Stokhof (Eds.), *Formal methods in the study of language* (pp. 277-322). Amsterdam: Mathematisch Centrum.
- KEEFE, D. E., & NEELY, J. H. (in press). Semantic priming in the pronunciation task: The role of prospective prime-generated expect-

- tancies. *Journal of Experimental Psychology: Learning, Memory, & Cognition*.
- KUČERA, H., & FRANCIS, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- KUNO, S. (1987). *Functional syntax: Anaphora, discourse, and empathy*. Chicago: University of Chicago Press.
- LASNIK, H. (1976). Remarks on coreference. *Linguistic Analysis*, 2, 1-22.
- LESGOLD, A., ROTH, S., & CURTIS, M. (1979). Foregrounding effects in discourse comprehension. *Journal of Verbal Learning & Verbal Behavior*, 18, 291-308.
- McKOOK, G., & RATCLIFF, R. (1980). The comprehension processes and memory structures involved in anaphoric reference. *Journal of Verbal Learning & Verbal Behavior*, 19, 668-682.
- McKOOK, G., & RATCLIFF, R. (1981). The comprehension process and memory structures involved in instrumental inference. *Journal of Verbal Learning & Verbal Behavior*, 20, 671-682.
- MEYER, D. E., & SCHVANEVELDT, R. W. (1976). Meaning, memory structure and mental processes. *Science*, 192, 27-33.
- MURPHY, G. L. (1985). Processes of understanding anaphora. *Journal of Memory & Language*, 24, 290-303.
- NICOL, J., & SWINNEY, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research*, 18, 5-20.
- NORRIS, D. (1987). Strategic control of sentence context effects in a naming task. *Quarterly Journal of Experimental Psychology*, 39A, 253-275.
- O'BRIEN, E. J., DUFFY, S. A., & MYERS, J. L. (1986). Anaphoric inference during reading. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 12, 346-352.
- OSTERHOUT, L. E., & SWINNEY, D. A. (1989, April). *The effects of context on elaborative inferencing during discourse processing*. Paper presented at the Annual Meeting of the Eastern Psychological Association, Boston.
- PARIS, S. C., & LINDAUER, B. K. (1976). The role of inference in children's comprehension and memory for sentences. *Cognitive Psychology*, 8, 217-227.
- POTTS, G. R., KEENAN, J. M., & GOLDING, J. M. (1988). Assessing the occurrence of elaborative inferences: Lexical decision versus naming. *Journal of Memory & Language*, 27, 399-415.
- REINHART, T. (1983). *Anaphora and semantic interpretation*. London: Croom Helm.
- RUMELHART, D. E. (1975). Notes on a schema for stories. In D. G. Bobrow & A. M. Collins (Eds.), *Representation and understanding: Studies in cognitive science* (pp. 211-236). New York: Academic Press.
- SACHS, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics*, 2, 437-442.
- SAG, I., & HANKAMER, J. (1984). Toward a theory of anaphoric processing. *Linguistics & Philosophy*, 7, 325-345.
- SCHANK, R. C. (1975). The structure of episodes in memory. In D. G. Bobrow & A. M. Collins (Eds.), *Representation and understanding: Studies in cognitive science* (pp. 237-272). New York: Academic Press.
- SCHANK, R. C., & ABELSON, R. P. (1977). *Scripts, plans, goals, and understanding*. New York: Academic Press.
- SEIDENBERG, M. S., WATERS, G. S., SANDERS, M., & LANGER, P. (1984). Pre- and postlexical loci of contextual effects on word recognition. *Memory & Cognition*, 12, 315-328.
- SINGER, M. (1979). Processes of inference during sentence encoding. *Memory & Cognition*, 7, 192-200.
- SINGER, M. (1980). The role of case-filling inferences in the coherence of brief passages. *Discourse Processes*, 3, 185-201.
- TANENHAUS, M. K., & CARLSON, G. (in press). Comprehending verb phrase anaphors. *Language & Cognitive Processes*.
- TILL, R. E., MROSS, E. F., & KINTSCH, W. (1988). Time course of priming for associate and inference words in a discourse context. *Memory & Cognition*, 16, 283-298.
- WALKER, C. H., & YEKOVICH, F. R. (1987). Activation and use of script-based antecedents in anaphoric reference. *Journal of Memory & Language*, 26, 673-691.
- WANNER, E. (1974). *On remembering, forgetting and understanding sentences: A study of the deep structure hypothesis*. The Hague: Mouton.
- WEBBER, B. L. (1979). *A formal approach to discourse anaphora*. New York: Garland.
- WEBBER, B. L. (1983). So what can we talk about now? In M. Brady & R. C. Berwick (Eds.), *Computational models of discourse* (pp. 331-371). Cambridge, MA: The MIT Press.
- WEST, R. F., & STANOVICH, K. E. (1982). Source of inhibition in experiments on the effect of sentence context on word recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 8, 385-399.
- WEST, R. F., & STANOVICH, K. E. (1986). Robust effects of syntactic structure on visual word processing. *Memory & Cognition*, 14, 104-112.

NOTES

1. "Discourse model" is used here in the same sense as that in which it is used by philosophers and computational linguists such as Heim (1983), Kamp (1981), and Webber (1979, 1983), when they describe the representations of properties and relations formed in response to the linguistic input. We realize that there is another sense of "discourse model" that evokes yet another level of representation and refers to a mental model of the situation or state of affairs that is being discussed. Webber (1983) gives a nice example that shows how texts identical at the level of the situational mental model may nonetheless be different at the level of the discourse model because the entities that can be extracted from the linguistic form of the one cannot be extracted from the linguistic form of the other. The example involves the following sentences:

1. John traveled around France twice.
??They were both wonderful.
2. John took two trips around France.
They were both wonderful.

Only the second version gives "trips" a representation in the discourse model such that it is available for subsequent reference by the anaphor "they." Although the concept of trips may be part of the mental model of the situation constructed by the listener on hearing Sentence 1, it cannot be assumed to be shared by both speaker and hearer.

Although it is true that in the testing of postlexical processing that we discuss in this paper, we were not able to make the fine distinction between a discourse model and a situational model, the processes we are studying, particularly anaphoric reference, involve entities most likely to be represented in a discourse model.

2. We used only one potential target (e.g., just "steak," rather than both "steak" and "hamburger"), but it was balanced across three different versions of Sentence 2—one in which it was appropriate, one in which it was inappropriate, and one in which it was unrelated. We did not believe that anything would be gained by running the other target through the same conditions, since our argument just hinges on differences in priming across the three versions.

3. The interpretation of inhibition for inappropriate targets is also supported if separate comparisons are performed at each delay. The comparison of inappropriate and control conditions is significant at the 0-msec delay in the subject analysis [$F_1(1,15) = 7.16$, $MS_e = 10,053$, $p < .05$] and approaches significance in the item analysis [$F_2(1,23) = 3.89$, $MS_e = 780$, $p = .06$]. At 600 msec, the comparison is significant in both the subject and the item analyses [$F_1(1,15) = 7.22$, $MS_e = 9,401$, $p < .05$, and $F_2(1,23) = 5.89$, $MS_e = 17,319$, $p < .05$]. The subject analysis reaches significance at 900 msec [$F_1(1,15) = 10.45$, $MS_e = 3,230$, $p < .01$], but not the item analysis [$F_2(1,23) = 1.57$, $MS_e = 31,938$]. Although the appropriate and inappropriate condition comparisons did not reach significance in the 0-msec delay [$F_1(1,15) = 1.6$, $MS_e = 16,866$, and $F_2(1,23) = 1.15$, $MS_e = 35,754$], the comparisons were significant at the 600-msec delay [$F_1(1,15) = 16.44$, $MS_e = 6,828$, $p < .005$, and $F_2(1,23) = 8.88$, $MS_e = 18,974$, $p < .01$] and at the 900-msec delay in the subject analysis [$F_1(1,15) = 5.56$, $MS_e = 5,944$, $p < .05$, and $F_2(1,23) = 1.42$, $MS_e = 34,927$]. Comparisons of the

appropriate and control conditions were not significant in either subject or item analyses at any delay.

4. Pairwise comparisons performed at the two delays show a similar picture. The difference between appropriate and control conditions was significant in the subject analysis but not the item analysis at both the 0-msec delay [$F_1(1, 15) = 10.55$, $MS_e = 792$, $p < .01$, and $F_2(1,23) = 1.5$, $MS_e = 8,404$, n.s.] and the 600-msec delay [$F_1(1,15) = 11.24$, $MS_e = 606$, $p < .01$, and $F_2(1,23) = 2.85$, $MS_e = 3,567$, n.s.]. The difference between the inappropriate and control conditions was not significant at the 0-msec delay (F_1 and $F_2 < 1$), although the amount of priming (14 msec) is in the right direction. The difference between inappropriate and control conditions was marginally significant, however, in the 600-msec delay condition [$F_1(1,15) = 4.14$, $MS_e = 1,153$, $p = .06$, and $F_2(1,23) = 2.3$, $MS_e = 3,086$, n.s.]. Appropriate and inappropriate conditions did not differ significantly at either the 0-msec delay [$F_1(1,15) = 2.7$, $MS_e = 983$, n.s., and $F_2(1,23) < 1$] or the 600-msec delay (F_1 and $F_2 < 1$).

5. Two concerns came to our attention after the experiments were done. The first was that the materials were not uniformly of the type that would require a discourse model for their comprehension; that is, some of the antecedents could be assigned on the basis of sentence-specific information and others on the basis of general world knowledge. In the case of sentence-specific information (i.e., where the discriminating adjective in the context sentence is repeated in the second sentence), subjects could conceivably access just the lexical form rather than the discourse model to make the antecedent assignment. Upon closer examination, we found only six items that permitted the use of sentence-specific information (Nos. 4, 7, 8, 10, 20, 24 in Appendix A). We ran ANOVAs with information type as a factor on the lexical decision and naming data (collapsing across target delays). We found no effect of type of information needed to make the inference [$F(1,22) = 1.57$, n.s., for naming and $F(1,22) = .15$, n.s., for the lexical decision data]. There was also no interaction of type \times appropriateness in either naming [$F(2,44) = .12$, n.s.] or lexical decision [$F(2,44) = .61$, n.s.]. Although there were only six items in the sentence-specific condition for this analysis, we felt that the lack of even a marginally significant pattern of differences indicated that there was no difference in level of representation access in these two sentence types.

The other concern was an alternative explanation for our results that does not involve invoking different levels of representation. This possibility, suggested by a reviewer, is that both lexical decision and naming are sensitive to postlexical checks, but that naming can be completed before anaphor resolution or inferencing are complete. This hypothesis would seem to be ruled out, though, by the results of Experiment 1C, where we do not find an appropriateness effect (indicating that antecedent assignment has taken place) with naming at a 600-msec delay. This is well after anaphora resolution should have been completed, as is evidenced by the fact that we find inhibition for the inappropriate antecedent at a 0-msec delay with lexical decision (Experiment 1A), using the same materials.

6. One particular concern in many probe recognition studies is the nature of the controls. In a typical study, two names are introduced in a sentence and then either of the names is repeated (the noun condition) or a pronoun is used. Probe recognition times are typically faster to both names after the pronoun condition as opposed to the noun condition. This result might simply be a confusion effect—that is, the repetition of a name would make the nonrepeated name harder to recognize, particularly in the context of an experiment that includes many names. The lack of an effect with different gender names might occur because names for the same sex are more confusable than names for different sexes.

7. The scores for the 14 verb-instrument pairs taken from Doshier and Corbett (1982) were compared with those for the 14 verb-instrument pairs specific to this experiment, to determine if there were any differences. Both sets of stimuli showed an advantage for appropriate over inappropriate items in both pronoun and instrument versions of the experiment. For Doshier and Corbett's stimuli in the instrument version, appropriate items elicited responses that were 62 msec faster than those for inappropriate items; for the newly generated pairs, appropriate items elicited responses that were 66 msec faster than those for inappropriate items. In comparison, the difference between appropriate and inappropriate items in the no-context Experiment 2A was -5 msec for Doshier and Corbett's instrument targets and -7 msec for the newly generated instrument targets. In the pronoun version of this experiment, Doshier and Corbett's items showed a 97-msec advantage for appropriate antecedents and the new items a 35-msec advantage for appropriate antecedents.

APPENDIX A

Stimuli for Experiments 1A-1C	Target
1. Walking through the woods Bob first spotted a skunk and then he spotted a bear. a. He was frightened by the smaller animal. b. He was frightened by the larger animal. c. He enjoyed the lovely scenery around him.	Skunk
2. Jamie wanted to be either a dentist or an artist. a. She chose the more lucrative profession. b. She chose the less lucrative profession. c. She was in her freshman year in college.	Dentist
3. The young boy was torn between learning to play the violin and the tuba. a. He finally chose the smaller instrument. b. He finally chose the larger instrument. c. He loved all kinds of music.	Violin
4. Charlie Brown wasn't sure which would make a better Christmas tree, the large fir or the small pine. a. He finally bought the smaller tree. b. He finally bought the larger tree. c. The lot had a wide selection.	Pine

APPENDIX A (Continued)

Stimuli for Experiments 1A-1C	Target
<p>5. Brick and wood are common materials from which to make a house.</p> <p>a. Builders tell us though that it is easier to work with the softer material.</p> <p>b. Builders tell us though that it is easier to work with the harder material.</p> <p>c. The cost of building a house varies greatly among different builders.</p>	Wood
<p>6. Mary couldn't decide whether to play tennis or to play basketball.</p> <p>a. She chose to go out for the individual sport.</p> <p>b. She chose to go out for the team sport.</p> <p>c. The sports program at her school was an excellent one.</p>	Tennis
<p>7. The gourmet was asked whether he preferred fresh water trout or salt water tuna.</p> <p>a. He said that he preferred the fresh water fish.</p> <p>b. He said that he preferred the salt water fish.</p> <p>c. The cook was an expert in preparing all kinds of good food.</p>	Trout
<p>8. Lying on the desk were a blue novel and a red dictionary.</p> <p>a. Fred walked over and picked up the red book.</p> <p>b. Fred walked over and picked up the blue book.</p> <p>c. The desk was dusty and looked as though it had not been used in quite a while.</p>	Dictionary
<p>9. Sarah could not decide whether to buy steak or hamburger.</p> <p>a. She finally chose the more expensive meat.</p> <p>b. She finally chose the less expensive meat.</p> <p>c. The store had a wide selection of foods.</p>	Steak
<p>10. Joyce was debating whether to take a serious biography or a light mystery on the train trip.</p> <p>a. She finally decided to take the more serious book.</p> <p>b. She finally decided to take the lighter book.</p> <p>c. She hoped to have plenty of time to read.</p>	Biography
<p>11. Couches and chairs are similar to each other in many ways.</p> <p>a. For pure comfort many people choose the larger piece of furniture.</p> <p>b. For pure comfort many people choose the smaller piece of furniture.</p> <p>c. Furniture may be expensive but it need not be replaced for years.</p>	Couch
<p>12. Mary and Betty had nothing to do so they thought about playing either poker or monopoly.</p> <p>a. After a long time they both decided to play the board game.</p>	

APPENDIX A (Continued)

Stimuli for Experiments 1A-1C	Target
<ul style="list-style-type: none"> b. After a long time they both decided to play the card game. c. After a long discussion they finally decided to go to a movie at the mall instead. 	Monopoly
<p>13. Bob was trying to decide whether to take a bus or a plane to Toronto for the weekend.</p> <ul style="list-style-type: none"> a. He finally decided on the faster form of transportation. b. He finally decided on the slower form of transportation. c. He had never been out of the United States before. 	
<p>14. Dogs make good pets and cats do too.</p> <ul style="list-style-type: none"> a. In South America most people can't tolerate the smaller animal. b. In South America most people can't tolerate the larger animal. c. A pet can be a good companion to a person who lives alone. 	Plane
<p>15. Bill could not decide whether to go out for hockey or golf.</p> <ul style="list-style-type: none"> a. He finally chose the contact sport. b. He finally chose the noncontact sport. c. He was very good at almost all sports. 	Cat
<p>16. The convicted criminal was given a choice between a fine and going to prison.</p> <ul style="list-style-type: none"> a. He chose the unconfining form of punishment. b. He chose the confining form of punishment. c. He had committed a fairly minor crime. 	Hockey
<p>17. John was debating whether to order fruit or a pie.</p> <ul style="list-style-type: none"> a. He decided to order the less fattening dessert. b. He decided to order the more fattening dessert. c. He had enjoyed the good meal at the restaurant. 	Fine
<p>18. At the zoo the children's favorite animals are the lions and the camels.</p> <ul style="list-style-type: none"> a. Sandra's favorite is the jungle animal. b. Sandra's favorite is the desert animal. c. We go to the zoo about once a year. 	Fruit
<p>19. Our family has taken vacations in Maine and in Texas for the past ten summers.</p> <ul style="list-style-type: none"> a. My mother enjoys only half our trips because she can't stand the climate of the Northern state. b. My mother enjoys only half our trips because she can't stand the climate of the Southern state. c. We always enjoy those trips because they allow us to spend time together as a family. 	Lion
	Maine

APPENDIX A (Continued)

Stimuli for Experiments 1A-1C	Target
20. Red grapes and bananas are both delicious fruits. a. In our house though everyone prefers the red fruit. b. In our house though everyone prefers the yellow fruit. c. Fruit is most abundant during the summer months.	Grapes
21. Both the refrigerator and the mixer needed to be repaired. a. Sam decided to fix the smaller appliance first. b. Sam decided to fix the larger appliance first. c. Sam was afraid that the repairs would be quite expensive.	Mixer
22. Bill was trying to decide whether to buy a motorcycle or an automobile. a. He finally chose the smaller vehicle. b. He finally chose the larger vehicle. c. His old car was finally beyond repair.	Motorcycle
23. Kathy was trying to decide whether to wear a dress or jeans to school. a. She finally chose the less casual attire. b. She finally chose the more casual attire. c. She had to hurry or else she would be late.	Dress
24. Brian wanted to give his fiancee yellow tulips and red roses. a. He could only find the red flowers. b. He could only find the yellow flowers. c. He thought she would be pleased with the gift.	Roses

APPENDIX B

Stimuli for Experiments 2A-3D

The stimuli for Experiments 2A and 2B were the ‘‘b’’ sentences in each set. In those experiments, all proper names were replaced by the pronouns *he* or *she*. The words underlined in the ‘‘a’’ sentences were the targets presented at the point marked by a slash in the ‘‘b’’ sentences.

The ‘‘a’’ sentences were the context sentences added in Experiments 3A and 3B. In the ‘‘c’’ sentences, prepositional phrases containing pronouns were substituted for phrases in ‘‘b’’ sentences to form the pronoun versions for Experiments 3A and 3B.

Sentences marked ‘‘d’’ were used in Experiment 3C and followed the ‘‘a’’ sentences. Slashes mark where targets were presented for lexical decision.

- 1a. There was a *broom* in the *closet* next to the kitchen.
 - b. John swept the floor every week / on Saturday.
 - c. John swept the floor with it / on Saturday.
 - d. John is the most organized person / I know.
- 2a. From her *purse* Mary got out the *key* to her apartment.
 - b. She unlocked the door cautiously / and stepped inside.
 - c. She unlocked the door with it / and stepped inside.
 - d. Suddenly, she realized her credit cards were missing / and began to worry.
- 3a. John took his *car* out of the *garage* when the snow melted.
 - b. He drove to work in a hurry / this morning.
 - c. He drove to work in it / this morning.
 - d. He wanted to sort through the garden tools / before spring.

- 4a. In the *woodshed* John found the *axe* that belonged to his father.
b. He decided to chop wood for a fire / that evening.
c. He decided to chop wood with it / that evening.
d. He decided to look for some more old mementos / while there was still light.
- 5a. John picked up a *hammer* in the *attic* where he was working.
b. He pounded a nail carelessly / and hit his thumb.
c. He pounded a nail with it / and hit his thumb.
d. He stopped to enjoy the solitude / before going back to work.
- 6a. Mary borrowed a *ladder* from the *backyard* of her friend.
b. She climbed up the tree quickly / to get her cat down.
c. She climbed up the tree with it / to get her cat down.
d. They stopped to chat about raising vegetables / before Mary left.
- 7a. John bought a *brush* from the *store* last week.
b. He painted a picture of his teacher / in art class.
c. He painted a picture with it / in art class.
d. He also inquired about a part-time job / for the summer.
- 8a. The medicine *cabinet* was where Mary found her *scissors* after a long search.
b. She trimmed her hair a little / before she got dressed.
c. She trimmed her hair with them / before she got dressed.
d. She saw that she was running out of contact lens cleaner / and made a note to buy more.
- 9a. Sally pulled some *string* from the top *drawer* of her dresser.
b. She tied a package securely / and mailed it.
c. She tied a package with it / and mailed it.
d. She noticed that she was down to her last pair of socks / and decided to do some laundry.
- 10a. On the *counter* there was a *spoon* which had just been cleaned.
b. Janet stirred her martini gently / before drinking.
c. Janet stirred her martini with it / before drinking.
d. Janet was reminded that she needed to leave cash / for the cleaning lady.
- 11a. Alice took a *needle* from a *box* on the shelf.
b. She sewed a button on her blouse / and then called a friend.
c. She sewed a button with it / and then called a friend.
d. She noticed a lot of dust / and got a rag to wipe it off.
- 12a. On the *dock* there was a large *crane* made of steel.
b. Harry lifted some freight from a truck / onto a ship.
c. Harry lifted some freight with it / onto a ship.
d. Harry noticed a freighter in the distance / and stopped to watch it enter the harbor.
- 13a. There was a *shovel* on the *patio* near the garden.
b. Paula dug a hole near the tomatoes / to plant seeds.
c. Paula dug a hole with it / to plant seeds.
d. The snow had melted / and it was time for spring planting.
- 14a. On the *table* there was a *pitcher* which was empty.
b. Sam poured the cocktails in preparation / for his party.
c. Sam poured the cocktails with it / for his party.
d. Mary decided she had to redecorate the living room / before the next party.
- 15a. The terrorist planted a *bomb* in the *room* next to the furnace.
b. He blew up the building in a rage / when his demands were refused.
c. He blew up the building with it / when his demands were refused.
d. Then he locked up / and hid outside in the bushes.
- 16a. Alice took the *towel* off the *hook* on the refrigerator.
b. She dried the dishes thoroughly / before going to sleep.
c. She dried the dishes with it / before going to sleep.
d. Then a mouse scampered across the floor / and frightened her.
- 17a. In the *kitchen* there was a microwave *oven* that belonged to Sally's roommate.
b. Sally baked dinner for her boyfriend / for the first time yesterday.
c. Sally baked dinner with it / for the first time yesterday.
d. The roommate was traveling in Europe this summer / and said Sally could use it.

- 18a. Out of the *case* John took his *binoculars* and adjusted the focus.
b. He scanned the horizon slowly / in search of geese.
c. He scanned the horizon with them / in search of geese.
d. Then he sat back in his seat / and waited for a plane to land.
- 19a. Gary picked up a *club* from the *corner* of the room.
b. He beat the man mercilessly / several times.
c. He beat the man with it / several times.
d. He had decided to do some housecleaning / before his wife returned from her trip.
- 20a. Paul keeps a *net* in his *backpack* because he is an insect collector.
b. He captured a butterfly in the woods / just last Tuesday.
c. He captured a butterfly with it / just last Tuesday.
d. He also carries a guide / for identifying rare species.
- 21a. From the *pantry* Mary got a *tray* given to her at her wedding.
b. She served breakfast to her sick husband / and then went to work.
c. She served breakfast with it / and then went to work.
d. She tripped over the kid's boots / and decided to put them somewhere else.
- 22a. Chester got a *book* down from a *shelf* in the living room.
b. He studied economics diligently / in preparation for his test.
c. He studied economics with it / in preparation for his test.
d. He heard the phone ring / and went into the next room to answer it.
- 23a. At the costume *store* Cindy bought a *mask* that was on sale.
b. She disguised her identity cleverly / in order to spy on someone.
c. She disguised her identity with it / in order to spy on someone.
d. She also asked the clerk for directions / to the museum.
- 24a. Dan drew *money* out of the *bank* using his charge card.
b. He donated a large amount generously / to a local charity.
c. He donated a large amount of it / to a local charity.
d. He really liked the services / that modern technology made possible.
- 25a. From the *ground* Janice picked up her *racket* with great enthusiasm.
b. She swung at the ball from center court / forcefully.
c. She swung at the ball with it / forcefully.
d. Even though she had fallen, / she had still won the point.
- 26a. Wendy set the *guitar* on her *lap* and opened the songbook.
b. She strummed a melody softly / for her boyfriend.
c. She strummed a melody with it / for her boyfriend.
d. She leafed through the pages / looking for a particular selection.
- 27a. With great *relief* the soldier took off his *boots* and relaxed.
b. He had marched miles on bad road / that day.
c. He had marched miles in them / that day.
d. He was finally a safe distance / from the enemy shelling.
- 28a. Gwen poured some *water* into the *sink* in the bathroom.
b. She washed her face vigorously / and then went to bed.
c. She washed her face with it / and then went to bed.
d. She noticed a stain / and went to get some cleanser.