REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	12/15/93	Final 1 Mar 90 - 31 Aug 93

4. TITLE AND SUBTITLE

Reading: Interaction with Memory

5. FUNDING NUMBERS

AFOSR-90-0246 (G)

6. AUTHOR(S)

Gail McKoon

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Northwestern University
Department of Psychology
2029 Sheridan Rd
Evanston, IL 60208-2710

8. PERFORMING ORGANIZATION REPORT NUMBER

0650-300-W401

AFOSR-TR- 94 0097

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

AFOSR/NL
Building 410
Bolling AFB DC 20332-6448

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Available to general public. Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The topic of the supported research was reading, and the ways information in memory can contribute to the inference processes that occur during reading. One source of information for inference processes is short-term memory for parts of a text that have already been read. Experiments investigated how this information is made available to allow, for example, inferences that decide the correct referent of a pronoun, or inferences that relate via causality two events described by the text. Experiments also examined the local representation constructed for a text, testing our proposal that locally available information is structured by the linguistic, semantic, and pragmatic means by which the information is expressed. A second line of research examined interactions between inference processes and well-known information from long-term memory, examining knowledge of the semantic structures of verbs, knowledge of what concepts are frequently associated with each other, and knowledge about how lexical items are used in various contexts.

14. SUBJECT TERMS

Redding; Memory; Language; Comprehension

15. NUMBER OF PAGES

2

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	

NSN 7540-01-280-5500

Standard Form 298 (Rev 2-89)
Prescribed by ANSI Std Z39-18
298-102

AFOSR 93-0240

Final Report

PI: Gail McKoon

One major project was to combine in a theoretical paper the general view of text processing that I have been advocating with a large set of accumulated empirical results. The view, a "minimalist" view of text processing, is controversial; the claim that readers perform only a limited amount of inferencing during reading is not widely accepted. In research on reading and language comprehension, it had long been believed that readers/listeners understood texts and discourse to the extent of constructing a complete mental model of the linguistically described situation. In 1992 in *Psychological Review* with Roger Ratcliff, I published a paper describing the minimalist hypothesis. According to this hypothesis, readers/listeners typically process only the information they need to in order to meet their immediate goals or needs; they do not construct many of the inferences that they could construct because they are time-consuming and likely to be unnecessary. This hypothesis has been met with considerable debate, and is considered by many to have given new life to the reading and text processing areas of research. It has formed the basis of many other people's current research. My hope is that my strong statement of the view will force further empirical tests and provide the impetus for further theoretical development.

One particularly interesting implication of the minimalist hypothesis is that readers/listeners often sacrifice accuracy of understanding for speed of understanding. So long as they get the gist right, details may be ignored. Roger Ratcliff, Steve Greene (Princeton University), and I have investigated this speed/accuracy tradeoff as it applies to the understanding of pronouns. Contrary to strong claims by other psycholinguists and especially by linguists, we find that readers can leave pronouns unresolved. We also find that the resolution of pronouns depends not, as hitherto thought, on the syntactic construction of a sentence but on other more general factors of the discourse or text as a whole.

Another important implication of the minimalist hypothesis is that readers/listeners will often depend for comprehension on information that is quickly and easily available to them. They sometimes make mistakes when easily available information leads to misunderstanding (many people, when reading about the animals that Moses took, two by two, onto the Ark, notice no problem). With Roger Ratcliff, Gregory Ward, and Richard Sproat, I have experimentally demonstrated the power exerted by two kinds of easily available information, the information that immediately precedes subsequent information and well-known long-term memory knowledge.

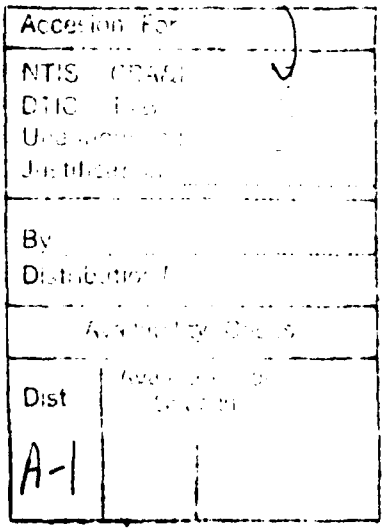
Investigation of the immediately available information in short-term memory has centered on the representation of discourse that is used in short-term memory during comprehension. Previous models have either assumed a syntactic, sentence-based, representation or a simple semantic structure that represents only the recency of concepts mentioned in the discourse and their relations to the topic of the discourse. Experiments in my lab have tested a new model, by which all concepts in a discourse have some degree of salience in memory, the degree of salience depending on a variety of factors, including the syntactic, semantic, and pragmatic contexts in which they were first mentioned. The experiments show strong effects of syntactic context that can be overridden by pragmatic manipulations.

Another project has had the goal of examining the "focus" of a discourse- what concepts are most salient at any point in the discourse. Experiments (done in collaboration with Steven Greene and Roger Ratcliff) show that pronouns are often assumed to refer to focused concepts. We draw the conclusion that pronouns are not devices that trigger a search for their referents (as has previously been proposed), but rather they serve as pointers to already focussed information. We are continuing this line of research with a wide variety of different kinds of concepts and pronouns referring to them. Currently we have found that comprehension of the referents of pronouns is markedly influenced by the verbs for which they are arguments: it is as though the verb, not the pronoun, sets up the appropriate referential structure. We have also found that prior knowledge of the people referenced by pronouns does not affect comprehension. The goal is to describe the different ways in which different kinds of verbs and other concepts interact with the pronouns used to reference them.

A new topic for investigation in psycholinguistics is the extent to which language comprehension can be investigated through statistics on language usage. Such statistics can be obtained from large corpora of texts; Roger Ratcliff and I have recently collected a corpus of about 100 million words of the New York Times, and developed software for access to it. Already, we have found information from statistics from the corpus that we could not have found in any other way. For example, linguists have claimed that the anaphor "do so" could not be used in any context for which there was not an explicit verb phrase as an antecedent. The presence of a number of counter-examples in the New York Times corpus shows this claim wrong. The linguistic claim was important because "do so" was the only anaphor thought to require an explicit surface structure representation of preceding discourse. Without "do so," it becomes possible to postulate a language comprehension system without an explicit surface structure level of representation.

Psycholinguistics has been strongly influenced by linguistic theory throughout its short history. While many psychologists would like to think of language comprehension as mainly driven by meaning, many linguists see it as equally guided by syntax. In another recent project (with Roger Ratcliff and Gregory Ward), I have found that one of the main results supporting the syntax position (a result claimed by Chomsky as an extremely important example of the benefits of cognitive science approaches) is artifactual. This finding will considerably alter the course of research into syntactic comprehension processes.

Still another project has tested two theories of memory retrieval and priming against each other. The long-held view in cognitive psychology is that memory consists of a network of concepts and pieces of knowledge, and retrieving one piece from another involves "activation" spreading from input information to other connected pieces of information. Ratcliff and McKoon (1988) proposed that the mechanism of retrieval from memory was a compound cue mechanism based on current global memory models; the compound cue mechanism was proposed as an alternative to the popular spreading activation process. McNamara has presented data that seem at first to contradict the compound cue mechanism. However, we have been able to show that his data can actually be handled quite well by compound cue theories, and that it is predictions of specific spreading activation theories, not compound cue theories, that are contradicted. Our proposal has led to five experimental papers plus an interchange in *Psychological Review*.



Selected publications supported by the Airforce/NSF grant to Gail McKoon.

Manuscript in press, Psychological Review.

Retrieving Information from Memory:

Spreading Activation Theories versus Compound Cue Theories

Roger Ratcliff and Gail McKoon

Northwestern University

Short Title: Spreading Activation vs. Compound Cue Theories

Address correspondence to Roger Ratcliff, Psychology Department, Northwestern University, Evanston, IL, 60208.

Abstract

McNamara (1992b) attacked compound cue theories on a number of grounds. Using free association as a measure of distance between concepts in memory, he argued that compound cue theories cannot explain mediated priming effects. We show that free association production probabilities do not accurately predict priming effects, either directly or in the context of current spreading activation models, and so remove the basis for McNamara's criticism. McNamara also claimed that compound cue theories cannot account for the sequential effects of items that precede a target item on responses to the target, but we show that sequential effects are consistent with compound cue models so long as the target item is weighted more heavily than the preceding items in the calculation of familiarity that determines response time and accuracy for the target. We conclude that compound cue and spreading activation theories are equally consistent with available data, and that each provides valuable impetus for the other in suggesting empirical investigations and theoretical developments.

Spreading Activation Theories versus Compound Cue Theories

Ratcliff and McKoon (1988) and Dosher and Rosedale (1989) proposed that information is accessed in memory via a process that combines the multiple cues present in the retrieval environment into a compound. In a critique of compound cue models, McNamara (1992b) addressed a large number of issues, contrasted compound cue models with their main competitors, spreading activation models, and concluded that compound cue models could do little more than "explain (experimental) results by questioning the methods or appealing to ad hoc processes." In this reply to McNamara's article, we respond to his main criticisms and show that his conclusions are misguided, and that in fact the two kinds of models are quite balanced in their abilities to account for data. We reiterate the claim made in our 1988 paper that compound cue models provide an alternative view that can be used to generate empirical investigations of retrieval that would not be suggested by spreading activation models.

The general assumptions of spreading activation theories are widely known and often thought to be intuitively clear. In contrast, compound cue theories are relatively new. An important difference between the two kinds of theories lies in their assumptions about how information presented to the retrieval system focuses on some subset of information in long-term memory. For the tasks discussed in this article, lexical decision and recognition, spreading activation theories propose that all the action in retrieval processing takes place in temporary changes to long-term memory: when an item is presented to the system, activation spreads from the representation of that item in long term memory to other nearby items in long-term memory. In compound cue theories, all the action takes place in short-term memory. Items presented to the retrieval system are assumed to join together into compounds in short-term memory. A compound is matched against information in long-term memory by a global and passive matching process. In spreading activation models, the result of retrieval processing is increased activation in long-term memory of items related to the input item. In compound cue models, the result of retrieval processing is a value indicating the familiarity of the cue compound to all the items in long-term memory. The two different sets of assumptions about retrieval offer two different ways to think about processing, about what experiments are interesting to perform, and about how to interpret data. In this way, each kind of theory is valuable to the other.

"Mediated" Priming?

In spreading activation models, items in memory vary in the number of links between them: flower and rose might be directly connected to each other whereas flower and thorn might be connected only by a mediating link through rose. Items connected by one or even two mediators should prime each other in tasks such as lexical decision because presentation of the prime word sends activation spreading to the target word, so that the target is already activated in advance of its actual presentation. In contrast, distance between items in terms of number of links is not meaningful for compound cue theories. In the SAM model for example (Gillund & Shiffrin, 1984), priming occurs when the strength value of the prime matched against some word(s) in memory is high and the strength value of the target matched against the same word(s) is also high. For example, if flower primes rose, it is because of the high strength values of both flower and rose when they are matched against flower in memory, and the high strength values of flower and rose matched against rose in memory (and perhaps also other items for which flower and rose both have high strength values). Thus, compound cue models predict priming only for items that are directly related by high strength values (or, in SAM, related via at most one other item with high strength values to both prime and target), but spreading activation models predict priming for items separated by multiple links. Because of these contradictory predictions, mediated priming has become a critical focus of the debate about the relative merits of spreading activation theories and compound cue theories.

The key issue in this debate is how the distance between two concepts in memory should be measured. A priming effect for a pair like flower-thorn contradicts compound cue theories only if it can be shown that flower and thorn are not directly related (or related by no more than one intervening item in SAM). McNamara (1992b) argued that the best measure of distance is free association production probability (page X10), and used this measure to account for priming effects which he claimed "pose difficulties to non-spreading-activation (compound cue) theories" (page X13). Specifically, he claimed on the basis of free association data that pairs of words such as flower-thorn are not directly related, which means that priming effects for these words contradict compound cue models. But this claim is wrong - free association production probability is not an accurate measure of distance for predicting priming effects. First, there exist pairs of words that prime each other even though connections between them are not produced in free association. For example, Fischler (1977), McKoon and Ratcliff (unpublished data, following Shelton & Martin, 1992), McKoon and Ratcliff (1992), and Seidenberg et al. (1984) have all shown priming for pairs of words that are not associated according to free association production measures. Second, even when free association does produce connections between words, the production probabilities do not correctly predict priming effects, as we demonstrate in the next section. Thus, free association is not a veridical measure of distance in memory and so, in the absence of data indicating otherwise, compound cue theories are free to explain priming effects for pairs like flower-thorn by assuming they are directly (albeit weakly) related, consistent with other measures such as cooccurrence statistics or relatedness judgments (McKoon & Ratcliff, 1992).[1]

Free Association Production Probabilities Do Not Accurately Predict Priming Effects

In this section, we discuss what procedures might be appropriate for using free association data to measure associative distance, and present data which allow comparison of free association production probabilities and priming effects. We then show that an explicit spreading activation model (ACT*, Anderson, 1983) cannot simultaneously account for both kinds of effects.

To present these issues, we (like McNamara, 1992b) center our discussion around two sets of pairs of words. We designate one set, from Balota and Lorch (1986) and McNamara and Altarriba (1988), the MA set, and the other set, from McKoon and Ratcliff (1992), the MR set. McKoon and Ratcliff (1992) found that the two sets of pairs gave priming effects of about the same size (14 ms and 13 ms). Primes and targets of the MA set were intended to be words connected by mediators; flower-thorn is an example. Primes and targets of the MR set were

originally intended to be words that were not connected by any mediator produced in free association; flower-root is an example. But McNamara (1992b) claimed that both sets of primes and targets did have mediators, and that the equivalent priming effects between the prime and target words of these pairs were predicted by equivalent probabilities that the primes and targets were linked through free associations. He used this to support his contention that free association is the best measure of distance between concepts in memory.

To obtain chains of mediating concepts by which primes and targets could be linked, McNamara (1992a; 1992b, Table 1) used what has been termed the "continued association" procedure (Postman & Keppel, 1970). He asked subjects to generate multiple free associates (e.g. as many as they could in 1 min) to each prime word, target word, and potential mediating word. Averaging the resulting production probabilities over all responses for both subjects and items, McNamara claimed that the chains linking primes and targets were about equally strong for the MR pairs as for the MA pairs. However, we question this claim because of McNamara's use of the continued association procedure.

Free association production probabilities can be used for a variety of purposes, including, for example, generating and norming materials to be used in experiments and for these uses the continued method may be appropriate. But when they are used to measure associative distances among concepts in memory as in McNamara's (1992b) studies, then the continued association procedure is problematic. In the earlier literature about free associations (Postman & Keppel, 1970, and precursors), it was generally accepted that this procedure allowed each next response generated from a single stimulus to be determined not only by the initial stimulus but also by the prior response or any of the other previously produced responses (see recent discussion by Nelson, Schreiber, & McEvoy, 1992). Moreover, the probabilities produced for a given stimulus by the continued procedure sum to more than 1 and so cannot be considered associative strengths for the purpose of modeling a network in which the total proportion of activation spreading from one node to each of its directly connected nodes must not sum to more than 1.0 (cf ACT*, Anderson, 1983).

The standard free association method for obtaining association strengths (avoiding the problems with the continued procedure, Postman & Keppel, 1970) is to ask subjects to give only a single response for each stimulus. We collected data with this procedure, asking subjects to generate free associates to all of the primes, potential mediators (from McNamara, 1992b), and targets for both the MA and the MR pairs. For the MA pairs, the prime and target are supposed to be linked by one mediating concept, a two-step chain. For some of the MR pairs, McNamara also proposed a two-step chain, and for others, a three-step chain. For both kinds of chains, Figure 1 shows the data we obtained, the mean first production probabilities for the directions indicated by the arrows.

Insert Figure 1 here

The important result is that the average probabilities for the two- and three-step MR chains are considerably lower than the average probabilities for the MA chains, contrary to McNamara's claims that the two kinds of pairs are equivalent. For example, for the two-step chains, the probability that a mediator is produced in response to its prime is 0.192 for the MA pairs but only 0.053 for the MR pairs. For a very simple spreading activation model, it might be assumed that when a prime is presented, some proportion of activation spreads from prime to mediator (p) and some proportion spreads from mediator to target (q), so that the activation passed from prime to target is pq. Using the production probabilities for each link to determine p and q, then multiplying along the links gives an activation value on a target of 0.0219 for the MA targets (0.192 * 0.114) but only 0.0025 for the MR two-step targets and only 0.0007 for the MR three-step targets (values from Figure 1). Over all the targets, the weighted mean value of activation for the MR targets (0.00175) is 13 times less than for the MA targets. Clearly, these values in this simple model cannot predict equivalent priming effects for the MA and MR pairs.

The difference between the MA and the MR pairs is even larger when an averaging artifact is taken into consideration. The averages just given were calculated by averaging across materials (e.g. averaging all prime to mediator links and averaging all mediator to target links) and then multiplying the averages to get activation for the target. A more appropriate way to average would be to multiply the probabilities for the chain for each item, and then average the resulting values of target activation. This way of averaging is more appropriate because for the MR pairs, it is often the case that the probability for one of the links, prime to mediator or mediator to target, is high while the other is very low. This second way of averaging increases the difference between the MA and MR pairs. For the MA pairs, multiplying probabilities before averaging gives a value of 0.0162 (cf 0.0219 above) and for the MR pairs (weighted average) a value of 0.00034 (cf 0.00175) leading to a ratio of 47:1. The same averaging problem applies to the data McNamara (1992a) collected with the continued association procedure. For the MR pairs given in his appendix, multiplying the probabilities before averaging gives a value of 0.037 as opposed to the product after averaging of 0.092 (compared with the MA value of 0.154). Thus instead of a ratio of MA to MR activation of about 1.7:1 (the value reported by McNamara), the ratio may be around 3:1 or 4:1 (we cannot calculate it exactly because we do not have the necessary MA data).

Modeling Priming Effects with an Explicit Spreading Activation Model

To develop the argument further, we examined one of the few explicit spreading activation models that has been used to make specific predictions about priming in memory. In ACT* (Anderson, 1983), activation reverberates among connected concepts. The strengths of the links from a prime to its target and the strengths from the target back to the prime both determine the total amount of activation that accrues at the target. The equations for asymptotic activation (i.e., when the system has settled to a final state) are:

$$0=n_i-p\,a_i$$

where a_i is the activation value of the ith node, p is a maintenance factor denoting the amount of activation transmitted to neighboring nodes (and usually set to 0.8 by Anderson), and n_i is the total activation to node i where

$$n_i=c_i+\Sigma_j r_{ji} a_j$$

and where r_{ji} are the connection strengths to node i and c_i is the input activation of node i. These equations appear simpler when converted to matrix form:

$$A=C+pRA,$$

and solving for A:

$$A=(I-pR)^{-1}C$$

where A is a vector (or list) of the asymptotic activation values, C is the vector of input activations, R is a matrix of connection strengths, and I is the identity matrix (a matrix with diagonal elements 1 and off diagonal elements 0). Using a system such as Mathematica, predictions for asymptotic activation values can be easily obtained using just six lines of code.

ACT* predictions[2] for relative amounts of priming were calculated for four different possible networks. The first (shown in Figure 2) was designed to represent the prime, mediator, and target along with some other nodes connected to them. The figure shows one mediator for a two-step chain between prime and target; the corresponding network for a three-step chain would have an additional mediator with three other nodes connected to it for a total of 18 nodes. The figure shows the strengths on the links leaving the prime, mediator, and target. The sums of the strengths leaving each of these nodes are set to 1.0, making the network consistent with the assumptions of ACT* (Anderson, 1983, p.22). In the matrix of connection strengths, this assumption is reflected in the fact that the strengths in each column add to 1.0. However, the network shown in Figure 2 would not be a completely acceptable representation of a semantic memory network because the nodes 4 through 14 send all of their strength back to the prime, mediator, or target (whichever of these nodes they are connected to). More realistically, each of the nodes 4 through 14 would be expected to be connected to other nodes. This means that the strength on the link from one of these nodes back to the prime, mediator, or target would have to be less than 1.0 because some of the strength leaving these nodes would have to go to their other connected nodes. So in the second possible network that was considered, it was assumed that the sum of the strengths returning from nodes 4 through 14 to P, M, or T was not 1.0 but instead that their strengths returning to P, M, or T were the same as the strengths leaving (s_a, s_b, or s_c). The third network provides a

check on results obtained from the second network; it was a larger, again more realistic, network in which each of the nodes 4 through 14 had 2 other nodes connected to them (making a total of 36 nodes). In this model, the strength of connection from the new nodes to nodes 4 through 14 was assumed to be 1.0, making a consistent network. The strengths from the nodes 4 through 14 back to P, M, or T were the same as in the second network. Finally, the fourth network, used for comparison, was a simple 3 node model with just the prime, mediator, and target (this corresponds to the top left hand 3x3 corner of the matrix in Figure 2 and should produce results similar to those obtained by simply multiplying probabilities together as was done above).

insert Figure 2 and Table 1 here

Computations from all of the networks assumed c set to 1.0 and b set to 0.8 (typical values used by Anderson, 1983). Connection strengths were derived from the production probabilities in Figure 1 for the MA pairs and the MR (two and three step) pairs. Predictions of relative amounts of priming are shown in Table 1. The first three rows show results for the first network, in which all activation returns from nodes 4 through 14 to P, M, or T, the next three rows show results for the second network, in which only some activation returns, the next two rows show results for the larger network, and the last rows show results for the simple three node network. The table shows the predicted amounts of activation on the target node after activation has been entered at one or more source nodes and the system has stabilized. We assumed as a baseline against which to measure the predicted amount of priming the case where only the target node was a source of activation, corresponding to the case where the target was presented to the system with an unrelated prime. Given this baseline, we could then predict "mediated" priming from prime to target, for which we assumed that the prime and target were sources of activation, and direct priming from the mediator to the target, for which we assumed that the mediator and target were sources of activation. Direct priming should always lead to more activation on the target than mediated priming, and this is what the predictions in the table show. For example, for the MA items, the prediction from the first network for activation on the target as a result of direct priming is 2.777, up 0.434 from baseline. The prediction for activation on the target as a result of mediated priming is 2.481, up only 0.138 from baseline. Comparing the two amounts of priming, the ratio of direct to mediated is 3.1 (shown in the fifth column of Table 1). Over all the four different networks, the ratio of direct priming for the MA pairs to mediated priming for the MA pairs is 3.1 or greater (ranging up to 6.5). Taking the low end of this range, the prediction is consistent with empirical data within typical standard errors (assuming a linear relationship between activation and reaction time, e.g., Anderson, 1983). For example, McNamara and Altarriba (1988) found 24 ms of direct priming and 10 ms of mediated priming.

The important results in Table 1 are the ratios of the predicted priming effects for the MA pairs and the MR pairs. First, the direct priming effect for the MA pairs can be compared to the mediated priming effects for the MR pairs. These predictions are not consistent with data. Over the different networks, the direct priming effect for the MA pairs (empirically 24 ms) is predicted to be from 18.9 to 234.8 times larger than the mediated priming effect for the MR pairs (which is 14 ms). But empirically, direct priming is only about 1.7 times larger. Second, the MA mediated priming effect and the MR mediated priming effect can be compared. Empirically, these effects are about the same size (about 14 ms). But ACT* predicts that MA priming should be anywhere from 5.6 to 36.0 times larger.

What can be concluded from this discussion? First, reiterating McKoon and Ratcliff's (1992) previous conclusion, free association production probabilities do not correctly predict priming effects. In this article, we demonstrate this for an explicit model, ACT*. Thus, in the context of current theories and data, free association data cannot be used to decide whether or not two items in memory are directly connected, and so, consistent with compound cue models and alternative measures of strength of connection (e.g., relatedness, co-occurrence), it is reasonable to suppose that all pairs of words that give priming are directly connected with some degree of strength. In consequence, contrary to McNamara's (1992b, p. X) claims, priming effects and free association production probabilities do not pose problems for compound cue models. But priming effects and free associations would pose problems for spreading activation models if the models assumed that free association probabilities should predict priming effects.

McNamara (1992b) acknowledges both that there are inherent problems in measuring distances between items in memory, and that measures like free association may not be definitive (page X14). It is important to understand why they are not definitive: It is not the case that free association is "probably" an accurate measure, if we could only get enough subjects to generate enough responses. Instead, as is exemplified by the exercise above with ACT*, free association clearly fails as a predictor of priming. As a result, both spreading activation and compound cue models need to provide a theoretical account of how free associations and priming effects can be related to each other and of how they can both be related to other variables such as semantic relatedness and co-occurrence frequencies which might be more direct predictors of priming effects (see McKoon & Ratcliff, 1992, for discussion of these variables).

Sequential Effects

McNamara (1992a, 1992b) argued that sequential (lag) effects among multiple lexical decision tests cannot be explained by compound cue theories. McNamara's argument began with a demonstration that, for a particular set of experimental procedures, a compound used to retrieve information from memory about a target word must contain the two items preceding the target as well as the target. McNamara demonstrated this by showing facilitation for a target when the related word that preceded it was separated by an intervening word (see also Ratcliff & McKoon, 1978; Ratcliff, Hockley, & McKoon, 1985). For example, for the sequence hammer, vase, nail, response time for the target (nail) was facilitated. McNamara's point was that the facilitation could only come about if the related word (hammer) were included in the compound cue, which means the compound must contain all three words in the sequence.[3]

Then McNamara considered sequences like hammer, nail, vase, in which the first and second words are related to each other but not to the third word. We label these words the preprime, prime, and target items, respectively. McNamara (1992a, 1992b) pointed out that response time for the target item in such a sequence should be facilitated, because the compound used to access memory for the target must also contain the related prime and preprime. When such facilitation was not found in his experiments, McNamara concluded that the compound cue prediction failed.

What is wrong with this conclusion is that it is based on assumptions in McNamara's application of compound cue theory that are not reasonable, assumptions about the relative weightings of the preprime, prime, and target in the calculation of the total familiarity value for the target. When more reasonable weightings are assumed, the amount of predicted facilitation is too small to have been detected in any experiments that have been conducted.

Insert Table 2 and Figure 3 here

Table 2 shows quantitative predictions for several kinds of sequences generated from a compound cue model based on SAM (Gillund & Shiffrin, 1984; Ratcliff & McKoon, 1988). The predictions were derived for the simplified memory structure shown in Figure 3 (see Table 1, Ratcliff & McKoon, 1988) in which each cue word is related with strength 1.0 to itself in memory, it is related with strength 1.0 to each of two related other words in memory (which are in turn related back to the cue word with strength 1.0), and it is related to all other items in memory (and they are related to it) with strength 0.2. To determine the familiarity value for the target (see Figure 3), the strength values for the preprime, prime, and target cue words are weighted differently, with most weight on strength values for the target because it is the word that actually requires a response, and the weighted strength values are summed over all items in memory.

To argue that SAM should predict facilitation for sequences in which the preprime and prime are related to each other but not the target, McNamara used a weighting scheme of 0.5 on the target, 0.3 on the prime, and 0.2 on the preprime. This scheme places a lot of weight on the preprime and prime relative to the target. It means that if the prime and preprime were nonwords and the target a word, equal weight would be given in the decision process to the nonwords (preprime and prime) as to the target word, and a 50% error rate on the target word would be expected. We believe that this is not a reasonable choice for a weighting scheme, and several others are presented in Table 2. The results show that McNamara's claim depended on the excessive weighting of the prime and preprime.

Table 2 shows familiarity values for a range of weighting schemes for several kinds of sequences, and the resulting predictions for priming effects (in the rightmost three columns). The empirical constraints that the predictions must meet are straightforward (from McNamara, 1992a, Experiment 2): First, the

familiarity value on the target should be lowest when neither preprime nor prime is related to it (baseline = UUU) and highest when the prime is related to it (URR). McNamara (1992a obtained a difference between these two conditions of 30 ms. The familiarity value on the target should also be higher than baseline when the preprime is related to it (RUR); McNamara obtained a difference for these two conditions of 14 ms in one experiment and 21 ms in another experiment. Most importantly, the familiarity value on the target should not be distinguishably higher than baseline when the preprime and prime are related to each other but not the target (RRU); for these two conditions, McNamara found no significant difference in response times.

With McNamara's weighting scheme (0.2, 0.3, 0.5), the URR priming effect in terms of familiarity value is 0.30, the RUR priming effect is 0.19, and the RRU effect is 0.10. The RRU effect is one third the size of the URR effect, and so should be observable empirically. But if the weight on the target is increased to 0.6 and the weights on the preprime and prime decreased accordingly, then the RRU effect is only about one ninth the size of the URR effect and it would be unlikely that this could be detected empirically. The URR effect is 30 ms, and one ninth of that would only be about 3 or 4 ms. The other weighting schemes shown in Table 2 also predict an RRU effect too small to be observed.

The conclusion to be drawn from the results displayed in Table 2 is clear: the difference predicted by a compound cue version of SAM between response times in the RRU condition and the baseline UUU condition is too small to be observable empirically (except possibly in an extremely large experiment with low variance). This conclusion holds for reasonable relative weights on preprime, prime, and target. Only when excessive weight is given to the preprime and prime does SAM predict an effect large enough to be observable. Thus, the data provided by McNamara (1992a) are not inconsistent with compound cue models.

The effect of related preprime and prime on target response times was one sequential effect with which McNamara (1992a, 1992b) criticized compound cue theory. A second effect was an inhibition on targets that appeared when the preprime item was a nonword. The four conditions that McNamara (1992a, 1992b) examined are shown in Table 3: the target word was preceded by either a related prime or an unrelated prime, and the prime was preceded by either a word or a nonword (a nonword is indicated in the table by an X). McNamara's results (1992a, Table 7) are given at the bottom of Table 3. He found that a nonword preprime slowed responses overall, but it did not significantly affect the amount of facilitation given by a related prime to a target (the two priming effects shown in Table 3, 26 ms and 33 ms, were not significantly different from each other).

Insert Table 3 here

McNamara (1992a, 1992b) claimed that compound cue theories could not accommodate this pattern of results, but again predictions depend on the weighting scheme for the preprime, prime, and target. Table 3 shows predictions with two different sets of weights (the same specific model was used as for the results in Table 2, and the strength connecting any cue word to a nonword in memory was assumed to be 0.1). The predictions fit the data remarkably well. The main effect of inhibition by a nonword preprime appears as lower values of familiarity in the XUU and XRR conditions which compares well with the observed increase in reaction times for these two conditions compared with UUU and URR. The priming effect is predicted to be only slightly larger when the preprime is a word than when it is a nonword, in accord with the null effect in McNamara's data. Simultaneously, SAM correctly predicts the relative size of the RUR priming effect. Thus, contrary to McNamara's claim, the SAM compound cue model gives an excellent fit to a complicated pattern of data (and may also apply to choice reaction time sequential effects, see McKoon & Ratcliff, 1992) while spreading activation models require the addition of an explicit reaction time model for sequential effects.[4,5]

Naming

Researchers interested in priming effects have often argued that theories designed to explain such effects should link priming in lexical decision with priming in the task of naming a word because both tasks involve accessing the lexicon and because similar experimental variables have been examined in the two tasks (cf McNamara, 1992b; Neely, 1991). In contrast, we have argued that priming in lexical decision has a natural affinity with priming in recognition memory. It is our strong bias to attempt to generalize research domains in terms of underlying theoretical mechanisms, and in theoretical terms, both lexical decision and recognition require an item to be encoded and compared with memory to produce a binary decision. Naming a word, on the other hand, is a task for which one out of tens of thousands of possible responses must be produced. McNamara (1992b) criticizes compound cue theories because they fail to explain priming effects in naming, but models that deal with naming and lexical decision could be similarly criticized because they do not deal with recognition memory.

Although we are biased against relating naming and lexical decision through empirical considerations, it may be possible to relate them theoretically by implementing a compound cue mechanism in models of naming. Memory models in which compound cue mechanisms have been implemented are parallel processing models. This characteristic suggests Seidenberg and McClelland's (1989) model for lexical decision and naming as a candidate to implement a compounding mechanism. In Seidenberg and McClelland's (1989) model, orthographic and phonological units each form two distinct levels of representation linked by a hidden layer of units. To model compounding, gradual (stochastic) replacement of one item by the next item (e.g., with exponential probability of a feature being replaced) would allow the representation at input to be a compound, a combination of features from the current and prior items, and this compound could percolate through the whole network. To produce semantic priming effects, it would be necessary to add an explicit (as yet unimplemented) semantic layer of information. Then the semantic layer could represent semantic feature overlap so that a compound of related items would produce a better match to memory and faster responses. To assess whether such a marriage of models could account for priming in naming, testing and data fitting would be required as well as development of a representation system for the semantic layer.

Conclusions

McNamara (1992b) claimed that compound cue theories could not account for mediated priming effects and sequential effects. We demonstrated that compound cue models could account for these effects by exploring them in the joint context of empirical data and specific models. We also found that the juxtaposition of spreading activation and compound cue models suggested new ways to view some empirical phenomena. Our findings can be summarized by the following points:

McNamara (1992b) claimed that some sequential effects are inconsistent with compound cue models. However when the familiarity of a sequence was calculated with reasonable weights on the strengths of the different items in the sequence, compound cue models fit the data quite well.

McNamara (1992a; 1992b) failed in his effort to demonstrate multiple-step priming because predictions derived from his method of measuring distances between concepts in memory (free association production probability) are not consistent with observed data.

Neither current spreading activation models (such as ACT*) nor compound cue theories can jointly predict free association production probabilities and priming effects. Variables other than free association, including semantic relatedness and cooccurrence measures, may predict priming effects but these measures need more investigation, both empirical and theoretical, in order to relate them to priming.

Words that prime each other may be directly related to each other in memory, and therefore priming effects among them are consistent with compound cue theories. Since we currently have no empirical method for measuring distance in semantic memory, words that seem far apart may instead be weakly directly related. A corollary of this point is that any individual word may have literally hundreds of associates, most of which are weakly but directly related. A memory system made up of large numbers of weak but direct associates is consistent with compound cue models of retrieval and with the intuition that any word can appear in many (perhaps hundreds) of familiar combinations with other words (see McKoon & Ratcliff, 1992).

Free association data suggest that a word in memory has many other words associated to it. When this is taken into account, the utility of spreading activation as a general retrieval mechanism must be viewed with suspicion. Suppose each word had 20 other words that it activated to a non-trivial degree (see Postman & Keppel, 1970). Then with 3-step priming in a spreading activation model, 20x20x20=8000 words would be activated; this is a good

proportion of the adult lexicon. Or, if a single word activated 40 other words, then 64,000 words would be activated by 3-step priming, about the number of words in the adult lexicon. Such rampant spread of activation through memory would severely reduce the utility of the spreading activation process as a general retrieval mechanism.

Spreading activation has been almost unchallenged as an explanation of priming phenomena, and has remained so despite the development of parallel processing and feature models that are inconsistent (to various degrees) with it. The debate represented in this article contributes to a long overdue examination of spreading activation, as well as additional evidence in support of compound cue theories as viable alternatives.

References

Anderson, J.R. (1983). The architecture of cognition. Cambridge: Harvard University Press.

Balota, D.A. & Lorch, R.F. (1986). Depth of automatic spreading activation: Mediated priming effects in pronunciation but not in lexical decision. Journal of Experimental Psychology: Learning, Memory, and Cognition, 12, 336-345.

Dosher, B.A., & Rosedale, G. (1989). Integrated retrieval cues as a mechanism for priming in retrieval from memory. Journal of Experimental Psychology: General, 2, 191-211.

Falmagne, J.C. (1965). Stochastic models for choice reaction time with applications to experimental results. Journal of Mathematical Psychology, 12, 77-124.

Fischler, I. (1977). Semantic facilitation without association in a lexical decision task. Memory and Cognition, 5, 335-339.

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

Joordens, S. & Besner, D. (1992). "Priming" effects that span an intervening unrelated word: Implications for models of memory representation and retrieval. Journal of Experimental Psychology: Learning, Memory, and Cognition.

LeSueur, L. L. (1990). On metaphors and associations. Unpublished doctoral dissertation, Vanderbilt University.

McKoon, G., & Ratcliff, R. (1979). Priming in episodic and semantic memory. Journal of Verbal Learning and Verbal Behavior, 18, 463-480.

McKoon, G. & Ratcliff, R. (1992). Spreading activation versus compound cue accounts of priming: Mediated priming revisited. Journal of Experimental Psychology: Learning, Memory, and Cognition.

McNamara, T.P. & Altarriba, J. (1988). Depth of spreading activation revisited: Semantic mediated priming occurs in lexical decisions. Journal of Memory and Language, 27, 545-559.

McNamara, T.P. (1992a). Theories of Priming: I. Associative distance and lag. Journal of Experimental Psychology: Learning, Memory, and Cognition. In press.

McNamara, T.P. (1992b). Priming and constraints it places on theories of memory and retrieval. Psychological Review.

Neely, J.H. & Durgunoglu, A. (1985). Dissociative episodic and semantic priming effects in episodic recognition and lexical decision tasks. Journal of Memory and Language, 24, 466-489.

Nelson, D. L., Schreiber, T. A., & McEvoy, C. L.. (1992). Processing implicit and explicit representations. Psychological Review, 99, 322-348.

Postman, L. & Keppel, G. (1970). Norms of Word Association. New York: Academic Press.

Ratcliff, R., Hockley, W.E., & McKoon, G. (1985). Components of activation: Repetition and priming effects in lexical decision and recognition. Journal of Experimental Psychology: General, 114, 435-450.

Ratcliff, R. & McKoon, G. (1978). Priming in item recognition: Evidence for the propositional structure of sentences. Journal of Verbal Learning and Verbal Behavior, 17, 403-417.

Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. Psychological Review, 95, 385-408.

Ratcliff, R., Sheu, C-F., & Gronlund, S. (1992). Testing Global Memory Models using ROC Curves. Psychological Review, 99, 518-535.

Remington, R.J. (1969). Analysis of sequential effects in choice reaction times. Journal of Experimental Psychology, 82, 250-257.

Seidenberg, M.S., & McClelland, J.L. (1989). A distributed, developmental model of word recognition and naming. Psychological Review, 96, 523-568.

Seidenberg, M.S., Waters, G.S., Sanders, M., & Langer, P. (1984). Pre- and postlexical loci of contextual effects on word recognition. Memory and Cognition, 12, 315-328.

Shelton, J., & Martin, R. How semantic is automatic priming? In press, Journal of Experimental Psychology: Learning, Memory, and Cognition.

Author Note

This research was supported by NIMH grants HD MH44640 and MH00871 to Roger Ratcliff and NSF grant 85-16350, NIDCD grant R01-DC01240, and AFOSR grant 90-0246 (jointly funded by NSF) to Gail McKoon. We thank Tim McNamara for extensive discussions which we hope have resulted in an agreement on the points of agreement and disagreement in this debate.

Footnotes

1. McNamara (1992b) suggests that an experiment by Ratcliff and McKoon (1978) provides evidence against cooccurrence as a predictor of priming. Ratcliff and McKoon measured the amount of priming due to temporal contiguity, that is, the nearness of words to each other in a sentence. They found that the amount of priming due to temporal contiguity was less than that due to propositional distance. McNamara (1992b) identified cooccurrence as being necessarily closely related to temporal contiguity and less related to propositional distance. However, cooccurrence as presently defined includes propositional, temporal, and even between-sentence effects, and so Ratcliff and McKoon's results currently have no implications for the use of cooccurrence measures.

2. ACT* relates link strength to node strength by requiring that link strength $r_{ij}=s_j/\sum_k s_k$ where s_k are all the nodes connected to node i (including s_j). The problem is that for most networks that are relatively interconnected, it is impossible to obtain node strengths for all the nodes in the network that satisfy this equation for all link strengths. This can be seen easily with a 3 node network and 6 links all set to different nonzero values with r_{ij} summing to 1 for the 2 links leaving node i. In this case, no solution can be found, and in general, unless there are fewer nonzero interconnection or link strengths than nodes, nontrivial solutions are not possible. This means that node strengths cannot be assigned on the basis of link strengths and so the input activation of a node c_i cannot depend on a value of node strength derived from link strengths, as assumed in ACT*. We have no independent measure of node strength for the items modeled here, so all node strengths were set to 1.

3. Joordens and Besner (1992) have criticized compound cue theory because, they claim, it cannot predict priming effects when an item intervenes between a related prime and target. This is clearly false; Ratcliff and McKoon (1988) showed exactly how compound cue models predict such effects (see also McNamara, 1992a, 1992b).

4. A third sequential effect that McNamara (1992b) marshals in his critique of compound cue theories involves sequences of only two items, not three. He

points out that compound cue theories should predict slower response times on a positive target when it is preceded by a negative test item because the negative item will cause the familiarity of its compound with the target to be low. Sequential effects have been demonstrated in choice reaction time (Remington, 1969; Falmagne, 1965) as mentioned above. McNamara cites two sets of data for which the predicted effect does not hold (LeSueur, 1990; Neely & Durgunoglu, 1985). However, there are other sets of data which do show the predicted effect (cf Ratcliff, Sheu, & Gronlund, 1992, Experiment 1, and also sequential effects in choice reaction time, Falmagne, 1965; Remington, 1969).

5. McNamara (1992b) also considered sequential effects that involve neutral prime items (a neutral prime is a word like ready, presented many times over the course of an experiment). Empirical results currently suggest that some effects of neutral primes may be different in lexical decision (McNamara, 1992 manuscript) and recognition.

Figure Captions

Figure 1: Free association production probabilities (means across subjects and items) from the single response procedure for the MA pairs (McNamara & Altarriba, 1988), the MR two-step pairs (McKoon & Ratcliff, 1992, with McNamara's, 1992, mediators), and the MR three-step pairs (McKoon & Ratcliff, 1992, with McNamara's, 1992, mediators).

Figure 2: A network for spreading activation computations for ACT* and a matrix of the strengths of connections between nodes. For ACT*, the weights leaving a node are assumed to sum to 1, so strengths in each column of the matrix sum to 1.

Figure 3: The retrieval structure for the SAM model used in modeling priming effects.

Table 2: Familiarity of Various Pre-prime, Prime and Target Relationships

	Preprime, Prime, Target				Familiarity minus baseline (UUU)		
	UUU	RRU	RUR	URR	RRU	RUR	URR
Weights 0.1, 0.2, 0.7 Strengths 1 and 0.2	3.58	3.61	3.73	3.90	0.03	0.15	0.32
Weights 0.15, 0.15, 0.7 Strengths 1 and 0.2	3.58	3.61	3.81	3.81	0.03	0.23	0.23
Weights 0.1, 0.3, 0.6 Strengths 1 and 0.2	3.46	3.50	3.57	3.86	0.04	0.11	0.40
Weights 0.15, 0.25, 0.6 Strengths 1 and 0.2	3.66	3.70	3.78	4.01	0.04	0.12	0.35
Weights 0.14, 0.29, 0.57 Strengths 1 and 0.2	3.41	3.47	3.56	3.77	0.06	0.15	0.36
Weights 0.2, 0.3, 0.5 Strengths 1 and 0.2	3.34	3.44	3.53	3.64	0.10	0.19	0.30
Weights 0.1, 0.4, 0.5 Strengths 1 and 0.2	3.39	3.45	3.47	3.84	0.06	0.08	0.45

Note. UUU means that none of the words are related, RRU means that the preprime and prime are related (e.g., hammer, nail, veil in a sequence), RUR means the preprime and target are related, and URR means the prime and target are related.

Table 1: Predictions from ACT* for Mediated and Nonmediated Pairs

	number of nodes	Baseline Activation	Mediated Activation	Direct Activation	Ratio MA Direct to Mediated	Ratio MA mediated to mediated
MA, all activ returns	14	2.343	2.481	2.777	3.1	1
MR 2-step, all activ returns	14	2.550	2.573	2.766	18.9	5.6
MR 3-step, all activ returns	18	2.288	2.302	2.589	31.0	9.1
MA, some activ returns	14	1.151	1.172	1.272	5.8	1
MR 2-step, some activ returns	14	1.169	1.172	1.246	40.3	7.0
MR 3-step, some activ returns	18	1.144	1.145	1.227	121.0	21.0
MA large network	36	1.315	1.345	1.468	5.1	1
MR 2-step, large network	36	1.333	1.337	1.399	41.4	8.1
MA small network	3	1.0103	1.0247	1.1042	6.5	1
MR 2-step, small network	3	1.0017	1.0033	1.0403	58.7	9.0
MR 3-step, small network	4	1.0073	1.0077	1.0706	234.8	36.0

Note. The ratio of direct priming to mediated priming and mediated to mediated is the ratio of the differences between the condition and baseline. Ratios based on the probabilities from free association (ratios of probabilities or ratios of products of probabilities) are MA mediated to direct, 5.2, ratio for MR 2-step mediated to MA direct is 44.8, and ratio of MR 3-step mediated to MA direct is 158.3. MA stands for the McNamara & Altarriba (1988) materials and MR for the McKoon & Ratcliff (1992) materials.

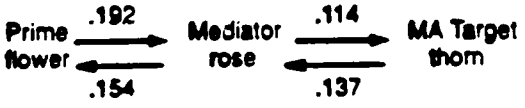
Table 3: Familiarity and Reaction Times for Priming Pairs Preceded by Words and Nonwords

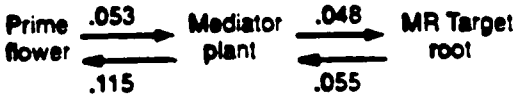
	Preprime, Prime, Target				Priming Effect: Condition-Baseline		
	UUU	URR	XUU	XRR	URR -UUU	XRR -XUU	RUR -UUU
Weights 0.1, 0.2, 0.7	3.58	3.90	3.25	3.54	0.32	0.29	0.15
Weights 0.14, 0.29, 0.57	3.41	3.77	2.96	3.28	0.36	0.32	0.15
Reaction Times (ms)	562	536	593	560	-26	-33	-14

Note: Reaction times are from McNamara's (1992a) Table 7. Note that the reaction times and familiarity value differences have opposite signs because smaller reaction times correspond to higher familiarity values for positive responses. The RUR condition comes from Table 2.

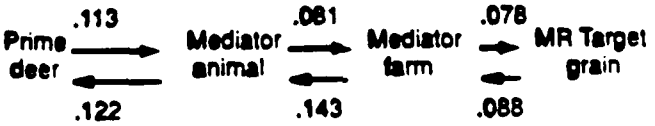
Free Association Data

Two-Step Chains





Three-Step Chains

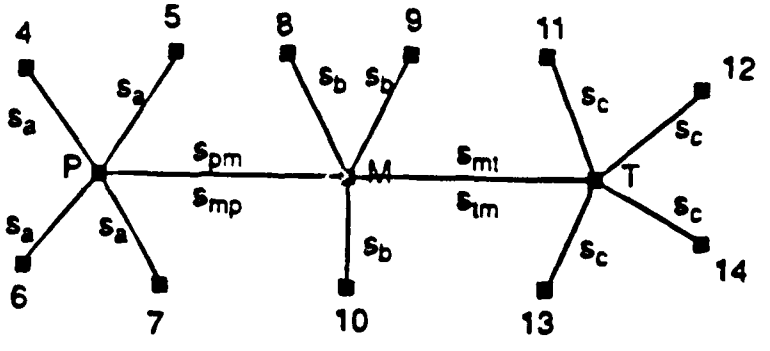


	Target									
Cue	1	2	3	4	5	6	7	8	9	10
1	1	1	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
2	1	1	1	0.2	0.2	0.2	0.2	0.2	0.2	0.2
3	0.2	1	1	1	0.2	0.2	0.2	0.2	0.2	0.2
4	0.2	0.2	1	1	1	0.2	0.2	0.2	0.2	0.2
5	0.2	0.2	0.2	1	1	1	0.2	0.2	0.2	0.2
6	0.2	0.2	0.2	0.2	1	1	1	0.2	0.2	0.2
7	0.2	0.2	0.2	0.2	0.2	1	1	1	0.2	0.2
8	0.2	0.2	0.2	0.2	0.2	0.2	1	1	1	0.2
9	0.2	0.2	0.2	0.2	0.2	0.2	0.2	1	1	1
10	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	1	1
11	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
12	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1

Note that cues 11 and 12 are assumed to be nonwords with strengths 0.1, the residual strengths from word cues to other words are assumed to be 0.1 and the strengths of words connected to each other is assumed to be 1. Familiarity is computed from

$$F(\text{cue i, cue j, cue k}) = \sum_{l} S_{il}^{w1} S_{jl}^{w2} S_{kl}^{w3}$$

where S_{il}^{w1} is the strength of cue i to target l with weight w1.



From Node

To Node	P	M	T	4	5	6	7	8	9	10	11	12	13	14
P	0	s_{mp}	0	1	1	1	1	0	0	0	0	0	0	0
M	s_{pm}	0	s_{tm}	0	0	0	0	1	1	1	0	0	0	0
T	0	s_{mt}	0	0	0	0	0	0	0	0	1	1	1	1
4	s_a	0	0	0	0	0	0	0	0	0	0	0	0	0
5	s_a	0	0	0	0	0	0	0	0	0	0	0	0	0
6	s_a	0	0	0	0	0	0	0	0	0	0	0	0	0
7	s_a	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	s_b	0	0	0	0	0	0	0	0	0	0	0	0
9	0	s_b	0	0	0	0	0	0	0	0	0	0	0	0
10	0	s_b	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	s_c	0	0	0	0	0	0	0	0	0	0	0
12	0	0	s_c	0	0	0	0	0	0	0	0	0	0	0
13	0	0	s_c	0	0	0	0	0	0	0	0	0	0	0
14	0	0	s_c	0	0	0	0	0	0	0	0	0	0	0

A PRAGMATIC ANALYSIS OF SO-CALLED ANAPHORIC ISLANDS

GREGORY WARD — *Northwestern University*
RICHARD SPROAT — *AT&T Bell Laboratories*
GAIL MCKOON — *Northwestern University*

It is commonly assumed that words are grammatically prohibited from containing antecedents for anaphoric elements, and thus constitute 'anaphoric islands' (Postal 1969). In this paper, we argue that such anaphora—termed OUTBOUND ANAPHORA—is in fact fully grammatical and governed by independently motivated pragmatic principles. The felicity of outbound anaphora is shown to be a function of the accessibility of the discourse entity which is evoked by the word-internal element and to which the anaphor is used to refer. The morphosyntactic status of the antecedent is but one factor affecting the accessibility of that entity. A series of psycholinguistic experiments support the analysis.*

INTRODUCTION

1. For over twenty years, various attempts have been made to rule out word-internal antecedents for anaphoric elements. The first such attempt is found in Postal 1969, where contrasts such as the one between 1a and 1b are discussed (p. 230):

(1) a. Hunters of animals tend to like them. [them = animals]
b. *Animal hunters tend to like them.

To account for the deviance of examples like 1b, Postal argued that words such as *animal hunters* constitute a type of ANAPHORIC ISLAND—'a sentence part ... which cannot contain the antecedent structure for anaphoric elements lying outside' (1969:205). In particular, he proposed the following constraint on what he termed OUTBOUND ANAPHORA: for any word (W1), no anaphor could have as an antecedent another word which is either 'part of the sense of' W1 or morphologically related to W1.

While Postal's observations concerning so-called anaphoric islands were originally cited as evidence for the theory of Generative Semantics, these observations have more recently been cited as evidence for particular views of the relation between morphology and syntax. What is common to these disparate theories is the assumption that there exists some kind of GRAMMATICAL prohibition against the kind of anaphora illustrated in 1b.

In this paper we argue that outbound anaphora is not ruled out by any principle of grammar: morphemes in word-internal positions, for example, may serve as antecedents for subsequent anaphora. Our analysis presupposes a sharp distinction between syntax and pragmatics. In particular, we assume that a genuinely ungrammatical construction is ungrammatical in all (nonmetalinguistic) contexts, and cannot be 'amnestied' by pragmatic or discourse factors. Given this assumption, we maintain that outbound anaphora is fully gram-

* We wish to thank the following people for useful comments and data: Betty Birner, Mary Dalrymple, Julia Hirschberg, Judy Levi, Beth Levin, Janet Pierrehumbert, Roger Ratcliff, Mats Rooth, audiences at Northwestern University and the University of Pennsylvania, and two anonymous reviewers. This research was supported in part by NSF grant BNS85-16350 to Gail McKoon and by AFOSR grant #90-0246 (jointly funded by NSF) to Gail McKoon.

matical and governed by independently motivated pragmatic principles. In this way, our approach is similar to that of Reinhart 1983, in which it is argued that, aside from cases of bound anaphora, the grammar need not make any special statement about the referential possibilities of anaphoric elements.

For the purposes of this study, we adopt a conventional view of the notion 'word'. We will consider a word to be any combination of a stem and affixes (normally written as one orthographic word in English), or any compound (which may consist of more than one orthographic word in English). This usage of the term is consistent with most of the work in morphology, including Matthews 1974, Aronoff 1976, Bauer 1983, and Mohanan 1986, inter alia.

We begin with a review of previous studies of anaphoric islands in general and outbound anaphora in particular, pointing out inadequacies. Next, we present our pragmatic account of outbound anaphora, and argue that the interpretability of an anaphor is a function of the relative accessibility of the discourse entity to which the anaphor is used to refer; the morphosyntactic status of the antecedent of the anaphor is only one factor which affects the relative accessibility of that entity. As part of our discussion we will review the results of a series of psycholinguistic experiments that support our analysis.

Previous literature

2.1. Anaphoric islands and Generative Semantics. To the best of our knowledge, Postal (1969) was the first to claim that—as he put it—reference both into and out of words is ungrammatical. Consider his examples of outbound anaphora in 2:[1]

(2) a. *Max is an *orphan* and he deeply misses *them*. (orphan = 'a child whose *parents* have died') (Postal 1969:206, ex. 3a)
b. *The best *pork* comes from young *ones*. (pork = 'meat from *pigs*') (Postal 1969:226, ex. 100b)
c. *Max wanted to *glue the boards together* but Pete wanted to *do so* with tape. (glue = '*fasten* with glue'). (Postal 1969:212, ex. 35b)
d. **McCarthy*ites are now puzzled by *his* intentions. (Postal 1969:213, ex. 42b)
e. *The best *wombat*meat comes from young *ones*. (Postal 1969:226, ex. 100a)
f. **Smoker*s really shouldn't *do so*. (Postal 1969:217, ex. 65b)

[1] In these and all subsequent examples, we shall adopt the convention of italicizing intended coreferential expressions, with the following stipulations: (i) whenever a word-internal expression is phonologically or orthographically unmodified within the containing word, we italicize just the portion of the word which corresponds to the intended antecedent (e.g. *Bush* supporters, *flut*ist, *New Yorker*, *smoker*); (ii) if the containing word is not so clearly segmentable, we italicize the entire containing word (e.g. *Belgian*, *Glaswegian*, *second*). Furthermore, we shall represent greater than normal intonational prominence (where relevant) with small capitals. Finally, in our review of previous studies we shall be using the annotations of unacceptability used by the original authors (usually '*'). Elsewhere, however, we shall be using the symbol for pragmatic deviance ('#'), given our claim that outbound anaphora involves no grammatical violation.

On the basis of such data, Postal concluded that coreferential pronouns (e.g. 2a), 'identity of sense' pronouns (e.g. 2b), and the pro-VP *do so* (e.g. 2c) cannot be anaphorically related to words that constitute 'part of the meaning' of another word in the sentence. Even if a word is morphologically present within another word, Postal claimed, it still cannot serve as an antecedent for these anaphoric elements, as illustrated in 2d–f.

Postal also argued that anaphoric elements themselves may not occur as part of the sense of a word, nor may they be morphologically incorporated into a word. Such anaphora, which he termed INBOUND ANAPHORA, is exemplified in 3:

(3) a. *The grolf wanted to visit *Max*. (grolf = 'one who has written the biography of *X*') (Postal 1969:206, ex. 11a)
b. *The boy who owned a flark made fun of Max's *gorilla*. (flark = 'a device for removing the pelt of *one*') (Postal 1969:210, ex. 25a)
c. *The fact that Max plorbed Betty did not convince Pete to *kiss* her on the lips. (plorb = '*do so* on the lips') (Postal 1969:213, ex. 39a)
d. **McCarthy* was glad that *him*ites were the majority in the room. (Postal 1969:214, ex. 50a)
e. *Harry was looking for a rack for *magazines* and he found a *one*-rack. (Postal 1969:216, ex. 60b)
f. *People who *smoke* like other *do so*ers. (Postal 1969:217, ex. 69a)

In 3a–c we see that anaphors may not occur as part of the sense of a word, while in 3d–f we see that anaphors may not be morphologically incorporated in lexical items. Thus, both simple and derived morphological forms are claimed to be anaphoric islands with respect to both outbound and inbound anaphora.

As Postal noted, some of these data seemed problematic for the theory of Generative Semantics and would appear to provide good support for the alternative theory of Interpretive Semantics then under development. Recall that in Generative Semantics it was posited that a word such as *orphan* might actually be represented syntactically by the phrase *a child whose parents have died*. It was therefore something of a puzzle that one could not refer to the deceased parents with an anaphor, as illustrated in 2a. By contrast, in Interpretive Semantics words were not decomposed into underlying syntactic representations; this theory was therefore not required to explain examples of ill-formed outbound anaphora like those in 2a–c or the absence of words with the characteristics required to yield examples like those in 3a–c.

Interestingly, Postal marshaled the anaphoric-island data as evidence FOR rather than against Generative Semantics. First, while Interpretive Semantics could explain the lack of inbound anaphora in cases like 3a–c, it could not explain the absence of forms like **himite*, **oner*, or **do soer* in 3d–f without some additional constraint. Generative Semantics, however, coupled with an anaphoric-island constraint applying late in the derivation of sentences, could give a uniform account of why ALL such cases of inbound anaphora are ill-

formed. Similarly, while Interpretive Semantics could handle cases of outbound anaphora like 2a–c, it could not without additional stipulation account for those in 2d–f. For example, given that *McCarthy* is morphologically present in *McCarthyites*, there should be no reason on an interpretive account why it could not function as an antecedent for the anaphor in 2d. Again, with an additional late anaphoric island constraint, Generative Semantics could provide a uniform account of all of the examples in 2. Given these assumptions, *McCarthyites* and *orphan* are treated alike, since both would be marked as anaphoric islands late in the derivation and both would be equally 'inaccessible' to subsequent anaphora. Finally, Postal argued that a late anaphoric-island constraint was in fact required on independent grounds. He presented evidence that relational adjectives such as *American* in *the American attempt to invade Cuba* are derived from underlying full NPs (see also Levi 1978); indeed, as this example shows, the underlying NP can evidently serve as the antecedent for the deleted subject of the embedded clause *to invade Cuba*. Yet such adjectives nonetheless constitute islands, according to Postal, who offered as evidence the examples in 4 (1969:223):

(4) a. **Her* enemies were pleased by the *American* invasion of Vietnam.
b. **America* praised the *it*an invasion of Cuba.

Thus, Postal concluded, there must be some kind of constraint that marks simple and derived words as anaphoric islands fairly late in the derivation of sentences, at least after the application of the rule converting noun phrases into relational adjectives. Given that a late anaphoric-island constraint appeared independently necessary, Generative Semantics stood in a better position than Interpretive Semantics to account for these data; only the former could readily explain parallels between words that only underlyingly 'contained' antecedents or anaphors and words that morphologically contained antecedents or anaphors. It was thus taken to be an advantage of Generative Semantics that it is only on the surface that, say, *pork* and *wombatmeat* consist respectively of one and two morphemes; the anaphoric-island constraint treats them identically with respect to outbound anaphora.

Ross attempted to pinpoint the stage in the derivation at which the anaphoric-island constraint applies, claiming that 'it is perfectly possible for pronouns to appear in the course of a derivation which refer to NPs "inside" words, as long as these pronouns do not eventually appear in surface structures' (1971:599). For example, in 5 the ellipted VP is *justify herself*, where *herself* clearly has *Britain*, part of *British*, as its antecedent (Ross 1971:599, ex. 2):

(5) I approve of *America*'s attempt to justify *herself*, but I don't approve of the British attempt (to).

To handle such data, Ross suggested that the anaphoric-island constraint is triggered only by pronouns which are present in surface structure. The fact that the implicit reference to *Britain* in 5 is possible was taken by Ross to be further support for Generative Semantics.[2]

[2] It is interesting that Ross appears to have overlooked the fact that the omitted *herself* does not have *America* or *British* as a direct antecedent, at least not in the theory of transformational syntax assumed at the time (nor, for that matter, in current Government-Binding theory). Rather,

Before we proceed further with the discussion, it is worth bearing in mind two points concerning grammatical theory at the time of the early discussions of anaphoric-island phenomena. First, most researchers in generative syntax then had little interest in morphology per se; hence, there was often no attempt to distinguish cases in which an antecedent is morphologically contained within another word from cases in which the two words are merely morphologically RELATED (though see the discussion of Browne 1974 below). Second, early studies in the generative framework viewed anaphora as a relationship—either a transformational one or one involving some sort of indexing—between two positions in a syntactic structure. The view that words were anaphoric islands therefore constituted, in effect, a syntactic constraint. While we do not deny that syntax may constrain at least one kind of anaphora, namely bound anaphora, we shall assume, as argued in Reinhart 1983, that unbound pronouns are not indexed or otherwise structurally related to their antecedents. Rather, following Karttunen 1976, Grosz 1977, Morgan 1978, Webber 1979, Sidner 1979, and Grosz & Sidner 1986, inter alia, we assume that such reference is more accurately seen as a relation between language and discourse entities, which constitute part of a speaker's (continuously updated and revised) model of the ongoing discourse.

2.2. THE GRADIENT NATURE OF OUTBOUND ANAPHORA. Subsequent work on so-called anaphoric islands revealed outbound anaphora to be a gradient phenomenon, rather than the categorical one originally described by Postal.

Tic Douloureux 1971, for example, observed that certain 'unmentionable' body substances may be felicitously referred to with an anaphor even when those substances are not explicitly evoked in the preceding discourse. Consider the examples in 6, in which no explicit antecedent for the anaphor occurs (Tic Douloureux 1971:46):

(6) a. John *bled* so much *it* soaked through his bandage and stained his shirt. (bleed = 'to emit *blood*')
b. When Little Johnny *threw up*, was there any pencil-eraser in *it*? (throw up = 'to emit *vomit*')

To account for such data, Tic Douloureux proposed the following 'grammatical' principle (1971:48): 'Whenever a sentence has a semantic interpretation making reference to an action or event that (inferentially) results in the production of an unmentionable bodily substance, such a substance can be referred to by a pronoun *it* within the sentence...' Significantly, this principle makes no reference to any morphological or syntactic relation between anaphor and ante-

---

the antecedent for *herself* is the deleted subject of the VP *to justify herself*, given that the verb *attempt* is an EQUI-verb, and that the related noun *attempt* is an EQUI-controlling noun; in current parlance, the subject of *attempt* controls the PRO of the embedded clause. Curiously, however, while *French* can apparently control the PRO in (i), as Postal 1969 noted in connection with similar examples, an explicit anaphor—which should permit coindexing with the subject PRO—is odd in this context, as seen in (ii):

(i) the French attempt PRO to regain the former colonies
(ii) ?the *French* attempt PRO to regain *her* former colonies

cedent. However, as we shall see, the inferential process alluded to in Tic Douloureux's principle extends far beyond unmentionable bodily substances.

Lakoff & Ross (1972) proposed a set of principles designed to account for some of the gradations in acceptability for outbound anaphora. First, they suggested that examples of outbound anaphora are improved if the intended antecedent is morphologically related to the surface word that contains it. Thus, 7b is correctly predicted to be more acceptable than 7a (Lakoff & Ross 1972:121):

(7) a. *The *orphan* misses *them*.
b. ?*A *guitar*ist bought *one* yesterday.

Second, they claimed that an even greater improvement can be achieved if the derived lexical item containing the antecedent does not COMMAND the pronoun.[3] Thus 8a is worse than 8b, they claimed, because in 8a the word containing the antecedent (*guitarist*) commands the pronoun (*it*), while in 8b it does not (1972:121):

(8) a. ?*The *guitar*ist thought that *it* was a beautiful instrument.
b. ?John became a *guitar*ist because he thought that *it* was a beautiful instrument.

On the basis of these observations, Lakoff & Ross proposed the following three degrees of deviance for outbound anaphora:

(9) a. '*' if the lexical item and the antecedent are not morphologically related;
b. '?*' if the lexical item and the antecedent are morphologically related and if the lexical item commands the pronoun;
c. either '?' or 'ok' if the lexical item and the antecedent are morphologically related and if the lexical item does not command the pronoun.

However, it is not the case that morphological unrelatedness necessarily results in infelicitous outbound anaphora. Consider the example in 10, where the containing word *second* is clearly not morphologically related to the intended antecedent *two*:[4]

(10) This is the *second* time in *as many* weeks.

Another problem is that Lakoff & Ross's command condition 9b would assign the second degree of deviance to the naturally-occurring examples in 11:

(11) a. The *Senator Bradley* forum has been canceled due to *his* need to be in Washington for the budget vote.
(note on poster at AT&T Bell Labs; September 26, 1990)
b. Last night's *Sinead O'Connor* concert at the Garden will be *her* last.
(WNBC 6:00 News; August 25, 1990)

[3] Node A commands node B if neither node dominates the other and if node B is dominated by the first S node above A (Ross 1986:201).

[4] As we explain in §3.4, what is required for the felicitous outbound anaphora exemplified in 10 is the existence of a well-instantiated lexical—rather than morphological—relationship between the containing word and the intended antecedent.

c. I was reading this *Peggy Noonan* book on *her* years at the White House...
(Julia Hirschberg in conversation; November 9, 1990)

In all these examples, the lexical item containing the antecedent commands the pronoun, yet none seems particularly infelicitous.

Watt (1975) discussed a number of factors that, he claimed, serve to improve the 'penetrability' of outbound anaphora. First, he noted that such anaphora is facilitated when the antecedent bears contrastive stress, as in 12 (Watt 1975:106):

(12) All the *Nixon*ites I know are for putting all the Agnewites in cold storage till 1976; but *HE HIMSELF* doesn't care a fig.

Here, it is claimed that the contrast between Nixon and Agnew—marked prosodically by a pitch accent on *Nixon*—'exposes' the antecedent in a way the deaccented antecedent would not. Watt argued that exposed antecedents result in reduced processing effort (1975:105):

> 'In the case of an 'impenetrable', exposure of (= penetration to) the contained anaphorical antecedent would thus be possible at the point in hearing the sentence when only the antecedent had been heard, rather than, retrospectively, when the anaphor was heard, perhaps much later. A reduction of processing effort should result, and so a gain of acceptability.'

Thus, for Watt, accent on *Nixonites* in 12 serves to expose the substring *Nixon*, rendering the NP 'available' for subsequent reference. However, as noted by Wilson & Sperber (1979), Prince (1981b, 1986), Rooth (1985), Hirschberg & Pierrehumbert (1986), and Pierrehumbert & Hirschberg (1990), among others, the function of pitch accent is not to expose linguistic strings, but rather to highlight, or focus, the discourse entities to which those strings refer. Such an analysis of accent is consistent with our view of reference as a relation between language and entities in a discourse model, rather than as a relation between linguistic objects. Furthermore, we argue that what is relevant for felicitous outbound anaphora is not accent per se, but rather the relative accessibility of the discourse entity which may be evoked AS A RESULT OF a speaker's use of accent. Nonetheless, we agree with Watt that accent is relevant to the interpretation of outbound anaphora, though it is but one of many factors that contribute to the relative accessibility of discourse entities.

Another factor contributing to felicitous outbound anaphora, according to Watt, is the degree to which the anaphor is 'specific' to the particular antecedent. To illustrate, Watt offered the examples in 13 (1975:102):

(13) a. ??Whenever Otis meets a lifelong *New Yorker* he says he thinks *it*'s the worst city in the world.
b. +Whenever Otis meets a lifelong *New Yorker* he says he wouldn't live *there* on a bet.[5]
c. +Whenever Otis meets a lifelong *New Yorker* he says he would never visit *such a place*.

Here Watt claimed that, as an anaphor becomes increasingly specific (i.e. from the least specific, *it*, to the most specific, *such a place*), the corresponding

[5] Watt used '+' to mean 'the antithesis of "0", however interpreted' (1975:101).

islands become increasingly 'penetrable'. While we disagree with Watt about the infelicity of 13a, we nonetheless agree that in general the more descriptive the anaphor, the greater the possibility of successful reference.[6]

Watt's set of conditions under which penetration into islands is more or less possible constituted the first attempt of which we are aware to describe what would now be called pragmatic factors that affect the well-formedness of outbound anaphora. However, Watt adopted the contemporary prevailing view of anaphora as essentially a relation between linguistic elements: 'The bond joining anaphor and antecedent is sensitive to whether or not both anaphor and antecedent are present in the given sentence as 'words', but this sensitivity is very mutable' (1975:101). This contrasts with the more modern (and more accurate) view of anaphora as a relation between a linguistic anaphor and its nonlinguistic referent in the discourse model.

Corum (1973) presented additional evidence in support of a gradient, rather than categorical, constraint on outbound anaphora. She argued that, in some cases, pronouns MUST be allowed to refer to an antecedent that is contained in the semantic structure of another word. She further suggested that the gradient nature of the constraint—i.e. that anaphors can refer AT ALL to items within words—is evidence for a Generative Semantic as opposed to an Interpretive approach. Browne 1974, however, argued that Corum's idea of (semantic) containment must be weakened to 'semantically related', because an anaphor's antecedent can either contain or be contained in the surface form. As evidence, Browne provided the examples in 14 (1974:620):

(14) a. Mary knows *Kurd*ish, because she is *one*.
b. John is a *Kurd*, and his children can speak *it*.

In 14a the antecedent of *one* (*Kurd*) is semantically and morphologically contained within the word *Kurdish*, while in 14b the intended antecedent of *it* (*Kurdish*) actually contains the surface word *Kurd*. In fact, all of Browne's examples involve surface words which are both morphologically AND semantically related to the intended antecedent (cf. Lakoff & Ross's 1972 formulation concerning MORPHOLOGICAL relationship).

We note in passing that, assuming the examples in 14 are well-formed, Browne's argument has an undesirable consequence for the Generative Semantics position. If *Kurdish* is represented as 'the language spoken by Kurds' in 14a, and if *Kurd* is represented as 'people who speak Kurdish', as 14b would seem to suggest, then a representational infinite regress results.

2.3. OUTBOUND ANAPHORA AND RECENT THEORIES OF MORPHOLOGY. While the outbound-anaphora data were originally offered as evidence for Generative Semantics, such data have also been cited in support of a number of claims about morphology. For example, Levi (1978) argued that the data supported her position that complex nominals (e.g. compound nouns) are categorially

[6] A better example to illustrate Watt's point in 13a is presented in (i):

(i) Whenever Otis meets a lifelong *New Yorker* he says he thinks *it*'s dirty.

Without the predicate in Watt's example (*the worst city the world*), the *it* of (i) is difficult to interpret.

nouns rather than noun phrases. More recently, anaphoric-island data have been reinterpreted in the context of the theory of Lexical Phonology and Morphology. An important principle of lexicalist theories of morphology (e.g. Pesetsky 1979, Kiparsky 1982, and Mohanan 1986, inter alia) is the LEXICAL INTEGRITY HYPOTHESIS. Under this hypothesis, syntactic processes do not have access to the internal structure of words. Movement transformations, for instance, are prevented from moving morphemes either into or out of words. According to Pesetsky 1979 (and subsequent work, e.g. Mohanan 1986), such lexical 'integrity' is derivable from an important construct of Lexical Phonology and Morphology, namely BRACKETING ERASURE. Bracketing erasure deletes word-internal brackets at certain points in the derivation of a word (at the end of each cycle, in most versions of the theory). Crucially, word-internal brackets are also deleted at the end of a word's derivation, prior to lexical insertion. Bracketing erasure thus prohibits postlexical (e.g. syntactic) processes from having access to word-internal components; no syntactic process, for example, may make reference to the morpheme *truck* in the compound *truck driver*. Hence, such a compound would be as unanalyzable as *orphan* with respect to syntactic operations.

Under the assumption that anaphora involves a syntactic relationship between word strings, Simpson 1983 noted that the existence of anaphoric islands follows from the lexical integrity hypothesis. Because word-internal components are not visible to syntactic operations, there would be no way for an anaphor to be coindexed with a word-internal antecedent. Outbound anaphora is thus predicted to be categorically ungrammatical.[7] However, as we have seen, outbound anaphora is not, contra Simpson, a categorical phenomenon. Furthermore, while Simpson's approach makes a strong (but untenable) prediction concerning cases of sentence-internal anaphora, it is unclear what prediction it would make in a case where the anaphor is in a different sentence from its (word-internal) antecedent. Compare, for example, 15a–b:

(15) a. #Yesterday, I met this really odd *truck* driver who lives in *it*.
b. Yesterday, I met this really odd *truck* driver. #He lives in *it*.

Assuming that intersentential coreference is not governed by syntactic coindexation, Simpson's theory rules out 15a, while making no claim about the equally infelicitous 15b.

Sproat (1985, 1988) argued that Postal's prohibition against both inbound and outbound anaphora is derivable without appealing to the notion of lexical integrity. Instead, he suggested that the constraint could be derived from considerations concerning the kinds of antecedents that anaphors may have. He argued that previous work on anaphora within generative syntax has implicitly assumed that an antecedent for a pronoun must be a maximal projection. So it has been assumed, for example, that *him* in 16 cannot be coindexed with

[7] Note that this is similar to Postal's 1969 notion that the anaphoric-island constraint applies late in the derivation: in both cases, a principle applies that renders morphologically complex words indistinguishable from monomorphemic words with respect to postlexical processes (including anaphora).

just the head noun *man*, but only with the maximal projection of the head noun, i.e. the NP *the large man* (Sproat 1988:294):

(16) *The large *man* had a hat with *him*.

Sproat proposed that anaphors such as pronouns or the pro-VP *do so*—both maximal projections under his analysis—must have as antecedents phrases that are likewise maximal projections. Thus, he argued, one can derive structural constraints on outbound anaphora by appealing to the prohibition on maximal projections within words in English, as evidenced by the ungrammaticality of **a [The Bronx] hater*, where a maximal projection (*The Bronx*) occurs word-internally (Fabb 1984). Under such an analysis, *truck* in *truck driver* could not serve as the antecedent for a pronoun simply because it is not of the right syntactic form. In this way, both Sproat (1985, 1988) and Simpson (1983) argued that no anaphoric-island constraint per se is necessary, with Sproat pointing out that so-called anaphoric islands do not, contra Simpson, provide evidence for the lexical integrity hypothesis. However, both Sproat's and Simpson's approaches, like Postal's original analysis, treated anaphoric islands as a categorical phenomenon, which, as we have seen, is not supported by the data.

Like Lakoff & Ross 1972, Lieber 1984 suggested that structural configuration plays a significant role in the acceptability of outbound anaphora. Appealing to Government-Binding theory (Chomsky 1981), Lieber claimed that R-expressions (i.e. nonpronominal referring expressions) may not be bound, and hence that pronouns may not c-command their R-expression antecedents.[8] This constraint, she claimed, could account for the contrast illustrated in 17 (1984:188):

(17) a. *McCarthy*ites are now puzzled by *him*.
b. **He* distrusts *McCarthy*ites.

Specifically, Lieber attributed the unacceptability of 17b—where *he* c-commands the R-expression *McCarthy*—to a violation of Condition C of the binding theory, which states that an R-expression may not be bound. By appealing to binding theory, Lieber attempted not only to account for the ill-formedness of 17b, but also to argue against the lexical integrity hypothesis; since, she claimed, the syntactic principles of binding theory must have access to word-internal elements in order to rule out 17b, it follows that the lexical integrity hypothesis cannot be correct.

However, the problem with Lieber's example 17b is not that *McCarthy* is c-commanded by the subject pronoun; rather, its deviance results from the fact that there is no antecedent for the anaphor in the context provided. We would not expect *he* to specify McCarthy in this example any more than we would expect *he* to specify McCarthy in, say, *he left*. In an appropriate context, Lieber's example—slightly modified—is fine. Consider the constructed example in 18a, as well as the naturally-occurring example in 18b, from a report of an interview with Salman Rushdie:

[8] There are various definitions of c-command. For Lieber's—and our—purposes the following definition (taken from Radford 1988:115) will suffice: X c-commands Y iff the first branching node dominating X dominates Y, and neither X nor Y dominates the other.

(18) a. After *McCarthy* had undergone a change of heart and issued a public apology, *he* began to distrust the very McCarthyites who previously had been so fiercely loyal.

b. *He* has called editors to tell them *Rushdie* jokes ... (*New York Times Magazine*, 'Rushdie in hiding'; November 3, 1990, p. 68)

The felicity of these examples argues against any attempt to provide an exclusively structural account of outbound anaphora.

Finally, Sproat & Ward (1987) challenged the claim that the unacceptability of so-called anaphoric islands involving outbound anaphora is the result of a violation of some syntactic or morphological principle.[9] They argued that pragmatic factors such as contrast and topicality serve to increase the salience of a referent evoked by a word-internal element to a level where outbound anaphora is felicitous. In this paper we develop some of the suggestions introduced in this earlier work, and present the results of a series of psycholinguistic experiments that support these suggestions.

2.4. SUMMARY. Anaphoric-island data were first offered in support of the decompositional approach of Generative Semantics. Although Postal's original 1969 formulation of the anaphoric-island condition included a categorical prohibition on reference 'into and out of' words, it was soon noted (Lakoff & Ross 1972, Watt 1975) that the conditions on well-formed outbound anaphora were in fact gradient. The phenomenon was subsequently recast in terms of lexical integrity, a key principle of lexicalist morphological theory. The earlier anaphoric-island stipulation was argued to be derivable from a more general prohibition against syntactic access to lexical structure (implemented as bracketing erasure in Lexical Phonology and Morphology). Sproat (1985, 1988) argued against this approach and suggested instead that there was a syntactic condition on the kinds of phrases which could serve as possible antecedents for anaphors.

With few exceptions, previous approaches have assumed that outbound anaphora is to be ruled out by some morphological or syntactic principle. In what follows we shall suggest, as in the studies of Simpson 1983 and Sproat 1985, 1988, that there is no specific anaphoric-island restriction. However, unlike Simpson or Sproat, we shall argue that the degree to which outbound anaphora is felicitous is determined by the relative accessibility of the discourse entities evoked by word-internal lexical elements, and not by any principle of syntax or morphology.[10] While some previous studies (e.g. Tic Douloureux 1971, Watt 1975) have acknowledged the importance of pragmatic factors in the acceptability of outbound anaphora, most others have taken the alternative position that outbound anaphora is ungrammatical, and only occasionally ameliorated through contextual manipulations. In the following section we reject this 'ungrammatical-but-salvageable' view of outbound anaphora, and present our pragmatic analysis of the phenomenon.

[9] The sole exception is outbound anaphora with the pro-VP *do so*, on which see §3.3 below.

[10] Nor by any principle derivable from other morphological or syntactic principles, such as lexical integrity.

3. As noted in §1, we shall assume that a genuinely ungrammatical construction is ungrammatical in all (nonmetalinguistic) contexts, and cannot be 'amnestied' by pragmatic or discourse factors.[11] Given such an assumption, it would be inconsistent for a construction to be ruled out by syntactic considerations and, at the same time, be acceptable under certain discourse conditions. Rather, we would maintain that such a construction is syntactically well-formed, but restricted to certain discourse contexts for pragmatic reasons.

In our study of inbound and outbound anaphora, we will concentrate on cases where the antecedent (in the case of outbound anaphora) or anaphor (in the case of inbound anaphora) is morphologically 'contained' within a word.[12] Specifically, we propose that:

(19) A. Inbound anaphora is ruled out by a grammatical principle that prohibits pronominal elements from appearing in word-internal positions.

B. Outbound anaphora is not ruled out by any grammatical principle—with the exception of outbound anaphora involving *do so* (see §3.3).

First, we claim that inbound anaphora is ungrammatical: word-internal anaphors are categorically ruled out by independently motivated morphosyntactic principles. There are a number of ways in which this prohibition could be derived, but for the purposes of this discussion we present the simplest of these (see Sproat 1985, 1988 for a different explanation). Pronouns are closed-class items, and as such do not freely allow further morphological derivation (Paul Kiparsky, personal communication, 1990). Thus forms like **himite* or **themhater* are ruled out by the same morphological constraint that generally prevents formations like **withing* or **overer*.

Given our assumption that ungrammatical constructions cannot be amnestied by pragmatic factors, it follows that inbound anaphora should not be possible in ANY (nonmetalinguistic) discourse context. Indeed, we know of no contexts in which such anaphora is well-formed. We thus conclude that inbound and outbound anaphora are, contra Postal 1969, distinct in that only the former is governed by morphosyntactic principles. Crucially, however, inbound anaphora is not ruled out because words are anaphoric 'islands', but rather because pronouns are categorically barred from word-internal positions.[13]

Second, we claim that there is no principle of grammar that explicitly pre-

[11] For a contrasting view, see Shibatani & Kageyama (1988), who argue for an Anaphoric Island Constraint, while conceding that violations may occur as a result of 'some kind of pragmatic inference rather than by a direct coreferential relation' (1988:473, n. 12). However, they provide no criteria to distinguish between these two possibilities. As we will argue in the following discussion, such a distinction is both unmotivated and unnecessary.

[12] Examples where no morphological containment is involved, e.g. 2a, are discussed in §3.4 below.

[13] One might also point out that some languages do allow incorporated pronouns within verbs (see, for instance, Bresnan & Mchombo 1987). As far as their anaphoric behavior is concerned, incorporated pronouns in languages that have them are exactly like nonincorporated pronouns in

vents word-internal antecedents for pronominal anaphors.[14] As initial evidence, consider the naturally-occurring data in 20, drawn from our corpus of outbound anaphora (part of which is presented in the Appendix).

(20) a. For a *SYNTAX* slot, I'd rather see someone with more extensive coursework in *it*. (Judith Levi discussing various subdisciplines of linguistics; January 18, 1987)

b. Patty is a definite *Kal Kan* cat. Every day she waits for *it*. (Television advertisement for Kal Kan; January 28, 1987)

c. There's a *Thurber* story about *his* maid ... (Michael Riley in conversation; September 7, 1988)

d. [illegible]nt up to *Constable* country; we stayed in the village *he* was [illegible]n in. (Kenneth Sproat in conversation; October 11, 1988)

e. I refer you to the *Schachter* paper; *he*'s very proud of it ... (Mark Baker in response to a question at NELS; November 12, 1988)

f. Well, action is still needed. If we're to finish the job, Reagan's Regiments will have to become the *Bush* Brigades. Soon *he*'ll be the chief, and he'll need you every bit as much as I did. (Ronald Reagan, farewell speech, January 11, 1989, reported in Associated Press Newswire)

g. Millions of *Oprah Winfrey* fans were thoroughly confused last week when, during *her* show, she emotionally denied and denounced a vile rumor about herself. (*Chicago Tribune*, column by Mike Royko; May 22, 1989; cited in James McCawley's '1989 linguistic flea circus' as an example of reflexive usage—not as an example of outbound anaphora)

h. I had a *paper* route once but my boss said I took too long deliverin' *'em*. ('L. A. Law'; 1987)

i. I'm a *mystery-story* buff and read (and watch on PBS) a lot of *them*. (Northwestern University electronic bulletin board; January, 1989)

---

a language like English. Again, this does not affect our argument here: it seems that English MORPHOLOGICALLY rules out any kind of pronoun 'incorporation', and it is this grammatical fact which accounts for the inbound anaphora data. If English did allow incorporated pronouns, we would expect them to behave like free pronouns with respect to their anaphoric behavior, just as they do in languages that allow them.

[14] Following previous work on anaphoric islands, we shall restrict our analysis of outbound anaphora to nonepithet anaphors. However, we note that anaphoric epithet NPs, illustrated in (i) and (ii), also participate in such anaphora:

(i) The *Philadelphia* Inquirer beseeched its readers through a series of editorials last summer to stop giving to beggars, especially drug and alcohol abusers, who the paper claimed were driving away tourists and threatening the economic survival of *the city*'s downtown. (*Chicago Tribune* article, 'Beggar's bounty: Deaf ear, cold shoulder'; May 13, 1990)

(ii) Health Secretary Louis Sullivan said Monday he was outraged that 'unAmerican' protesters prevented him from being heard at an *AIDS* conference, but the incident would not reduce his commitment to fight *the disease*. (*Chicago Tribune* article, 'AIDS protest angers health secretary'; June 26, 1990)

j. In the distance, we heard the sound of an *ambulance* siren. Within a minute or so *it* arrived and stretcher bearers took the boy away. (*New York Times Magazine*, 'The tragedy of Detroit'; July 29, 1990, p. 25)

k. Officials in the Danish capital believe they've found a way to stop *bicycle* thefts—let people use *them* for free. (Associated Press Newswire; November 10, 1990)

l. I was reading this *Peggy Noonan* book on *her* years at the White House ... (= 11c)

If one takes the position that outbound anaphora violates a principle of grammar, one will have to allow for frequent pragmatic amnestying in order to accommodate the well-formedness of data such as those in 20. In the absence of any account of the conditions under which such amnestying is possible, it is not clear how to evaluate this position. Moreover, such an account would also have to explain why cases of truly ungrammatical inbound anaphora fail to be rendered acceptable under ANY circumstances. For example, if one were to argue that 20a can be amnestied because the anaphor is interpretable by some kind of 'pragmatic inference', one would have to explain why the same sort of pragmatic inference fails to salvage the following example, where there is clearly no difficulty in interpreting the anaphor:[15]

(21) *I'll eat *oysters* on occasion, but I'm really not much of a *them* lover.

On the basis of such data, we reject the view that outbound anaphora is ungrammatical and argue instead for a pragmatic analysis of the phenomenon. From this, it follows that the many examples of ill-formed outbound anaphora discussed by Postal (1969) and others are not syntactically UNGRAMMATICAL, but rather pragmatically INFELICITOUS.

Before proceeding, we first lay out some assumptions concerning the pragmatic framework that we will be adopting. As we have noted, one of the problems with previous accounts of outbound anaphora has been the assumption that anaphora—indeed, reference in general—involves a direct relation between LINGUISTIC objects. As discussed above, Postal's original formulation of the problem in terms of anaphoric islands involved morphosyntactic restrictions on possible antecedents for anaphoric elements: 'Outbound anaphora is the relation between a [sentence] chunk, part of which is interpreted as antecedent, and some anaphor outside of that chunk' (1969:206). Watt 1975 furthermore talks of 'penetrating' a word or phrase in order to arrive at a pronoun's antecedent.

In contrast, we maintain that a more adequate account of outbound anaphora

[15] One might argue that, on the one hand, constructions like **them lover* violate a strong morphosyntactic constraint, whereas instances of outbound anaphora violate only weak morphosyntactic constraints and are therefore more readily amnestied by pragmatic factors. While this is a possible theory, it is not clear how one would distinguish it empirically from the pragmatic approach we present below. Furthermore, we would argue that the pragmatic factors affecting the acceptability of outbound anaphora are factors that are relevant to anaphora in general; thus, the idea that outbound anaphora is even weakly ungrammatical serves no apparent purpose.

is possible once reference is viewed as a relation that holds between language and one or more entities in a constructed representation, or model, of the ongoing discourse (see Karttunen 1976, Grosz 1977, Webber 1979, and Sidner 1979, inter alia). Under this view, pronouns and other anaphors are used to refer to discourse entities rather than to linguistic antecedents. The felicity of a particular instance of anaphora, then, is a function of the relative accessibility of the discourse entity to which the anaphor is intended to refer, as well as the type of anaphor used to refer (Watt 1975). As is well known, pronouns are the most pragmatically constrained type of anaphor in that their felicitous use requires that the hearer has (or could appropriately come to have) the referent of the pronoun 'in consciousness' at the time of the hearing or processing of the utterance (see Chafe 1976, Sidner 1979, Prince 1981a, Gundel & Hedberg 1990, inter alia). That is, felicitous use of a pronominal referring expression requires that the entity to which the pronoun is being used to refer is accessible for the hearer at the time of the utterance.

We intend to demonstrate that outbound anaphora is sensitive to the same types of pragmatic constraints as are other types of pronominal reference. Specifically, we claim that word-internal morphemes may felicitously serve as antecedents for subsequent anaphora just in case the discourse entity evoked by the antecedent in question is sufficiently accessible at the time of the utterance. In those cases where the discourse entity evoked by the word-internal antecedent is not sufficiently accessible, we predict that outbound anaphora will be infelicitous.[16]

In §3.1 we discuss some of the morphosyntactic and semantic factors that affect the accessibility of discourse entities, and thus the felicity of outbound anaphora. We show that the infelicity of at least some types of outbound anaphora is derivable from various semantic and syntactic properties of words, given certain assumptions about the effects those properties have upon discourse entities introduced by word-internal morphemes. In §3.2 we consider some of the pragmatic factors that affect the felicity of outbound anaphora, and in §3.3 we argue that the VP anaphor *do so*, unlike other anaphors, is governed by morphosyntactic principles and does not participate in outbound anaphora. In

[16] An examination of our corpus of naturally-occurring data reveals that antecedents in word-internal positions evoke discourse entities of one of three types: a kind (in the sense of Carlson 1977), a mass term, or a specific set of one or more individuals. By far the largest class of examples in the corpus involves reference to particular individuals that are evoked by proper-name antecedents. Curiously, DiSciullo & Williams (1987:50–51) claim that words are 'referential islands' for proper names and that proper names within words are not 'truly referential'. From this, they claim, it follows that (for example) the property of admiring Nixon is not an essential property of a Nixon admirer. Thus, they argue that (i), unlike (ii), is not a contradiction (we include DiSciullo & Williams' judgments, 1987:51):

(i) John is a Nixon admirer in every sense except that he does not admire Nixon.
(ii) *John admires Nixon in every sense except that he does not admire Nixon.

If one can construe a Nixon admirer as being a person with a reliable set of traits (e.g. is clean-shaven, always wears three-piece suits, and carries an attaché case), then (i) might not be construed as a contradiction. But whether or not *Nixon* in *Nixon admirer* can be used referentially is beside the point.

all three of these sections we present psycholinguistic evidence in support of our claims. Finally, in §3.4 we discuss cases of outbound anaphora whose antecedents are not morphologically present.

3.1. MORPHOSYNTACTIC AND SEMANTIC FACTORS THAT AFFECT THE FELICITY OF OUTBOUND ANAPHORA. A key factor in determining the felicity of outbound anaphora is the semantic transparency of the word containing the antecedent of the anaphor (cf. Lieber 1984). The containing word must be sufficiently transparent for the word-internal morpheme to successfully evoke an accessible discourse entity. Consider the following examples:

(22) a. Although casual *cocaine* use is down, the number of people using *it* routinely has increased. (WCBS 11 O'clock News; December 20, 1990)

b. Patty is a definite *Kal Kan* cat. Every day she waits for *it*. (= 20b)

In 22a, *cocaine use* is a semantically transparent synthetic compound: the right-hand member is a deverbal nominal and the lefthand member is readily interpretable as the internal argument of the verb *use*. Thus, *cocaine use* means 'use of cocaine'. To arrive at this interpretation, a hearer must access the meanings of both *cocaine* and *use*, and it is in part this decomposition process, we claim, that renders the discourse entity *cocaine* accessible in the context of 22a. To understand the compound *Kal Kan cat* in 22b, the hearer must figure out the intended relation between cats and the substance Kal Kan. In the course of determining this relation, the hearer must access the referent of the brand name *Kal Kan* along with the denotation of the common noun *cat*. Again, such semantic decomposition serves to render accessible the relevant discourse entity.

However, it is well known that morphologically complex words tend to acquire idiosyncratic, institutionalized meanings over the course of time (Aronoff 1976, Bauer 1983). As a result, some morphologically complex words have become semantically opaque in that they can no longer be straightforwardly interpreted on the basis of their component parts. As the following examples illustrate, semantic opacity generally inhibits outbound anaphora.

(23) a. Fritz is a *cow*boy. #He says *they* can be difficult to look after.

b. Roberta is an ordained *Luther*an minister. #She's currently studying the early years of *his* life.

c. #Ironically, Paula had a *Caesar*ean while writing a book on *his* rise to power in early Rome.

d. Dom's clothes are absolutely *elephant*ine. #Indeed you could almost lose *one* in them.

Consider first the compound *cowboy* in 23a, a word that has become institutionalized. Because of institutionalization a hearer may access the meaning of the compound directly, i.e. without morphologically decomposing it. Thus *cow*, despite its morphological presence, would not generally evoke an accessible discourse entity when *cowboy* is uttered. The examples of derivational affixation in 23b–d illustrate the same point: elements within semantically opaque or institutionalized constructions do not evoke accessible discourse entities,

and thus do not generally permit felicitous outbound anaphora. In 23b, for instance, *Lutheran* is clearly related to *Luther* morphologically, yet it is to some extent only accidental that the former means 'the branch of Protestantism adhering to the views of Martin Luther'. Of course, the distinction between transparent words and opaque or institutionalized words is gradient rather than categorical. We would therefore expect word-internal morphemes to evoke discourse entities with a greater or lesser degree of accessibility depending, inter alia, upon the relative transparency of the containing word.

While semantically transparent compounds do allow felicitous outbound anaphora, it is also true that anaphora involving antecedents within compounds is, other things being equal, more difficult to construe than anaphora involving non-word-internal antecedents. One explanation for this difference may lie in the semantic difference between modifiers and predicates. First, we assume that compounds are modifier-head constructions (see, for instance, Levi 1978). That is, in the compound *Kal Kan cat*, *Kal Kan* can be said to modify *cat* in much the same way as the adjective *hostile* modifies *aunt* in the adjective-noun sequence *hostile aunt*. Let us further assume, following Wilson & Sperber 1979, that adjectives functioning as modifiers (in prenominal position, for example) are more backgrounded, i.e. less salient, than adjectives functioning as predicates. Given these assumptions, we can account for the infelicity of many instances of outbound anaphora involving compounding with the following hypothesis: discourse entities evoked by modifiers are, ceteris paribus, less accessible than entities evoked by predicates.

In fact, this hypothesized difference between modifiers and predicates has some empirical support. In an experiment reported fully in McKoon et al. 1990, it is shown that adjectives functioning as modifiers are generally less salient than the same adjectives functioning as predicates. Consider the sentences in 24, from McKoon et al. 1990:

(24) John doesn't like to visit his relatives very much.

a. His intolerable aunt is hostile.

b. His hostile aunt is intolerable.

He never has a very good time.

McKoon et al. (see also Rothkopf et al. 1988) found that adjectives were more available when presented in a later memory test if they had appeared in the text as predicates (e.g. *hostile* in 24a) than if they had appeared as (prenominal) modifiers (e.g. *hostile* in 24b). This finding suggests that, other things being equal, modifiers are generally less salient than predicates. In this way, we can account for the relative infelicity of outbound anaphora involving anaphors whose antecedents are functioning as compound-internal modifiers.

3.2. PRAGMATIC FACTORS THAT AFFECT THE FELICITY OF OUTBOUND ANAPHORA. In this section we discuss some pragmatic factors that affect the accessibility of discourse entities, and hence affect the felicity of outbound anaphora. We also review a series of psycholinguistic studies that provide empirical support for our analysis.

The accessibility of discourse entities is sensitive to a number of pragmatic

factors. In particular, a discourse entity seems to be more accessible (and subsequent outbound anaphora more felicitous) when the entity stands in salient opposition to some other discourse entity (see Watt 1975). Examples of such contrast are provided in 25:

(25) a. Well, action is still needed. If we're to finish the job, Reagan's Regiments will have to become the *BUSH* Brigades. Soon *he*'ll be the chief, and he'll need you every bit as much as I did. (= 20f)

b. For a *SYNTAX* slot I'd rather see someone with more extensive coursework in *it*. (= 20a)

c. Cliff Barnes: Well, to what do I owe this pleasure?
Ms. Cryder: Actually, this is a *BUSINESS* call, and I'd like to get ri$_{\text{e}}$ht down to *it*. ('Dallas'; 1987)

In 25a then-President Reagan is contrasting his regiments with soon-to-be inaugurated President Bush's brigades. As a result of this contrast, we claim, the discourse entity corresponding to *Bush*, being in salient opposition to the discourse entity evoked by *Reagan*, is rendered more accessible. Similarly, in 25b the speaker is contrasting syntax with other subdisciplines of linguistics, and in 25c the second interlocutor contrasts business with pleasure. As is the case with contrast in general, contrast in these examples is realized intonationally with a pitch accent on the word or morpheme that evokes the discourse entity being contrasted (cf. Watt's 1975 claim—discussed in §2.2—that accent can 'expose' a word-internal antecedent).

Related to the notion of contrast is the notion of discourse topic (Chafe 1976 and Reinhart 1981, inter alia). We have observed that topical discourse entities evoked by word-internal elements facilitate outbound anaphora more than nontopical discourse entities do. Consider the following token, from a story about violence in Detroit:

(26) In the distance, we heard the sound of an *ambulance* siren. Within a minute or so *it* arrived and stretcher bearers took the boy away. (= 20j)

Here the pronoun *it* can felicitously be used to refer to a specific ambulance, which was evoked by a word-internal morpheme in the previous sentence. One of the topics of the magazine article in question was the dramatic increase of crime-related injuries in Detroit. We maintain that, in this context, ambulances are relatively topical, and this topicality renders the example in 26 felicitous.

To investigate the effects of contrast and topicality on outbound anaphora, a series of psycholinguistic experiments was recently conducted (McKoon et al. 1990). It was hypothesized that these pragmatic factors would serve to increase the accessibility of discourse entities evoked by word-internal elements, and thus facilitate outbound anaphora. Below we present an overview of the experiments, beginning with a discussion of how accessibility was manipulated and how ease of comprehension was measured.

Accessibility was manipulated in two ways: syntactically, by varying morphosyntactic structure, and pragmatically, by varying topicality and contrast.

In the first experiment, a set of 24 texts was used, each with four versions; an example is provided in Table 1. The last sentence of each version of each text contained a pronominal anaphor. In two of the four versions, the antecedent of this anaphor appeared in a nominal compound in the penultimate sentence, and in the other two versions the antecedent appeared in a verb phrase. It was hypothesized that discourse entities evoked by compound-internal antecedents would be less accessible than entities evoked by antecedents not contained in compounds, and that this difference could be attributed to the fact that the antecedent in the NP versions appeared as a modifier within the compound (see above, §3.1). Therefore, it was predicted that comprehension of the anaphor in the final sentence would be facilitated in the VP versions relative to the NP versions. In Table 1, for example, comprehension of the pronoun *they* in the final sentence was predicted to be facilitated when its antecedent *deer* appeared as a verbal argument (*hunting deer*) relative to when it appeared as a compound-internal modifier (*deer hunting*).

Compound/Non-Topical

Sam has many interests in the outdoors. He's an avid skier, and each winter he takes about a month off from work to ski in Colorado. In the summertime, he visits his parents in Montana, where he has a chance to do some mountain climbing. Lately, he's taken up deer hunting.

And he thinks that they are really exciting to track.

Compound/Topical

Sam likes the outdoor life. Having grown up in rural Kentucky, he knows a lot about nature and is an expert at fishing and shooting. He goes on hunting trips as often as he can. He used to hunt just small game, like rabbit and quail. However, lately he's taken up deer hunting.

And he thinks that they are really exciting to track.

Verbal complement/Non-Topical

Sam has many interests in the outdoors. He's an avid skier, and each winter he takes about a month off from work to ski in Colorado. In the summertime, he visits his parents in Montana, where he has a chance to do some mountain climbing. Lately, he's taken up hunting deer.

And he thinks that they are really exciting to track.

Verbal complement/Topical

Sam likes the outdoor life. Having grown up in rural Kentucky, he knows a lot about nature and is an expert at fishing and shooting. He goes on hunting trips as often as he can. He used to hunt just small game, like rabbit and quail. However, lately he's taken up hunting deer.

And he thinks that they are really exciting to track.

TABLE 1. Examples of texts with pronominal anaphors.

In addition to varying morphosyntactic structure, McKoon et al. also varied the accessibility of the referent of the antecedent in the final sentence by ma-

nipulating the contrast between the referent and other discourse entities, as well as the relation between the referent and the overall topic of the text. The texts in which the referent of the intended antecedent was designed to be topical and/or contrastive were labeled 'topical' versions. In the topical versions of the texts in Table 1, for example, the discourse is largely about fishing and hunting, and includes mention of particular animals that have been hunted; in this context, deer are relatively topical. In the nontopical versions, the discourse is about the outdoors in general with no mention of animals, and thus deer in particular are less topical. Under our view of discourse comprehension, we predicted that the topical versions would render the referent more accessible than the nontopical versions, and that this increased accessibility would facilitate comprehension of the pronoun in the final sentence.

Measuring the difficulty of comprehension for the pronoun requires a model of the comprehension processes involved (see, for instance, van Dijk & Kintsch 1983 and McKoon & Ratcliff 1989). For the purposes of this discussion, we describe only the most minimal model, sufficient to allow interpretation of our experimental results (cf. Greene et al. 1990 and Ratcliff & McKoon 1988). The first assumption of the model is that comprehension of a pronoun begins with a process that matches the grammatical features of the pronoun (i.e., in English, gender, number, and person) against the corresponding features of all the entities that have been recently evoked in the discourse model. Discourse entities will vary in the degree to which they match the features of a pronoun, depending upon the accessibility of the entities in question as well as the extent to which the semantic features of the entities correspond to the features of the anaphor. This matching process can have one of several results. If the discourse is not well constructed, there may be no entity that matches to a sufficient degree for the pronoun to be interpreted as referring to that entity. In this situation, other kinds of processing might be initiated, perhaps involving a conscious (as opposed to an automatic) search for the referent, or else the attempt at comprehension could be abandoned altogether, leaving the pronoun without an interpretation. Another possible result of the matching process would be for several candidate entities to match to a high degree, requiring additional contextual information or further processing to decide among them. Finally, if one entity matches the pronoun better than all others, this entity can be interpreted as the intended referent, with the information about the referent being combined with information about the pronoun. All other things being equal, more accessible discourse entities will be matched to a greater degree and more quickly than less accessible ones.

This model can be applied in a straightforward way to the pronouns in the final sentences of the texts used in the experiments. We assume that the grammatical features of the pronoun in a final sentence are matched against (the features of) all of the entities in the text. The most recently evoked entities will all match to some degree; however, the texts in the experiment were constructed in such a way as to rule out, by means of feature mismatches or semantic implausibility, all referents except the intended one. It is the accessibility of this referent that will presumably determine the speed and outcome

of the matching process. The more accessible the referent, the more likely it is that there will be a successful interpretation of the pronoun, and the more quickly this outcome can be achieved.

Given such a model, the experiments reported in McKoon et al. 1990 were designed to measure whether the pronouns in the final sentences were understood as referring to the intended discourse entity, and, if they were so understood, whether the speed of understanding was affected by the relative accessibility of that referent. The texts in the experiments were presented to subjects on a CRT screen. A subject initiated each text by pressing the space bar on the keyboard. This caused the first line of the text to be displayed. When the subject finished reading this line, another press of the space bar brought up the next line of the text, and so on until the final line of the text appeared. When the subject pressed the space bar after the final line of the text, a single test word was displayed on the screen. Subjects were instructed to respond 'yes' or 'no' (by pressing keys on the keyboard) according to whether the test word had or had not appeared in the text that had just been presented. For the 24 texts exemplified in Table 1, the test word was always the (intended) antecedent of the pronoun in the final sentence (e.g. *deer*), and the correct response to this test word was 'yes'. Test words for which the correct response was 'no' were presented after the final lines of filler texts.

This procedure provided two measures, as shown in Table 2. The first measure is the reading time for the final sentence containing the pronoun, and the second is the response time for the test word. The response times for the test words can be used to decide whether the pronouns were equally well understood across the four conditions. Assuming that the successful interpretation of a pronoun leaves its referent highly accessible, decisions on the test word (which corresponds to the referent) should be relatively fast and accurate. So, if the pronouns are equally well understood in all conditions, then response times to the test word should be equally fast and accurate in all conditions, exactly as shown in the results in Table 2: there are no significant differences among the response times, and accuracy rates are all above 95%. Given equal comprehension of pronouns across conditions, any differences in reading times

TEXT VERSION	READING TIMES	RESPONSE TIMES
COMPOUND/NONTOPICAL:		
... Lately, he's taken up deer hunting.	2117ms	907ms
And he thinks that they are really exciting to track.		
COMPOUND/TOPICAL:		
... However, lately he's taken up deer hunting.	1785ms	870ms
And he thinks that they are really exciting to track.		
VERBAL COMPLEMENT/NONTOPICAL:		
... Lately, he's taken up hunting deer.	1868ms	893ms
And he thinks that they are really exciting to track.		
VERBAL COMPLEMENT/TOPICAL:		
... However, lately he's taken up hunting deer.	1738ms	886ms
And he thinks that they are really exciting to track.		

TABLE 2. Results for texts with pronominal anaphors.

for the final sentences can therefore be attributed to differences in difficulty of comprehension. McKoon et al. predicted (a) that comprehension would be relatively more difficult for the nontopical versions than for the topical versions, and (b) that comprehension would be relatively more difficult for the compound versions than for the VP versions. The data confirmed these predictions. For antecedents in both compound and noncompound structures, reading times were significantly slower with the nontopical versions, showing a clear pragmatic effect of topicality and contrast on both outbound and nonoutbound anaphora. Also, for the nontopical versions, reading times were significantly slower when the antecedents had appeared in nominal compounds than in verbal complements. However, for the topical versions, there was no significant effect of morphosyntactic structure on reading times. (Both the main effect of topicality and the main effect of morphosyntactic structure, as well as the interaction of the two, were significant by analyses of variance.) Apparently, for these versions, the accessibility of the referent was already sufficiently high that it could not be significantly increased by having the antecedent in a verbal complement.

These results support our pragmatic account of outbound anaphora in three ways. First, there is a significant effect of whether the intended antecedent is word-internal or not: in the absence of topicality, reading times were slower for the compound versions than for the VP versions. This observation is consistent with the results of the experiments described in §3.1, which showed that adjectival modifiers are generally less accessible than predicate adjectives. Given that compounds are also instances of modifier-head constructions, we are in a position to provide a unified account of both sets of data. All other things being equal, modifiers—of any grammatical category—are less accessible than predicates and complements. Second, the topical versions facilitated comprehension of the anaphor; indeed, in the topical versions there was no significant difference in comprehension between the compound version and the VP version, suggesting that topicality and contrast might in effect make accessibility high enough to be impervious to the effects of morphosyntactic structure. Third, both syntactic versions were affected by manipulations of topicality and contrast, suggesting that outbound anaphora is sensitive to the same types of pragmatic factors as anaphora in general.

Our interpretation of the results from this first experiment depends crucially on the assumption that the lack of differences in response times to a test word across conditions indicates a lack of differences in levels of comprehension for the pronoun across conditions. That is, we assume that the referent of the pronoun was correctly identified in all conditions. In several follow-up experiments (also reported in McKoon et al. 1990), this assumption was tested. For these experiments a new final sentence was written for each text, in which the pronoun was replaced by a nominal that had not previously appeared in the text. For example, the new final sentence for the text in Table 1 was *And he thinks bears are really exciting to track* (cf. *And he thinks they are really exciting to track*). With the new nominal, there is no pronominal reference to deer in the final sentence, and therefore there should be no facilitation of re-

sponse times to *deer* when it appears as a test word. That is, response times to the text word should be facilitated when the final sentence contains the pronoun, relative to when the sentence contains a new nominal, if the referent of the pronoun was actually identified during reading. This pattern, of course, should only obtain for the original test word (e.g. *deer*). With some other test word from the text (e.g. *trips*), response times should not be affected by the substitution of a new nominal for the original pronoun. The results of these follow-up experiments fully supported these predictions, thereby justifying the assumption that the test-word response times in the original results (Table 2) do indicate that the pronouns in question were understood across conditions, and that, consequently, reading times did in fact reflect comprehension difficulty.

In this section, we have argued that outbound anaphora is a fully grammatical anaphoric process of English whose felicity—like that of all grammatical phenomena—is determined by discourse context. Outbound anaphora thus contrasts sharply with inbound anaphora, which has been shown to be categorically ungrammatical. In the next section we discuss another grammatical restriction on anaphora in English.

3.3. OUTBOUND ANAPHORA INVOLVING *DO SO* AND *DO IT*. In distinguishing between 'deep' and 'surface' anaphora, Sag & Hankamer 1984 argued that surface anaphors are 'syntactically controlled' in that they require an explicit linguistic antecedent, while deep anaphors, being 'pragmatically controlled', do not.[17] Consider, for example, the contrast in 27 between the surface anaphor *do so* and the deep anaphor *do it* (examples from Sproat & Ward 1987:331):

(27) a. A: I'm going to lift this 500 lb. barbell.
B: With your back, do you think you should {do it, do so}?
b. [A bends down to lift a 500 lb. barbell]
B: With your back, do you think you should {do it, *do so}?

From these examples, we see that the explicit occurrence of a (VP) antecedent is required for felicitous use of *do so*. No such morphosyntactic restriction applies to the deep anaphor *do it*; indeed, there need be no explicit antecedent at all.

Sproat & Ward 1987 noted that, contra Postal, reference to an action evoked by a verb contained within a nominal is felicitous with the anaphor *do it*, but not with *do so*.[18] Consider first the following examples of felicitous *do it* anaphora:

(28) a. Mary is a heavy *smoker*—even though her doctor keeps telling her not to *do it*.

[17] The terms 'deep' and 'surface' anaphora—first introduced in Hankamer & Sag 1976—are replaced in Sag & Hankamer 1984 by the (more accurate) terms 'model-interpretive anaphora' and 'ellipsis', respectively. However, the original terms are still the ones generally used in the literature to describe the distinction between the two types of anaphoric processes, even by Sag & Hankamer themselves in 1984.

[18] We assume, following Webber 1979, that verb phrases that denote actions or events can evoke discourse entities.

b. In response to his wife's strenuous objections, Bill isn't much of a *sportscar racer* any more, but he still manages to *do it* every once in a while.

The surface anaphor *do so*, which requires an explicit VP antecedent, does not pattern in the same way: the examples of *do so* anaphora in 29, corresponding to the examples of *do it* in 28 above, are much worse:

(29) a. *Mary is a heavy *smoker*—even though her doctor keeps telling her not to *do so*.

b. *In response to his wife's strenuous objections, Bill isn't much of a *sportscar racer* any more, but he still manages to *do so* every once in a while.

Note that the corresponding examples of *do so* anaphora with full-VP antecedents are fully acceptable, as illustrated in 30:

(30) a. Mary *smokes* heavily—even though her doctor keeps telling her not to *do so*.

b. In response to his wife's strenuous objections, Bill doesn't *race sportscars* very much any more, but he still manages to *do so* every once in a while.

Unlike other anaphors, then, *do so* is highly constrained in terms of the morphosyntactic form of possible antecedents (Hankamer & Sag 1976, Sag & Hankamer 1984). Assuming that this constraint is a grammatical one, and given our working assumption that truly ungrammatical violations cannot be salvaged by pragmatic factors, it follows that no discourse context will render *do so* anaphora felicitous with non-VP antecedents. The examples in 29 illustrate the categorical unacceptability of such anaphora.

This distinction between surface and deep anaphora makes a number of empirically testable predictions. If we assume, following Sag & Hankamer 1984, that deep VP anaphors such as *do it* are understood with respect to a discourse model, then their interpretation should be sensitive to pragmatic factors, presumably the same kinds of pragmatic factors to which pronomial outbound anaphora was found to be sensitive. Furthermore, under this assumption deep VP anaphors should be sensitive to morphosyntactic factors only to the extent that these factors indirectly affect the accessibility of the referent event in the discourse model.[19] By contrast, a surface VP anaphor such as *do so*, being sensitive to the linguistic representation of its antecedent, should be more sensitive to morphosyntactic factors than to pragmatic ones.

These hypotheses were also tested in the series of psycholinguistic experiments described above (McKoon et al. 1990). The same experimental design used to investigate pronominal anaphora was used to investigate surface versus deep anaphora, first with the deep anaphor *do it* used in place of the pronominal anaphor (see Table 3). The accessibility of the referent event for the VP anaphor

[19] Murphy 1985 and Tanenhaus & Carlson 1990 have shown that syntactic parallelism between a deep VP anaphor and its antecedent does appear to affect comprehension difficulty for the anaphor. However, with the materials used in their experiments, parallelism probably affected the discourse-level representation of the antecedent event.

Deep anaphor (do it)	Surface anaphor (do so)
Nominalized antecedent/Non-Topical	**Nominalized antecedent/Non-Topical**
Joe does not have a very good sense of reality. Last year, he told everybody that he was going to go to law school. He isn't. In fact, he'll soon be dropping out of college. Next, he said he was dating a Vogue model. He wasn't. Now, he claims to be a good basketball player.	Joe does not have a very good sense of reality. Last year, he told everybody that he was going to go to law school. He isn't. In fact, he'll soon be dropping out of college. Next, he said he was dating a Vogue model. He wasn't. Now, he claims to be a good basketball player.
But, in fact, he's never done it.	But, in fact, he's never done so.
Nominalized antecedent/Topical	**Nominalized antecedent/Topical**
Joe is generally considered to be the best athlete Central High School has ever had. He swims; he's the star pitcher of the baseball team; and he is a defensive end on the varsity football team. And since he's 6'6", people naturally assume that he's a basketball player.	Joe is generally considered to be the best athlete Central High School has ever had. He swims; he's the star pitcher of the baseball team; and he is a defensive end on the varsity football team. And since he's 6'6", people naturally assume that he's a basketball player.
But, in fact, he's never done it.	But, in fact, he's never done so.
VP antecedent/Non-Topical	**VP antecedent/Non-Topical**
Joe does not have a very good sense of reality. Last year, he told everybody that he was going to go to law school. He isn't. In fact, he'll soon be dropping out of college. Next, he said he was dating a Vogue model. He wasn't. Now, he claims to play basketball well.	Joe does not have a very good sense of reality. Last year, he told everybody that he was going to go to law school. He isn't. In fact, he'll soon be dropping out of college. Next, he said he was dating a Vogue model. He wasn't. Now, he claims to play basketball well.
But, in fact, he's never done it.	But, in fact, he's never done so.
VP antecedent/Topical	**VP antecedent/Topical**
Joe is generally considered to be the best athlete Central High School has ever had. He swims; he's the star pitcher of the baseball team; and he is a defensive end on the varsity football team. And since he's 6'6", people naturally assume that he plays basketball.	Joe is generally considered to be the best athlete Central High School has ever had. He swims; he's the star pitcher of the baseball team; and he is a defensive end on the varsity football team. And since he's 6'6", people naturally assume that he plays basketball.
But, in fact, he's never done it.	But, in fact, he's never done so.

TABLE 3. Examples of texts with VP anaphors.

was manipulated in the same way that accessibility was manipulated for the referent of the pronoun. Topicality was varied by manipulating either the contrast between the referent event and other discourse events or the relation between the referent event and the overall topic of the text. As in the other experiments, the topical contexts were predicted to make the referent event more accessible than the nontopical contexts, thus facilitating comprehension of the deep anaphor in the final sentence. Morphosyntactic structure was varied as before, with the antecedent occurring either within a nominalization or as a verb phrase. There was no reason to believe that these two structures differed with respect to the accessibility they contributed to the relevant event in the discourse model; therefore, the two structures were predicted not to differentially affect comprehension of a deep anaphor specifying that event.

Both predictions for *do it* were supported by the data. In the topical versions reading times for the final sentences averaged 1504 ms, while in the nontopical

versions they averaged 1552 ms (a significant difference by analyses of variance that did not interact with morphosyntactic structure); this shows the predicted effect of pragmatic factors on deep anaphora. However, the morphosyntactic structure of the antecedent did not significantly affect reading times of the final sentences (1532 ms for the nominalization vs. 1524 ms for the verb phrase). Apparently, the two structures did not differentially affect the accessibility of the referent event.

McKoon et al. 1990 established that there were no significant differences in comprehension of the anaphors across experimental conditions in the same way as in the experiment described in §3.2, using test words taken from the antecedent of the VP anaphor (e.g. *basketball* in Table 3). As expected, response times for these test words did not differ significantly across conditions.

Next, the surface anaphor *do so* was tested by replacing the *do it* anaphors in the previous experiment with *do so* anaphors. If it is true that surface anaphors are understood with direct reference to a linguistic representation, and only indirectly with reference to events in the discourse model, then replacing *do it* with *do so* should alter the effects of the pragmatic and morphosyntactic variables that were obtained in the earlier experiment. Whereas comprehension of the *do it* anaphor was affected by the topicality of the referent event in the discourse model more than by the morphosyntactic form of its antecedent, comprehension of *do so* should be affected more by linguistic form than by topicality. Again, the results were as predicted: when the antecedent for *do so* was contained within a nominalization, reading times for the final sentences averaged 1740 ms; when the antecedent for *do so* was a verb phrase, reading times averaged 1601 ms, demonstrating a significant effect of morphosyntactic structure (by analyses of variance) that did not interact with topicality. Reading times in the topical versus nontopical versions did not differ significantly (1686 ms vs. 1654 ms, respectively), indicating that topicality had no effect on comprehension of the surface anaphor. Overall, reading times for the *do so* sentences were slower than for the *do it* sentences, but the absence of significant differences in response times to test words selected from the antecedent indicated that there were no significant differences in comprehension of the anaphors across experimental conditions.

This psycholinguistic evidence supports our claim that outbound anaphora involving the pro-VP *do so* is not rendered more felicitous by the same pragmatic factors that facilitate other types of outbound anaphora. This result is predicted from the existence of a GRAMMATICAL restriction on the antecedent of *do so* and further supports our general contention that true morphosyntactic violations cannot be amnestied by pragmatic factors.

3.4. OUTBOUND ANAPHORA WITHOUT MORPHOLOGICAL CONTAINMENT. Up to now, we have dealt primarily with outbound anaphora involving antecedents that are morphologically contained within words. In this final section we would like to consider cases in which the antecedent of the pronominal anaphor is not morphologically contained in, or in some cases even morphologically related to, the words that introduce them. Consider the examples in 31:

(31) a. 'I heard someone say,' he began, 'that you are a *New Zealander*. I was out *there* as a small boy.' (Ngaio Marsh, *Night at the Vulcan* (1951:207). New York: Jove)
b. Jean is a *Frenchman*, though he hasn't lived *there* for many years.
c. This is the *fourteen*th time in *as many* weeks.
d. This is the *second* time in *as many* weeks. (= 10)
e. Mary is a *physicist*: she says *it*'s an exciting field.
f. Bill is a *linguist*: he says *it*'s an exciting field.

These data suggest that in some cases of outbound anaphora the morphological relationship between the word containing the antecedent and the antecedent itself need not be regular or even apparent. So, while *New Zealand* is clearly morphologically contained within *New Zealander*, the same cannot be said of the pair *France* and *Frenchman*. And while *fourteenth* is presumably derived from *fourteen* by suffixation of *-th*, there is no morphophonological relationship between the forms *two* and *second*. Finally, although *physicist* may be morphologically derived from *physics*, the relationship between *linguistics* and *linguist*, from a surface morphological point of view, appears to go in the opposite direction.

What the examples in 31 have in common is the fact that the link between the containing word and the intended antecedent is in each case an example of a well-instantiated LEXICAL relationship. Specifically, the pairs *New Zealand/New Zealander* and *France/Frenchman* are examples of the relationship between names of countries and names for inhabitants of those countries. This relationship is well instantiated in that it is quite generally the case that there is a term of provenance—usually unique within a given register—associated with each country name. Although there are subregularities, this relationship is by no means generally expressed in a morphologically regular fashion, as seen in 32:

(32)

COUNTRY	PROVENANCE TERM
France	Frenchman
New Zealand	New Zealander
Canada	Canadian
Brazil	Brazilian
America	American
Spain	Spaniard
Thailand	Thai
Denmark	Dane

However, the SEMANTIC relationship expressed by these examples is entirely regular and predictable: all of the nouns in the righthand column refer to a person living in or originating from the corresponding country in the lefthand column. Similarly, the pairs *fourteen/fourteenth* and *two/second* (31c–d) are particular instances of the well-instantiated—indeed, completely productive—relationship between a cardinal number and its associated ordinal. Again, the morphology is irregular for some of the more common cases (*first, second, third, fifth, twelfth*), but the semantics is entirely regular. And finally, *physics/*

physicist and *linguistics/linguist* (31e–f) are examples of the relationship between a field and a practitioner in that field.

To account for such cases, we would like to suggest that outbound anaphora is sensitive to the productivity (and semantic predictability) of the relationship between an anaphor's antecedent and the lexical item containing that antecedent. That is, *Frenchman* can evoke France in 31b precisely because the relationship between Frenchmen and the country France is sufficiently transparent due to the well-instantiated relationship of which the pair *France/Frenchman* is an instance.[20] Similarly, *second* can evoke the number two in 31c because of the well-instantiated and semantically transparent relationship between cardinal numbers and their associated ordinals.[21] Felicitous outbound anaphora, then, does not appear to require a morphological relationship in the strictest sense; a sufficiently clear and well-instantiated lexical relationship will suffice.

The lexical relationships exemplified in 31 are reminiscent of traditional inflectional paradigms (see, for instance, Matthews 1974:156). In both cases, there is a sense in which a word fills a particular 'slot' in a paradigm that expresses some relationship between word forms.[22] In the case of the English past-tense paradigms, for example, *compiled* fills the past-tense slot of *compile*. Irregular forms are full-fledged members of the paradigm; the suppletive form *went* is as much the past-tense form of *go* as *compiled* is of *compile*. In a similar vein, *Frenchman* could be said to fill the provenance slot in a paradigm relating it to the place term *France* (as in the set of paradigms in 32 above); despite the irregular morphology, it is no less a provenance term than the regular form *New Zealander*. Although the notion of paradigm has traditionally been used in the description of inflectional morphology, there is no a priori reason for that restriction; the lexical relationship expressed in the examples in 32 is quite similar to the relationship among inflectional verb forms.

However, not all instances of outbound anaphora are best analyzed in terms

[20] A similar well-instantiated relationship seems to hold between a place and the language spoken there. Consider the naturally-occurring token in (i):

(i) I had *French* for eight years and I've never been *there*. (Prospective apartment renter in conversation; April 12, 1987)

[21] Watt 1975 also discusses the possibility of felicitous outbound anaphora with cases like *two* and *second*, but offers a very different analysis (see §2.2 above). Watt argues that, when a 'hidden antecedent [is] so circumscribed as perforce to be one particular word', then outbound anaphora is possible. In the case of 'the second blow for freedom in as many weeks' (1975:111), Watt argues that the 'number anaphor' *as many* requires a numerical antecedent, and since the 'set of possible antecedents is so circumscribed that it has only the one member, "two" itself (1975:112)', outbound anaphora is possible, indeed 'forced'. However, the explanation appears not to be that *as many* forces any particular antecedent (although it certainly does do that), but that *second* is so transparent. In our terms, *second* is transparently related to the cardinal number two, and therefore its use will serve to render it sufficiently accessible for subsequent reference.

[22] While paradigm slots are usually filled by a unique word form, some slots are occasionally filled by more than one form, e.g. the English plural forms *cacti* and *cactuses*. However, as Aronoff (1976) and others have noted, there is a strong tendency for the existence of a filled slot to 'block' additional forms.

of the paradigmatic lexical relationships exemplified in 31. In 22b, for example, it seems specious to analyze *Kal Kan cat* as filling a slot in a paradigm that relates things to cats liking those things; there is no well-instantiated 'cat__which__likes__x' paradigm in the English lexicon. Instead, as is the case with most of the examples discussed in this paper, outbound anaphora is felicitous in 22b because the discourse entity *Kal Kan* is sufficiently accessible to permit subsequent anaphoric reference to it, due in part to the morphological presence of the brand name *Kal Kan*. Thus, we suggest that there are in fact two sources for the contained antecedent in examples like 31a: one is the paradigmatic relationship that the containing-word/contained-word pair instantiates, and the other is the actual morphological presence of the contained word; *New Zealander* both morphologically contains *New Zealand* and is paradigmatically related to it qua provenance term.

Given this analysis, outbound anaphora is predicted to be generally infelicitous when there exists neither a morphological relationship between an anaphor's antecedent and the lexical item containing that antecedent, nor a paradigmatic lexical relationship of the kind exemplified in 31. This prediction appears to be borne out by the data. Consider again Postal's classic orphan example in 33.

(33) #Max is an *orphan* and he deeply misses *them*. (cf. 2a)

First, it is clear that *orphan* and *parents* are not morphologically related. Second, although the words *orphan* and *parent* might be formally related, given certain assumptions about the lexicon, it is clear that they do not form part of a well-instantiated lexical relationship. So, while one can find (or construct) an appropriate provenance term for a given country or city term, there is no general pattern such that for some term *x*, there is a word meaning 'person whose *x* has died'; only a few such pairs exist in English, namely *orphan/parent*, *widow/husband*, and *widower/wife*. It is this lack of morphological or paradigmatic lexical relationship, we claim, that renders 33 infelicitous.

However, in a more suitable context, even anaphora paralleling that in 33 is possible:

(34) 'That depends on whose mother she is,' Fitz told him. 'Mine has brown hair—hardly a bit of grey in it. Your mother's hair probably turned white in a night long ago.'
'I haven't got a mother,' said Johnny pathetically, staring at his ham sandwich. 'I'm an orphan.'
'Why, that's terrible, Johnny, when did it happen? You never told me you were an *orphan*.' Fitz was deeply concerned.
'I'm getting sort of used to it. *They* died when I was three.' (Elswyth Thane, *Ever after* (1945:155). New York: Hawthorn Books; noted by Beth Levin)

Here, in a context in which the existence of one's parents is under discussion (without an explicit mention of *parents*), subsequent pronominal anaphora is possible despite the absence of any morphological or lexical relationship.

CONCLUSION

4. Previous accounts of outbound anaphora have attempted to rule it out by means of various morphological and syntactic principles. Instead, we have argued that outbound anaphora is fully grammatical and that, like anaphora in general, its felicity is a function of the accessibility of the discourse entity to which the anaphor in question is used to refer. We have identified a number of morphosyntactic, semantic, and pragmatic factors that increase the accessibility of discourse entities—and therefore the felicity of outbound anaphora. Our analysis is supported by a series of psycholinguistic studies which show that topicality and contrast facilitate comprehension of word-internal anaphors.

APPENDIX

Below are some naturally-occurring tokens of outbound anaphora classified according to the type of discourse entity evoked. Specifically, we have classified the examples according to whether the word-internal antecedent:

- is a proper name or common noun which evokes a specific referent in the discourse corresponding to that name or noun;
- is a common noun and evokes an individual corresponding to a kind in the discourse;
- is a common noun and evokes an individual corresponding to a mass in the discourse.

On the italicization conventions for indicating coreference, see note 1.

1. SPECIFIC REFERENTS:

1. RS: Well, she got an *LSA* paper out of it.
   JH: Yes, she was *there*.
   (Julia Hirschberg and Richard Sproat in conversation: January 30, 1987)
2. A: It has something to do with *Suez* prices.
   B: Did it mean anything to you?
   A: I dunno. His father was a general *there*.
   ('Still Crazy Like a Fox': April 5, 1987)
3. I had *French* for eight years and I've never been *there*.
   (Prospective apartment renter in conversation: April 12, 1987)
4. GW: Excuse me, sir, but what's the *tray* situation?
   CW: I'll bring *them* right out.
   (Gregory Ward and cafeteria worker, Ida Noyes Hall, University of Chicago: April 23, 1987)
5. A 508-page manuscript of nine *Mozart* symphonies written in *his* own hand in Salzburg in the 1770's, before the composer's 20th birthday, was auctioned yesterday by Sotheby's in London for $4.34 million. (*New York Times* article, 'Record Price for Mozart Manuscript': May 23, 1987)
6. There's a *balance sheet* concern—we've never had to read *it* before. (Arno Penzias: September 29, 1987)
7. Thanks for the *Philly* dirt—I have never been *there* but if I ever do [sic] I'll let you know. (Message on electronic bulletin board: 1988)
8. Our *postscript printer* room had some water problems (under the floor) this weekend. *They* should be back up by 10:00am [sic] ... (Don Bock in email: 1987)
9. There's a *Thurber* story about *his* maid ... (Michael Riley in conversation: September 7, 1988)
10. I didn't know you had a *Joan Miller*-fan. Was this *her* office? (Michael Riley in conversation: September 12, 1988)

11. We went up to *Constable* country; we stayed in the village *he* was born in. (Kenneth Sproat in conversation; October 11, 1988)
12. You couldn't find a stronger *Dukakis*-supporter. The only way I wouldn't vote for *him* ... (Michael Riley in conversation; October 18, 1988)
13. ... that Mario Biaggi could not survive a long *jail*-sentence; that he would die *there*. (WINS; November, 1988)
14. RS: You don't know *Chinese*, I assume?
    PCS: I've been *there*, but I don't speak it.
    (Richard Sproat in conversation with prospective MIT Coop student)
15. *Bush* supporters would stay home, figuring *he*'d already won. (Julia Hirschberg in conversation; November 9, 1988)
16. I refer you to the *Schachter* paper; *he*'s very proud of it ... (Mark Baker in response to a question at NELS; November 12, 1988)
17. A cheer went up at *Mulroney* headquarters in *his* hometown of Baie-Comeau, Quebec, when the CBC made its first projection. (Associated Press Newswire; November 21, 1988)
18. I was an *IRS*-agent for about 24 years ... I stopped working for *them* ... (Radio ad for AARP heard December 31, 1988)
19. Well, action is still needed. If we're to finish the job, Reagan's Regiments will have to become the *Bush* Brigades. Soon *he*'ll be the chief, and he'll need you every bit as much as I did. (Ronald Reagan, farewell speech, January 11, 1989; reported in Associated Press Newswire)
20. *Museum* visitors can see through *its* big windows the 900-year-old Tower of London and the modern office blocks of the City financial district. (Associated Press Newswire; July 5, 1989)
21. 'Sometime [sic] they say, "Get away from me, I don't want to hear that *Jesus* stuff."' he said. 'But I think deeply of *him*. He's always with me and I want other people to know he can be with them, too.' (Associated Press Newswire; August 29, 1989)
22. *Rolling Stones* fans: clear your calendars! *They*'re adding more concert dates. (WCBS 11 O'clock News; September 26, 1989)
23. Spokesmen for the federal prosecutor's office in Karlsruhe said they viewed the letter as an authentic claim of responsibility from the Red Army Faction, which had been dormant for three years until the *Herrhausen* assassination. *His* armored Mercedes was blown up by a remote-control bomb in Bad Homburg, where he lived, as he was being driven to work Nov. 30. (Associated Press Newswire; December 5, 1989)
24. Millions of *Oprah Winfrey* fans were thoroughly confused last week when, during *her* show, she emotionally denied and denounced a vile rumor about herself. (*Chicago Tribune*, column by Mike Royko; May 22, 1989; cited in James McCawley's '1989 linguistic flea circus,' as an example of reflexive usage—not as an example of outbound anaphora)
25. The *Paris* idea holds a lot of charm, 'cuz I used to live *there*, y'know. (Greg McKenna in conversation; March 14, 1990)
26. Do *parental* reactions affect *their* children? (from Jill Burstein, uttered by one of her students; March 15, 1990)
27. 'My daughter knows the *German* people. She's been *there* [...]' (*Chicago Tribune* article, 'Holocaust revisionism: A family's strident salute'; May 8, 1990)
28. 'I heard someone say,' he began, 'that you are a *New Zealander*. I was out *there* as a small boy.' (Ngaio Marsh, *Night at the Vulcan*. 1951:207. New York: Jove)
29. JL: So, what's your *child* situation?
    DS: *He*'s 4.
    (Judith Levi and Deborah Schiffrin in conversation; May 22, 1990)

30. You know, this is a *Pilgrim* town here: *they* came into this harbor... (ABC 'Nightline'; July 4, 1990)
31. Nancy: The whole thing—it was like suddenly being caught in a *Diane Arbus* picture. Do you ever get that feeling?
    Elliot: Hourly.
    Nancy: Me too.
    Elliot: Yeah, we never shoulda had *her* take our wedding picture.
    ('thirtysomething'; July 17, 1990)
32. Another *Nixon* Summit, at *His* Library. (Title of article in *New York Times*; July 20, 1990)
33. In the distance, we heard the sound of an *ambulance* siren. Within a minute or so *it* arrived and stretcher bearers took the boy away. (*New York Times Magazine*, 'The tragedy of Detroit'; July 29, 1990, p. 25)
34. The *Senator Bradley* forum has been canceled due to *his* need to be in Washington for the budget vote. (Note on poster at AT&T Bell Labs; September 26, 1990)
35. I was reading this *Peggy Noonan* book on *her* years at the White House ... (Julia Hirschberg in conversation; November 9, 1990)
36. There's no reason to become a *California* citizen, unless I'm gonna live *there*. (Ken Baime to Gregory Ward in conversation; August 8, 1990)
37. I used a *gutter* person before, but just to clean *them*. (Julia Hirschberg in conversation; October 13, 1990)
38. *Saudi* anti-aircraft guns fired on *Iraqi* planes along *their* common border. (NBC Nightly News; August 11, 1990)
39. Last night's *Sinead O'Connor* concert at the Garden will be *her* last. (WNBC 6:00 News; August 25, 1990)
40. I think if I were a *Peruvian* I wouldn't want to live *there* for the next couple of years. (John Kingston in conversation; September 6, 1990)
41. Heisenberg had bitter words to say about the lack of funds and materials, and the drafting of scientific men into the services. Excerpts from *American* technical journals suggested that plenty of technical and financial resources were available *there* for nuclear research. (Albert Speer, *Inside the Third Reich*, translated by Richard Winston and Clara Winston (1970:225–26). New York: Collier)
42. AMA: Cut *AIDS* 'protection'
    Doctors want *it* handled like other sexual diseases (Title of article in *Chicago Tribune*; December 6, 1990)
43. 'That depends on whose mother she is,' Fitz told him. 'Mine has brown hair—hardly a bit of grey in it. Your mother's hair probably turned white in a night long ago.'
    'I haven't got a mother,' said Johnny pathetically, staring at his ham sandwich. 'I'm an orphan.'
    'Why, that's terrible, Johnny, when did it happen? You never told me you were an *orphan*.' Fitz was deeply concerned.
    'I'm getting sort of used to it. *They* died when I was three.' (Elswyth Thane, *Ever After* (1945:155). New York: Hawthorn Books)
44. Our neighbors, who are sort of *New York City*-ites, they have jobs *there* ... (Ginny Beutnagel in conversation; December 30, 1990)

2. KINDS:

    1. GW: So, Roger, do they even HAVE venison in New Zealand?
       RR: Oh yes. They have a real *deer* problem. *They*'ve been running around eating all the forests.
       (Gregory Ward and Roger Ratcliff in conversation; 1987)
    2. I had a *paper* route once but my boss said I took too long deliverin' *'em*. ('L.A. Law'; 1987)

3. ... play the *Cutlass-Supreme* Game and win *one* ... (Radio ad heard on WINS; May 20, 1988)
4. 47th-St. Photo announces its *microwave oven* sale, just when you need *it* for your apartment ... (Radio ad heard on WINS; November 11, 1988)
5. I'm a *mystery-story* buff and read (and watch on PBS) a lot of *them*. (Northwestern University electronic bulletin board; January, 1989)
6. ... the only way to solve this *homeless* problem, say those who work with *them* ... (WCBS 11 O'clock News; January 4, 1989)
7. We asked *Saab 9000-CD* owners about *its* road-handling ... (Television ad for Saab; March 12, 1989)
8. If you're a *small business* owner, or interested in starting *one* ... (TV ad; June 14, 1989)
9. Game show host: So, I hear you're a real *cat*-lover. *How many* do you have now? ('Jeopardy'; July 24, 1989)
10. Euripides—described by Sophocles as a *woman*-hater in his tragedies, but very fond of *them* in bed—complained that they were always having other women 'coming into the house gossiping.' ... (Reay Tannahill, *Sex in history* (1982:95). New York: Stein & Day)
11. CHECK *VISA* REQUIREMENTS FOR INTERNATIONAL TRAVEL
    Though many popular destinations don't require *them*—including Canada, Mexico, England, much of the Caribbean and Europe, Japan, Thailand, and Hong Kong—the majority of countries still do. (Column in *Money Magazine*; January 1990, p. 144)
12. Officials in the Danish capital believe they've found a way to stop *bicycle* thefts—let people use *them* for free. (Associated Press Newswire; November 10, 1990)

3. MASS TERMS:

1. For a *SYNTAX* slot, I'd rather see someone with more extensive coursework in *it*. (Judith Levi discussing various subdisciplines of linguistics; January 18, 1987)
2. Patty is a definite *Kal Kan* cat. Every day she waits for *it*. (TV ad for Kal Kan; January 28, 1987)
3. There does not seem to me to be a serious *snow* problem. There is *some*, but no large accumulation is forecast. (Mark Liberman in email; 1987)
4. It's awfully *foggy* tonight so you people out there driving better watch out for *it*. (Heard on Chicago radio station; April 16, 1987)
5. Cliff Barnes: Well, to what do I owe this pleasure?
   Ms. Cryder: Actually, this is a *BUSINESS* call, and I'd like to get right down to *it*. ('Dallas'; 1987)
6. Chang Ching-hui uttered a third saying when the Japanese were making so many compulsory *grain* purchases that the peasants of the Northeast had *none* left. (Aisin-Gioro Pu Yi, *From emperor to citizen*, translated by W. J. F. Jenner (1987:287). Oxford: Oxford University Press)
7. MR: How did you become an *AI* person?
   JH: I got a degree in *it*.
   (Michael Riley and Julia Hirschberg in conversation; October 4, 1988)
8. I know, you probably get eight gazillion jokes from *pragmatics* students each semester you teach *it*, but maybe this one you haven't seen: (Ellen Prince in email; October 5, 1988)
9. I don't rejoice in the *stock* market going down so much, now that we started owning *some*. (Dan Hirschberg in conversation; November 11, 1988)
10. At the same time as coffee beans were introduced, the Arabs made changes in *coffee* preparation that greatly improved *its* flavor. (J. Schapira, D. Schapira and K. Schapira, *The book of coffee und tea* (1982:7). New York: St. Martin's Press)

11. Jo Ann Smith is a *beef* person. She grew up on *it* and remains a great fan of the standing rib roast. (Associated Press Newswire; July 6, 1989.)
12. 'Anyhow,' he said, 'it is nearly *Luncheon* Time.' So he went home for *it*. (A. A. Milne, *Winnie-the-Pooh* (1926:41). London: The Reprint Society)
13. 'It must be getting on for *luncheon* time,' he remarked to the Otter. 'Better stop and have *it* with us ...' (Kenneth Grahame, *The wind in the willows* (1908:92). London: The Reprint Society)
14. They're afraid it's the *Gas and Electric* man come to turn *that* off. (Interview with Baltimore politician, ABC 'Nightline'; March 23, 1990)
15. Very well. But I warn you that if you continue in such foolishness you'll be the last *paleontologist* alive by the time you retire. There's no future in *it*. (Stephen Jay Gould, 'In touch with Walcott', *Natural History*, July 1990:16)
16. Although casual *cocaine* use is down, the number of people using *it* routinely has increased. (WCBS 11 O'clock News; December 20, 1990)

REFERENCES

ARONOFF, MARK. 1976. Word formation in generative grammar. Cambridge, MA: MIT Press.

BAUER, LAURIE. 1983. English word-formation. Cambridge: Cambridge University Press.

BRESNAN, JOAN, and SAM MCHOMBO. 1987. Topic, pronoun and agreement in Chichewa. Lg. 63.741–82.

BROWNE, WAYLES. 1974. On the topology of anaphoric peninsulas. Linguistic Inquiry 5.612–20.

CARLSON, GREG N. 1977. Reference to kinds in English. Amherst, MA: University of Massachusetts dissertation.

CHAFE, WALLACE. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. Subject and topic, ed. by Charles Li, 25–55. New York: Academic Press.

CHOMSKY, NOAM. 1981. Lectures on government and binding. Dordrecht: Foris.

CORUM, CLAUDIA. 1973. Anaphoric peninsulas. Chicago Linguistic Society 9.89–97.

DISCIULLO, ANNA MARIA, and EDWIN WILLIAMS. 1987. On the definition of word. Cambridge, MA: MIT Press.

FABB, NIGEL. 1984. Syntactic affixation. Cambridge, MA: MIT dissertation.

GREENE, STEVEN; GAIL MCKOON; and ROGER RATCLIFF. 1990. Discourse models and anaphoric reference. Evanston, IL: Northwestern University, MS.

GROSZ, BARBARA. 1977. The representation and use of focus in dialogue understanding. Berkeley, CA: University of California dissertation.

——, and CANDACE SIDNER. 1986. Attention, intentions and the structure of discourse. Computational Linguistics 12.175–204.

GUNDEL, JEANETTE, and NANCY HEDBERG. 1990. Givenness, implicature, and the form of referring expressions in discourse. Berkeley Linguistics Society 16. 442–53.

HANKAMER, JORGE, and IVAN A. SAG. 1976. Deep and surface anaphora. Linguistic Inquiry 7.391–426.

HIRSCHBERG, JULIA, and JANET PIERREHUMBERT. 1986. The intonational structuring of discourse. Proceedings of the 24th annual meeting, 136–44. New York: Association for Computational Linguistics.

KARTTUNEN, LAURI. 1976. Discourse referents. Syntax and semantics VII: Notes from the linguistic underground, ed. by James McCawley, 363–86. New York: Academic Press.

KIPARSKY, PAUL. 1982. Lexical phonology and morphology. Linguistics in the morning calm, ed. by In Seok Yang, 3–91. Seoul: Hanshin.

LAKOFF, GEORGE, and JOHN ROSS. 1972. A note on anaphoric islands and causatives. Linguistic Inquiry 3.121–25.

LEVI, JUDITH N. 1978. The syntax and semantics of complex nominals. New York: Academic Press.
LIEBER, ROCHELLE. 1984. Grammatical rules and sublexical elements. Papers from the parasession on lexical semantics, Chicago Linguistic Society 20.187–99.
MATTHEWS, PETER. 1974. Morphology. Cambridge: Cambridge University Press.
MCKOON, GAIL, and ROGER RATCLIFF. 1989. Semantic association and elaborative inference. Journal of Experimental Psychology: Learning, Memory, and Cognition 15.326–38.
——; GREGORY WARD; ROGER RATCLIFF; and RICHARD SPROAT. 1990. Morphosyntactic and pragmatic manipulations of salience in the interpretation of anaphora. Evanston, IL, and Murray Hill, NJ: Northwestern University and AT&T Bell Laboratories, MS.
MOHANAN, K. P. 1986. The theory of lexical phonology. Dordrecht: Reidel.
MORGAN, JERRY. 1978. Toward a rational model of discourse comprehension. Proceedings of TINLAP-2, ed. by David Waltz, 109–14. New York: Association for Computing Machinery.
MURPHY, GREGORY. 1985. Psychological explanations of deep and surface anaphora. Journal of Pragmatics 9.785–813.
PESETSKY, DAVID. 1979. Russian morphology and lexical theory. Cambridge, MA: MIT, MS.
PIERREHUMBERT, JANET, and JULIA HIRSCHBERG. 1990. The meaning of intonational contours in the interpretation of discourse. Intentions in communication, ed. by Phil Cohen, Jerry Morgan, and Martha Pollack, 271–311. Cambridge, MA: MIT Press.
POSTAL, PAUL. 1969. Anaphoric islands. Chicago Linguistic Society 5.205–39.
PRINCE, ELLEN F. 1981a. Toward a taxonomy of given-new information. Radical pragmatics, ed. by Peter Cole, 223–55. New York: Academic Press.
——. 1981b. Topicalization, focus-movement, and Yiddish-movement: A pragmatic differentiation. Berkeley Linguistics Society 7.249–64.
——. 1986. On the syntactic marking of presupposed open propositions. Papers from the parasession on pragmatics and grammatical theory, Chicago Linguistic Society 22.208–22.
RADFORD, ANDREW. 1988. Transformational grammar. Cambridge: Cambridge University Press.
RATCLIFF, ROGER, and GAIL MCKOON. 1988. A retrieval theory of priming in memory. Psychological Review 95.385–408.
REINHART, TANYA. 1981. Pragmatics and linguistics: An analysis of sentence topics. Philosophica 27.53–94.
——. 1983. Anaphora and semantic interpretation. Chicago: Chicago University Press.
ROOTH, MATS. 1985. Association with focus. Amherst, MA: University of Massachusetts dissertation.
ROSS, JOHN. 1971. The superficial nature of anaphoric island constraints. Linguistic Inquiry 2.599–600.
——. 1986. Infinite syntax. Norwood, NJ: Ablex.
ROTHKOPF, ERNEST; MARY KOETHER; MARJORIE BILLINGTON; and BARBARA BIESENBACH. 1988. Why are certain sentence constructions mnemonically robust for modifiers? New York: Columbia University, MS.
SAG, IVAN A., and JORGE HANKAMER. 1984. Toward a theory of anaphoric processing. Linguistics and Philosophy 7.325–45.
SHIBATANI, MASAYOSHI, and TARO KAGEYAMA. 1988. Word formation in a modular theory of grammar: Postsyntactic compounds in Japanese. Lg. 64.451–84.
SIDNER, CANDACE. 1979. Towards a computational theory of definite anaphora comprehension in English discourse. Cambridge, MA: MIT dissertation.
SIMPSON, JANE. 1983. Aspects of Warlpiri morphology and syntax. Cambridge, MA: MIT dissertation.
SPROAT, RICHARD. 1985. On deriving the lexicon. Cambridge, MA: MIT dissertation.

——. 1988. On anaphoric islandhood. Theoretical morphology, ed. by Michael Hammond and Michael Noonan, 291–301. Orlando: Academic Press.
——, and GREGORY WARD. 1987. Pragmatic considerations in anaphoric island phenomena. Chicago Linguistic Society 23.321–35.
TANENHAUS, MICHAEL K., and GREG N. CARLSON. 1990. Comprehension of deep and surface verb phrase anaphors. Language and Cognitive Processes 5.257–80.
TIC DOULOUREUX, P.R.N. 1971. A note on one's privates. Studies out in left field: Defamatory essays presented to James D. McCawley on the occasion of his 33rd or 34th birthday, ed. by Arnold Zwicky et al., 45–52. Edmonton & Champaign: Linguistic Research, Inc.
VAN DIJK, TEUN A., and WALTER KINTSCH. 1983. Strategies of discourse comprehension. New York: Academic Press.
WATT, W. 1975. The indiscreteness with which impenetrables are penetrated. Lingua 37.95–128.
WEBBER, BONNIE. 1979. A formal approach to discourse anaphora. New York: Garland Press.
WILSON, DEIRDRE, and DAN SPERBER. 1979. Ordered entailments: An alternative to presuppositional theories. Syntax and semantics XI: Presupposition, ed. by Choon-Kyu Oh and David Dinneen, 299–323. New York: Academic Press.

Gregory Ward
Department of Linguistics
Northwestern University
2016 Sheridan Road
Evanston, IL 60208

[Received 30 July 1990;
accepted 2 November 1990.]

Testing Theories of Language Processing:

An Empirical Investigation of the On-Line Lexical Decision Task

Gail McKoon, Roger Ratcliff, and Gregory Ward

Northwestern University

Short Title: On-Line Lexical Decision

Address correspondence to Gail McKoon, Psychology Department, Northwestern University, Evanston, IL, 60208.

Abstract

On-line lexical decision has been used to test major theoretical hypotheses about language comprehension. Contrary to several current models, Sharkey and Sharkey (1992) found that a word in a sentence did not give facilitation to an immediately following, highly associated test item. We show in this article that such facilitation can be obtained. Other theories have proposed that syntactic processes supply antecedents for implicit anaphors. Using a test item that was an associate of the antecedent of the anaphor, we were unable to replicate previous findings of facilitation at but not before the site of the anaphor. Across nine experiments, obtaining facilitation depended on the choice of control condition. This dependency raises questions about previous on-line lexical decision results that have been used to support the immediacy of syntactic processing.

Testing Theories of Language Processing:

An Empirical Investigation of the On-Line Lexical Decision Task

Theories of language comprehension vary widely in their goals. Some attempt to explain the moment-by-moment processes that construct meaning as one individual word is read after another (e.g. Kintsch, 1988). Others attempt to explain the processes that organize words into syntactic structures that show the roles played by the individual words (Fodor, in press; Frazier, 1987; Frazier & Rayner, 1982; Nicol & Swinney, 1989; Swinney & Osterhout, 1990; Rayner & Morris, 1991). Still others are concerned with inferences that might integrate the pieces of a text into a wholistic representation in memory (e.g. Glenberg, Meyer, & Lindem, 1987; McKoon & Ratcliff, 1992). Efforts to test all of these theories share a major problem: finding empirical procedures that allow investigation of the processes or structures of theoretical interest. In this article, we report the results of several experiments designed to analyze one empirical procedure that has frequently been employed: on-line lexical decision.

In on-line lexical decision experiments, the words of a text are presented to subjects one word at a time, either visually or auditorily. At some point in the text, a test string of letters is presented visually. The subject is asked to decide, as quickly and accurately as possible, whether the string of letters is a word. Reaction time and accuracy are recorded.

The on-line lexical decision technique has been used to investigate comprehension of both word meanings and syntactic structures. One of the first uses was by Swinney (1979), whose aim was to examine the processing of ambiguous words. In his experiments, subjects listened to sentences like "The man was not surprised when he found several spiders, roaches, and other bugs in the corner of his room", which contains the ambiguous word *bugs*. While listening, the subjects watched a fixation point on a CRT screen. Immediately after the ambiguous word, a test word replaced the visual fixation point. The lexical decision response for the test word was facilitated if it matched either of the meanings of the ambiguous word; for example, following *bugs*, responses were facilitated for both *spy* and *ant*.

More recently, on-line lexical decision has been used to test the claims of general theories of meaning comprehension. Kintsch (1988; see also Ratcliff & McKoon, 1988) has proposed that meaning is constructed from the words of a text by processes that first activate the associates of individual words and then integrate the activated concepts into a representation of the meaning of the whole text. When words are read, all of their associates- even those that will turn out to be irrelevant to the meaning of the text- are

activated (with varying degrees of strength). Then, through a repeated recycling of activation, concepts that are associated to other activated concepts are strengthened while concepts that are not associated to other activated concepts are weakened. Once this cyclic integration process stabilizes, the result is a representation of the meaning of the text.

It is fundamental to Kintsch's theory (and others such as Dosher & Rosedale, 1989; Ratcliff & McKoon, 1988) that relations among words be immediately available during reading. For example, if a sentence contains the word *boy*, the relation between *boy* and *girl* should be immediately available. Sharkey and Sharkey (1992) tested whether this was the case with on-line lexical decision. The words of sentences were presented visually, at a rate of 200 ms per word. When a test string was presented, it replaced the next word of the text, so that the interval between onset of the word preceding the test and onset of the test was 200 ms. Sharkey and Sharkey used test words that were strong associates of words in the text, and found that responses were not facilitated. In other words, when *girl* was tested 200 ms after *boy* was presented, Sharkey and Sharkey found no facilitation of the response to *girl*. If this result were supported with further empirical evidence, it would be problematic for any theory postulating the immediate availability of well-known relations among words. However, in the experiments reported in this article, we find, contrary to Sharkey and Sharkey, that relations among words do support immediate facilitation in on-line lexical decision.

From most theoretical viewpoints, our result is not surprising. That is, it is not surprising that the explicit mention of a word should lead to facilitation of associates of the word. A more controversial claim is that the *implicit* mention of a concept can also lead to facilitation of associates. Consider, for example, the sentence *The instructors held the skier that the waitress in the lobby blamed for the theft*. Complete understanding of this sentence requires knowing that the person who was blamed was the skier, not the waitress or an instructor. Current psycholinguistic theories (Fodor, in press; Nicol & Swinney, 1989; Swinney & Osterhout, 1990) claim that this knowledge is computed by syntactic processes. These processes compute a syntactic structure for the sentence, and in the computed structure of the sentence above, there is a "trace" following the verb *blamed*. This trace is an implicit anaphor for the object of *blamed*, and the only syntactically possible antecedent for the anaphor is *skier*, to which the anaphor should be syntactically bound. Thus, syntactic processing should associate the "gap" after *blamed* with its antecedent *skier*.

Several researchers (Fodor, in press; Nicol & Swinney, 1989; Swinney & Osterhout, 1990) have tested syntactic gap-filling with on-line lexical decision. They have hypothesized that the gap-filling process

results in "activation" of the antecedent word at the gap site. For example, in the *skier* sentence, *skier* would be hypothesized to be activated immediately after the verb *blamed*. This activation, in turn, is hypothesized to lead to activation of associates of the antecedent word (e.g. *snow* as an associate of *skier*).

To examine the syntactic gap-filling process, Nicol and Swinney (1989) used sentences like the *skier* sentence above. Sentences were presented to subjects auditorily, and lexical decision test items were presented visually. Test items were chosen so as to measure the availability of potential fillers at two sites: immediately after the verb in the relative clause (the gap site) and immediately before the verb. Nicol and Swinney's results were consistent with the gap-filling hypotheses. After the verb, but not before it, the lexical decision for an associate of the syntactically determined antecedent of the wh-trace was facilitated. Lexical decisions for associates of other nouns in the sentence were not facilitated. So, for the *skier* sentence, *snow* would be facilitated when tested after the verb, but *restaurant* would not be. The overall pattern of results- facilitation for an associate of the syntactically determined antecedent, and only this antecedent, and facilitation for this antecedent after but not before the verb- suggests that the intended filler does in fact become available at the gap site.

The research reported in this article was originally planned to extend the findings of Nicol and Swinney (1989) to other linguistic phenomena. However, we found that we could not replicate the original Nicol and Swinney results. This failure led us to explore the on-line lexical decision paradigm, and Experiments 1 through 9 report the results of our efforts.

Much theoretical weight has been placed on data collected with the on-line lexical decision procedure. Sharkey and Sharkey's (1992) result from on-line lexical decision stands virtually alone as data contradicting major models designed to account for relations among the meanings of words (Anderson, 1983; Dosher & Rosedale, 1989; Kintsch, 1988; Ratcliff & McKoon, 1988). These models accommodate large ranges of other kinds of data.

Similarly, the results of Nicol and Swinney (1989), Swinney and Osterhout (1990), and Fodor (in press) have been applied to important and controversial hypotheses about syntactic processing. First, facilitation of an associate of the correct antecedent at its gap site would indicate that some kind of syntactic processing is engaged early in sentence processing. Second, it has been claimed that this processing proceeds independently of other kinds of information: Swinney and Osterhout (1990) found facilitation at a gap site for the correct antecedent even when it was much less plausible than other nouns in the sentence. For example, in the sentence *Everyone watched the enormous heavyweight boxer that the small 12-year old*

boy on the corner had beaten so brutally, real-world knowledge would suggest *the boy* as the object of *beaten*. Yet facilitation was obtained only for the syntactically correct object *boxer* (Swinney & Osterhout, 1990). This result was offered in support of the highly influential notion of modularity proposed by Fodor (1983). According to this notion, syntactic processing proceeds independently of other kinds of information such as semantics or pragmatics. Third, on-line lexical decision results have formed part of the data base used to distinguish among different linguistic theories (cf Fodor, in press). Facilitation in lexical decision has been found for the kinds of traces postulated in some linguistic theories, but not for the kinds of traces postulated by other linguistic theories. Fourth, Fodor (in press) has used the difference in patterns of results between on-line lexical decision and other tasks as part of the support for a distinction between two levels of linguistic information, phonetic form and surface structure. Finally, Chomsky (1990) pointed to the significance of gap-filling results as a reason that linguists should take the empirical research of psychologists into account in their theorizing.

All of these claims are under debate and none of the debates has been resolved. It is not our intention to present a detailed review of these theoretical positions or to contribute to the theoretical debates except indirectly through evaluation of the lexical decision procedure and results. However, this evaluation should serve to promote increased methodological concern in the design of future experiments.

Experiments 1 - 5

As mentioned above, our experiments were originally designed to replicate and extend results from earlier experiments described by Nicol and Swinney (1989). Therefore, our procedures and materials were modeled on theirs. Experiments 1 - 5 are summarized in Table 1.

We used two sets of sentences, both of which consisted of sentences with object-gap relative clauses. One set, which we labeled "complex", is exemplified by the *skier* sentence: *Two instructors held the skier that the waitress in the lobby blamed for the theft.* The sentences of this set were designed to have the same syntactic structures as those used by Nicol and Swinney (1989), with a wh-trace after the verb of the relative clause. The second set of sentences was constructed in order to provide some generality of results across sentence types. These sentences were simplified versions of the complex sentences, formed by simplifying the noun phrases and eliminating the prepositional phrase in the relative clause. For example, the simplified version of the *skier* sentence was *Somebody held the skier that Doctor Hillcroft blamed for the theft.* The "simple" sentences had a gap in the same (post-object) position as the complex sentences and

contained the same verbs in the relative clauses as the "complex" sentences with the same antecedents for the wh-traces that followed the verbs. Another example of a pair of sentences is: *The nun hated the ballerina that the senator from the north nominated for the council*, and *John hated the ballerina that an old friend nominated for the council*. Each sentence had one test word, an associate of the antecedent of the wh-trace (e.g. *snow* for the antecedent *skier*, and *dance* for the antecedent *ballerina*).

In Experiments 1 and 2, sentences were presented visually, one word at a time on a CRT screen. In Experiments 3 - 5, sentences were presented auditorily. In all the experiments, the lexical decision test items were presented visually.

Across the experiments, three different test positions were used (see Table 2). A test word in the first test position was presented immediately after the antecedent of the wh-trace (immediately after *skier* in the example sentence). In the second test position, the test word immediately preceded the verb in the relative clause. In this test position, the test word always followed the object of the prepositional phrase in the complex sentences and it always followed the subject noun of the relative clause for the simple sentences. In the third position, the test word immediately followed the verb of the relative clause (this was the gap position).

INSERT TABLES 1 AND 2 ABOUT HERE

A critical feature of Experiments 1 - 5 is the choice of a baseline against which to measure facilitation for the associate of the antecedent of the wh-trace. For example, if *snow* was tested in position 1, immediately after *skier*, then we might expect to see facilitation of the response time to *snow*. But the question is: facilitation with respect to what control test word? We chose as a control test word the associate of the antecedent from some other sentence. For example, the associate test word for the *skier* sentence was *snow*, and the control test word might have been *dance*. Thus, the same words were used as test items in the two conditions: the associated condition, in which a sentence was tested with the test word associated to the antecedent for the wh-trace, and the control condition. The only difference was that in the control condition a sentence was tested with the associate of some other sentence. This choice for control test words has several design advantages: First, it controls for any characteristics of the individual test words that might affect lexical decision response times or accuracy rates. For example, the frequencies in English of the control test words are exactly the same as the frequencies of the associate test words because they are the same words. Second, the mean response times for associated test words represent means across exactly the same words as the mean response times for the control test words, again because they are exactly the same

words. Third, any interactions between test words and test positions are controlled. Some possible test words might be facilitated or inhibited because they somehow "fit" or failed to fit the test positions in ways other than those under study. For example, an inanimate test word might show inhibition in a test position immediately following a verb because most of the verbs in our sentences take animate objects. Once more, using the same test words in both conditions controls for this potential problem.

Method

Materials. The set of complex sentences contained 28 sentences of the form: noun phrase, verb, noun phrase, *that*, noun phrase, prepositional phrase, verb, adjunct phrase. These sentences averaged 15 words in length. Each complex sentence was changed into a simple sentence by simplifying the first and third noun phrases and deleting the prepositional phrase. The simple sentences averaged 12 words in length. The second noun phrase and the verb of the relative clause were the same in both the simple and complex versions. The test word for each sentence was an associate of the noun in the second noun phrase (which was the antecedent of the wh-trace following the relative clause verb). The complete set of antecedents and their associated test words was: skier-snow, journalist-news, ballerina-dance, architect-building, locksmith-key, gardener-flowers, secretary-typing, convict-prisoner, boy-girl, photographer-camera, woman-lady, millionaire-rich, sculptor-statue, victim-injury, writer-novel, duchess-duke, poet-verse, gangster-mob, soldier-army, cowboy-Indian, baker-bread, doctor-nurse, junkie-drugs, comedian-laugh, jockey-horse, zoologist-animals, cobbler-shoes, musician-song. The complete set of complex sentences is shown in Appendix 1. The simple sentences were used in Experiments 1-4 and the complex sentences in Experiment 5.

There were also 48 filler sentences, averaging 14 words in length. Each of the filler sentences had one test item; 14 of these were words and 34 were nonwords. The test positions for these items were scattered randomly through the sentences, so that subjects could not anticipate which word in a sentence would be followed by a test item.

Visual Presentation Procedure. Sentences and test items were presented on a CRT screen, with responses collected from the CRT's keyboard. Stimulus presentation and response recording were controlled by a real-time computer system.

In Experiments 1 and 2, the sentences and test items were presented visually. The experiments began with a practice list of 30 lexical decision test items (without any sentences) to familiarize subjects

with the response keys. Then the 28 experimental sentences and the 48 filler sentences were presented in random order, with the random order changed after each second subject. Each sentence began with an instruction displayed on the CRT screen to press the space bar on the keyboard to initiate a sentence. The words of a sentence were presented one at a time, with all letters in lower case except for the first letters of the first words of sentences and the first letters of proper nouns. Each word was displayed for 170 ms plus 17 ms multiplied by the number of letters in the word; then the word was erased from the screen, and the next word was displayed. Each word was displayed at the same location on the CRT screen. Test items were displayed five spaces to the right of the location for words of the sentences, and test items were marked with two trailing asterisks. There was no extra time between a word of a sentence and the test item that immediately followed it, so the stimulus onset asynchrony (SOA) between the word of the sentence and the test item was 170 ms plus 17 ms multiplied by the number of letters in the sentence word. Test items were displayed in lower case. A test item remained on the screen until subjects made a response, "?/" for "word" and "z" for "nonword." Then the test word was erased and the words of the sentence continued after a 170 ms pause. Subjects were instructed to respond quickly and accurately to the test items. To encourage the subjects to read the sentences, they were occasionally given a recall test: After eight randomly chosen sentences, subjects were asked to write down the last sentence they had read. One test item proved problematic with visual presentation: *indian* (used as the associate of *cowboy*) was presented without the first letter capitalized and, probably as a consequence, it showed slow responses overall, so it was deleted from the analyses of results.

Auditory Presentation Procedure. In Experiments 3 - 5, the sentences were presented auditorily via headphones, and the test items were presented visually on a CRT screen. The sentences were recorded by a male speaker at a natural speaking rate. Test positions for a sentence were located by examining an amplitude-time plot of the sentence; a test position following a word of the sentence was defined as the point of lowest activity between that word and the next word. If there was no single point at which activity was lowest, the test position was located at the end of the range of lowest activity farthest ~~the end farthest~~ from the preceding word, but never overlapping the next word.

The experiments began with the same 30 lexical decision practice items as for the visual presentation experiments. Then the 28 experimental sentences and the 48 filler sentences were presented in random order, the same random order for each subject. A row of plus signs was displayed on the CRT screen as a fixation point at all times except when a test item was presented. The sentences were presented one after

another with about a 2 s pause between each sentence. At the test position for a sentence, the plus signs were replaced by the test item, which remained on the screen either until the subject responded or until 1800 ms had elapsed. Auditory presentation of the sentence continued during the interval that the test item remained on the screen. Subjects were instructed to respond quickly and accurately to the test item, pressing the "?/" key for a word and the "z" key for a nonword. As in the visual experiments, they were asked to recall in writing eight randomly chosen sentences.

Subjects and Designs. In each experiment, there were 32 subjects participating for credit in an introductory psychology class at Northwestern University.

For the first experiment, there was one test position: immediately following the second noun of the sentence (which was the antecedent of the wh-trace), position 1 in Table 2. There were two experimental conditions: the test word for a sentence was either the associate of the second noun of the sentence (the associated condition) or the associate of the second noun of some other sentence (the control condition). These two conditions were combined with groups of subjects and groups of sentences in a Latin square design.

Experiments 2 through 5 all had the same design, each employing two test positions. In Experiment 3, these positions were immediately after the second noun (test position 1, as in Experiment 1) and immediately before the verb of the relative clause (test position 2, see Table 2). In Experiments 2, 4, and 5, the second and third positions (immediately before and after the verb of the relative clause) were used. In each case, there were four experimental conditions: the two test positions crossed with the two test word conditions (associated and control). The four conditions were combined with groups of subjects and groups of sentences in a Latin square design.

When a sentence was tested in the control condition, the test word was the associate of the antecedent of some other of the 28 experimental sentences. Which other sentence was chosen randomly (without replacement), with the randomization changed after every second subject. No test item was presented to a subject more than once.

Results

Slow outlier response times (times longer than 1500 ms) were excluded from the analyses; these made up about 1.5% of the data in each experiment. Means of correct responses were calculated for each subject and each test item in each condition, and means of these means are shown in Table 1. Analyses of

variance were performed on the means, with both subjects, F1, and items, F2, as random variables, $p < 0.05$.

The pattern of results is presented in Table 1. First, when a test word immediately followed its associate in a sentence (test position 1), response time was facilitated. This was true both in Experiment 1 with visual presentation and in Experiment 3 with auditory presentation. This finding stands in clear contrast to Sharkey and Sharkey's (1992) failure to find facilitation in a similar experiment.

Second, at the gap position (position 3) following the verb, where there is hypothesized to be a wh-trace to serve as an anaphor, there is little evidence of facilitation. In these experiments, implicit mention of the antecedent through its anaphor did not serve to significantly facilitate responses for the associate of the antecedent.

The only test position at which results are somewhat equivocal is test position 2, immediately before the verb of the relative clause. In Experiment 3, the associate of the antecedent was facilitated, but this was not the case in Experiments 2, 4, and 5. We cannot offer any reason for this discrepancy.

Analyses of variance confirmed the conclusions just stated. For the first test position, there was significant facilitation of response times in Experiment 1, $F1(1,31)=5.55$ and $F2(1,26)=4.03$. In Experiment 3, there was significant facilitation at both the first and second test positions, $F1(1,31)=7.33$ and $F2(1,24)=6.78$. A planned test confirmed facilitation at the first test position, $F1(1,28)=4.37$ and $F2(1,24)=5.36$.

There were no significant effects on response times of any other variables in any of the experiments ($F's < 2.7$) except that in Experiment 2, responses were significantly faster in test position 3 than in test position 2 in the analysis of the subject means, $F1(1,31)=4.51$ and $F2(1,26)=3.31$. There were no significant differences among error rates, $F's < 2.7$.

The standard errors of the response time means in the five experiments were, in order: 7.3 ms, 22.3 ms, 12.8 ms, 13.0 ms, and 11.6 ms. Response times and error rates for filler test items are shown in Table 4.

An additional analysis was performed on the data from test positions 2 and 3 to investigate the possibility that the failure to obtain a difference between the associated and control conditions at the second and third test positions was due to spuriously fast responses in the *control* condition. Fast responses could arise in the control condition if the test words in that condition happened, by random assignment, to be associated (against our intentions) to either the antecedent of the implicit anaphor or to other words in the sentences with which they were tested. To eliminate this possible explanation of the results, we eliminated

from the analyses all the test words that were associated to any words in any sentences other than their own sentence. We eliminated all the test words that were associated in any way we could think of, by even quite weak associations, a total of 16 test words (which eliminated data about equally across the four counterbalancing groups of items). For example, we eliminated the test word *girl* because it might be associated to words from other sentences than its own, such words as *secretary* or *woman*. If such associations had speeded responses in the control condition, then eliminating these test words should lead to slower responses in the control condition than the associated condition, but this did not happen. Responses in the two conditions were still virtually identical, differing by no more than 5 ms.

Discussion

The results obtained in Experiments 1 - 5 contradict previous findings. Contrary to Sharkey and Sharkey (1992), we found that a word in a sentence facilitated response time on an immediately following test of an associated word. Our result, unlike Sharkey and Sharkey's, is consistent with current models of the processing of relations among words. Models that postulate spreading activation processes predict that presentation of a word will facilitate subsequent decisions on other words related to it (Anderson, 1983; Kintsch, 1988). Models that postulate compound cue kinds of retrieval mechanisms similarly predict that relations among related words will be quickly available to facilitate decisions (Dosher & Rosedale, 1989; Ratcliff & McKoon, 1988).

We can only speculate about why we were able to demonstrate immediate facilitation and Sharkey and Sharkey (1992) were not. They used fewer subjects, and perhaps variance was higher in their experiment. This is plausible because a 45 ms effect in their experiment (due to the position in a sentence at which a test word was presented) was not significant. Also, in their experiment, lexical decision test items were distinguished from words of the sentences by color of the lettering, green versus white. Perhaps the green lettering served in some way to switch processing away from the words of the sentences.

Our results were also different from previous findings when we tested for facilitation due to an implicit presentation of an associate of a test word. Nicol and Swinney (1989) reported facilitation at the site of an implicit anaphor. In sentences with syntactic structures like our sentences, they found a pattern of facilitation at the wh-trace site following a verb but no significant facilitation before the verb. Our results show no evidence of this pattern.

We thought that the reason for our failure to find the previously reported pattern of facilitation

might be our choice of control condition. As explained in the introduction, we believed that using the same pool of words in both conditions, associated and control, was an optimal experimental design. However, the control condition that has been used by Nicol and Swinney (1989), Swinney and Osterhout (1990), and Fodor (in press) is different- they used a different pool of words in the two conditions. In their designs, there were two test words for any given sentence, always the same two words. One of the words is the associate of the antecedent of the trace (e.g., the associate *snow* for the antecedent *skier*). The other word, the control, is a word unrelated to the meaning of the sentence, with the same number of letters and the same frequency in the English language as the associated word. We thought that this difference in choice of control condition between our Experiments 1 - 5 and previous experiments might account for the difference in results, and we tested this hypothesis in Experiments 6 - 9.

Experiments 6 - 9

These four experiments are outlined in Table 3. Both the simple and complex versions of the sentences were used, and sentences were presented both auditorily and visually. The only difference from the comparable experiments in the first series (Experiments 1 through 5) was in the control condition. A new pool of control words was chosen, one word for each sentence, such that the control word for a sentence had the same number of letters and approximately the same frequency in English as the associate test word (according to Kucera & Francis, 1967).

Method

Materials and Procedure. The sentences and their associated test words were the same as in Experiments 1 - 5, and the only change was in the words used in the control condition. The procedures for the experiments were also the same as in Experiments 1 - 5. The antecedents with their new control words were: skier-uses, journalist-clay, ballerina-equal, architect-material, locksmith-add, gardener-evident, secretary-afloat, convict-symmetry, boy-trade, photographer-affect, woman-file, millionaire-camp, sculptor-morale, victim-define, writer-stone, duchess-buys, poet-marks, gangster-ads, soldier-list, cowboy-warren, baker-seeds, doctor-graph, junkie-dried, comedian-shots, jockey-doubt, zoologist-perfect, cobbler-grown, musician-dust.

Subjects and Design. There were 32 subjects in each of Experiments 6 and 7, 24 subjects in Experiment 8, and 20 subjects in Experiment 9, all from the same population as in Experiments 1 - 5. Except for the new control words, the designs of the experiments and randomization procedures were the same as

in the earlier experiments.

Results

The data were analyzed in the same manner as for Experiments 1 - 5, and the means are displayed in Table 3.

INSERT TABLES 3 AND 4 ABOUT HERE

In test position 2, responses to the associate test word were faster than responses to the control test word in every one of the experiments. The same is true for test position 3, except in Experiment 6. For Experiments 7, 8, and 9, responses to the associate are faster than responses to the control word at test position 3, but this pattern reverses in Experiment 6, for no apparent reason.

Analyses of variance confirmed these observations. For Experiments 7, 8, and 9, the main effect of faster responses for the associate than the control was significant; for these three experiments in order, $F1(1,31)=7.21$, $F2(1,26)=4.47$, $F1(1,23)=16.01$, $F2(1,27)=24.98$, $F1(1,19)=9.28$, $F2(1,27)=11.85$. Other effects on responses times were not significant, all F's <3.23. The standard errors for the means were, respectively, 10.5 ms, 14.2 ms, and 17.8 ms. There were generally more errors on the control words than the associates, and this effect was sometimes significant. For the three experiments in order: $F1(1,31)=5.74$, $F2(1,26)=2.90$, $F1(1,23)=4.02$, $F2(1,27)=1.84$, $F1(1,19)=6.33$, and $F2(1,27)=6.20$. All other effects on error rates were not significant, F's < 2.3. For all of Experiments 6 through 9, the standard errors on the error rates varied between 1.0 and 1.5%.

The pattern in Experiment 6 was different. The interaction between test word and test position was significant for response times, $F1(1,31)=7.36$ and $F2(1,26)=11.18$. The main effect of test word was also significant in the subjects analysis, $F(1,31)=8.80$, but not in the items analysis, $F2(1,26)=2.25$. The main effect of test position was not significant, F's < 2.05. The standard error of the response time means was 12.0 ms. There were marginally more errors on the control test words, $F1(1,31)=4.14$ and $F2(1,26)=3.58$. Other effects on error rates were not significant, F's < 1.85.

Two aspects of the data should be pointed out. First, over the series of nine experiments, which included 17 different comparisons of associate and control response times, results were inconsistent for two of the comparisons (test position 2 in Experiment 2 and test position 3 in Experiment 6). This suggests that any results from the on-line lexical decision procedure should be replicated across experiments to ensure a high degree of confidence in the general patterns that emerge. Second, the F values for significant effects

were always higher with auditory presentation of the sentences than with visual presentation. This might have come about for a variety of reasons, but it is worth bearing in mind for future research.

The conclusions from Experiments 6 through 9 and comparisons of their results with those of Experiments 1 through 5 are straightforward. The first five experiments used the same pool of words as test words in the associated and control conditions. For these experiments, in six out of seven cases there was no facilitation at test positions 2 or 3. The last four experiments used different pools of words as test words in the associated and control conditions. For these experiments, in seven out of eight cases there was facilitation at both of test positions 2 and 3. It appears that the choice of control word was critical in determining the results.

General Discussion

The experiments reported here were designed to investigate the use of on-line lexical decision tests in the study of sentence comprehension. Lexical decision test words were presented at one of several points during a sentence. In the associated condition, the test word was highly associated to one of the words in the sentence, and it was tested either immediately after the associated word in the sentence, or at one of two later positions in the sentence. The results of our experiments depended on the choice of control test words; whether the control test words were the same words as for the associated condition (simply switched to sentences for which they were not associated) or whether the control test words were different words from the associated test words. If the control words were the same as the associated words, then there was facilitation of response times for the associated words relative to the control words at the immediate test position but not at later test positions. If the control test words were different from the associated test words, then facilitation was observed at the later test positions. These two conclusions held up over 15 of the 17 comparisons afforded by the nine experiments.

The finding that an associated word is facilitated when it is tested immediately after a related word in a sentence is intuitively compelling and also not surprising from most theoretical viewpoints. It would be expected that a lexical decision test of *snow* immediately following the sentence fragment ...*the skier* would result in facilitation of response time to *snow*, and this is what we found. Although Sharkey and Sharkey (1992) recently failed to find immediate facilitation, their result may well be anomalous. The variance among response times in their experiment appears to have been high (as mentioned above), and their failure is inconsistent not only with the results described here but also with a considerable amount of previous

research. On-line facilitation has been found with lexical decision test positions at the ends of sentences or sentence fragments (McKoon & Ratcliff, 1989a; 1989b; O'Seaghdha, 1989; Till, Mross, & Kintsch, 1988) and with on-line text experiments that use a variety of other paradigms including measurements of word by word reading times, phoneme monitoring latencies, and naming latencies (cf Foss & Speer, 1991; McKoon & Ratcliff, 1981; 1989c; Simpson, Peterson, Casteel, & Burgess, 1989; Stanovich & West, 1981). On-line facilitation for associated test words is also consistent with on-line facilitation for the multiple meanings of ambiguous words (Onifer & Swinney, 1979; Swinney, 1979; Tanenhaus, Leiman, & Seidenberg, 1979). Furthermore, the finding of on-line facilitation for associated words gains considerable validation in another important way: consistency with a wide range of different kinds of data is established by virtue of its incorporation into comprehensive theories of memory (Anderson, 1983; Kintsch, 1988; Ratcliff & McKoon, 1988). Thus, a large body of previous research argues in favor of accepting the validity of our finding of immediate facilitation.

It is important to stress the differences among the theories with which immediate facilitation is consistent. According to spreading activation theories (e.g., Anderson, 1983; Kintsch, 1988), presentation of a word in a sentence activates the concept in memory that corresponds to the word. The activation spreads to other related concepts, so that they, in turn, become activated. If one of these activated concepts is then presented as a test word for lexical decision, its response time will be facilitated because it was already activated prior to its presentation. In these theories, activation spreads quickly, so that the response on a test word can be facilitated even if presentation of an associated word preceded it by as little as 100 ms. The main competitors for spreading activation theories are theories that assume memory retrieval is based on a compound cue mechanism (Dosher & Rosedale, 1989; Ratcliff & McKoon, 1988). In these theories, the process by which immediate facilitation occurs is very different than spreading activation. There is no anticipatory activation of the test word. Instead, words presented to the system are assumed to join together in short-term memory to form a compound cue. This cue has some degree of familiarity, where familiarity is determined by the strengths of associations between the compound in short-term memory and items in long-term memory. Familiarity is calculated by a matching process that matches the cue in short-term memory against all the items in long-term memory. The immediate facilitation observed in the experiments reported here is consistent with the compound cue view because a lexical decision for an associated test word will be facilitated by a high familiarity value for the cue made up of the test word and the immediately preceding word of the sentence. Recently, compound cue theories and spreading activation theories have been extensively tested against each other, but both still seem to be viable accounts of retrieval from long-

term-memory (McKoon & Ratcliff, 1992b; McNamara, 1992a; 1992b; Ratcliff & McKoon, submitted).

The implications of the immediate facilitation effect found in our experiments are quite different when viewed from the two different theoretical perspectives. For spreading activation, immediate facilitation would be taken to indicate that reading a word in a sentence makes related concepts in memory immediately available. But for compound cue theories, immediate facilitation does not, in itself, indicate what happens during reading of the words in sentences. No conclusions can be drawn about what would happen if the test word was not presented. The facilitation in response time is a reflection only of the situation in which short-term memory contains both the word of the text and the test word. What the two kinds of theories share is the assumption that, however the facilitation comes about, it should happen quickly, within about 100 ms.

While our finding of immediate facilitation for related text and test words is consistent with most previous work, the patterns of facilitation we obtained for tests of implicit anaphors are not. A number of researchers have reported testing for the availability of antecedents at several different kinds of gap sites (Fodor, in press; Nicol & Swinney, 1989; Swinney & Osterhout, 1990). For sentences like *Two instructors held the skier that the waitress in the lobby blamed for the theft*, Nicol and Swinney (1989) found that response times for an associate of the antecedent for the wh-trace following the verb of the relative clause were facilitated when tested immediately after the verb but not when tested immediately before the verb; that is, *snow* would be facilitated when tested after *blamed* but not when tested before *blamed*. This pattern of facilitation after the verb but not before is the finding that has been used to argue for the re-activation of the antecedent of the wh-trace. But in neither of our sets of experiments did we find this pattern. When we chose control test words from the same pool of words as the associated test words, we did not find facilitation either before or after the verb. When we chose control test words from a different pool of words than the associated test words, we found facilitation at both test points.

Why did we fail to replicate previous results? One possible answer to this question is suggested by the dramatic effect of the choice of control condition. We got very different patterns of facilitation with the two different control conditions. This logically opens up the possibility that with other sets of control words, other patterns of data might emerge. With another set of control words, we might have replicated exactly the pattern that has been obtained in previous experiments (e.g. Nicol & Swinney, 1989). The most serious issues raised by our results are how to choose the "right" set of control words, whether there is any one correct set, and how researchers might go about defending the choice of control words used in their

experiments over some other choice.

We can only offer tentative suggestions about why the choice of control words might be so important. We know that the syntactic fit of a test word to its test position can affect response times (Clifton, Frazier, & Connine, 1984; Wright & Garrett, 1984). In Wright and Garrett's experiments, a test word either fit the syntactic context of the sentence fragment that preceded it or it did not, and lexical decisions were slowed when it did not. This suggests that there might also be a host of other reasons why different words have different response times at different test positions in a sentence, including the words' meaningfulness values, concreteness values, likelihoods of appearing in sentences of the type used in the experiments, and so on. For example, consider the sentences used in our experiments; they almost all took the form that "some person *verbed* someone that another person *verbed*." Some words, because of their semantics or pragmatics, just will not easily fit in such sentences. *Marshmallow* is a case in point. In a context that includes sentences about an employer confronting a secretary that an accountant fired, *marshmallow* seems out of place. Moreover, there may be subtle interactions between the syntactic and semantic contexts of a sentence and test position. To give a few examples of verbs from our sentences, we cannot *blame, suspect, bribe, nominate, appoint, drive*, or *assault* a *marshmallow*, so *marshmallow* might fit particularly badly in a test position following a verb and perhaps less badly in a test position at the end of a phrase before the verb. Again, our current state of knowledge about these issues only allows speculation. The important point is that attention must be paid to the choice of control words in future experiments. As this issue is investigated further, we may be able to understand why previously used sets of control words have given the results they did, whether or not the control words in an experiment should come from the same pool of words as the associated words, and what the important variables are that govern the response time for a word tested in the middle of a sentence.

In conclusion, the theoretical implications from our results can be easily outlined. First, previous research on syntactic gap-filling and the suggestion from that research that syntactic processes occur early and fast are called into question. Until we understand better how control words should be chosen, it may be that the case for fast syntactically based gap filling processes will have to be made from other paradigms (cf Bever & McElree, 1988; Boland, Tanenhaus, & Garnsey, 1990; Foss & Speer, 1991; Frazier & Clifton, 1989; Garnsey, Tanenhaus, & Chapman, 1989; McElree & Bever, 1989; Rayner & Morris, 1991; Stowe, 1986). Second, theoretical enterprises that have depended on on-line lexical decision results (cf Fodor, 1989; in press; Nicol & Swinney, 1989; Swinney & Osterhout, 1990) will have to be reworked, either with

new lexical decision evidence or with reliance on other kinds of empirical evidence.

References

Anderson, J.R. (1983). *The architecture of cognition.* Cambridge: Harvard University Press.

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

Boland, J.E., Tanenhaus, M.K., & Garnsey, S.M. (1990). Evidence for the immediate use of verb control information in sentence processing. *Journal of Memory and Language, 29,* 413-432.

Chomsky, N. (1990). *Address to the Cognitive Science Society.* Boston, MA.

Clifton, C., Jr., Frazier, L., & Connine, C. (1984). Lexical expectations in sentence comprehension. *Journal of Verbal Learning and Verbal Behavior, 23,* 696-708.

Dosher, B.A., & Rosedale, G. (1989). Integrated retrieval cues as a mechanism for priming in retrieval from memory. *Journal of Experimental Psychology: General, 2,* 191-211.

Fodor, J.A. (1983). *The modularity of mind.* Boston: MIT Press.

Fodor, J.D. (1989). Empty categories in sentence processing. *Language and Cognitive Processes, 4,* 155-209.

Fodor, J. D. (in press). Processing empty categories: A question of visibility. *Language and Cognitive Processes.*

Foss, D., & Speer, S. (1991). Global and local context effects in sentence processing. In R. Hoffman & D. Palermo (Eds.), *Cognition and the symbolic processes: Applied and ecological perspectives.* Hillsdale, N.J.: Erlbaum.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and Performance 12: The Psychology of Reading.* London: Erlbaum Assoc.

Frazier, L., & Clifton, C. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Process, 4,* 93-126.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in analysis of structurally ambiguous sentences. *Cognitive Psychology, 14,* 178-210.

Garnsey, S.M., Tanenhaus, M.K., & Chapman, R.M. (1989). Evoked potentials and the study of sentence comprehension. *Journal of Psycholinguistic Research, 18,* 51-60.

Glenberg, A.M., Meyer, M., & Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory and Language, 26*, 69-83.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*, 163-182.

Kucera, H., & Francis, W. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

McElree, B., & Bever, T.G. (1989). The psychological reality of linguistically defined gaps. *Journal of Psycholinguistic Research, 18*, 21-35.

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

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

McKoon, G., & Ratcliff, R. (1989b). Assessing the occurrence of elaborative inference with recognition: Compatibility checking vs. compound cue theory. *Journal of Memory and Language, 28*, 547-563.

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

McKoon, G., & Ratcliff, R. (1992a). Inference during reading. *Psychological Review, 99*, 440-466.

McKoon, G., & Ratcliff, R. (1992b). Spreading activation versus compound cue accounts of priming: Mediated priming revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 1155-1172.

McNamara, T.P. (1992a). Theories of Priming: I. Associative distance and lag. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 1173-1190.

McNamara, T.P. (1992b). Priming and constraints it places on theories of memory and retrieval. *Psychological Review, 99*, 650-662.

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

Onifer, W., & Swinney, D.A. (1981). Accessing lexical ambiguities during sentence comprehension:

Effects of frequency of meaning and contextual bias. *Memory and Cognition, 9*, 225-236.

O'Seaghdha, P.G. (1989). The dependence of lexical relatedness effects on syntactic connectedness. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 73-87.

Rayner, K., & Morris, R.K. (1991). Comprehension processes in reading ambiguous sentences: Reflections from eye movements. In G. Simpson (Ed.), *Understanding word and sentence*. Amsterdam: North Holland Press.

Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review, 95*, 385-408.

Ratcliff, R., & McKoon, G. (1993). *Retrieving information from memory: Spreading activation theories versus compound cue theories*. Manuscript submitted for publication.

Sharkey, A., & Sharkey, N. (1992). Weak contextual constraints in text and word priming. *Journal of Memory and Language, 31*, 543-572.

Simpson, G., Peterson, R., Casteel, M., & Burgess, C. (1989). Lexical and sentence context effects in word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 88-97.

Stowe, L.A. (1986). Parsing WH-constructions: Evidence for on-line gap location. *Language and Cognitive Processes, 1*, 227-245.

Swinney, D.A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior, 18*, 645-659.

Swinney, D. & Osterhout, L. (1990). Inference generation during auditory language comprehension. In A. Graesser and G. Bower (Eds.), *The Psychology of Learning and Motivation, 25*, 17-33, New York: Academic Press.

Stanovich, K.E. & West, R.F. (1981). The effect of sentence context on ongoing recognition: Tests of a two-process theory. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 658-672.

Tanenhaus, M., Leiman, J., & Seidenberg, M. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior, 18*, 427-440.

Till, R.E., Mross, E.F., & Kintsch, W. (1988). Time course of priming for associate and inference words in

a discourse context. *Memory and Cognition, 16*, 283-298.

Wright, B. & Garrett, M. (1984). Lexical decision in sentences: Effects of syntactic structure. *Memory and Cognition, 12*, 31-45.

Appendix 1

Two instructors held the skier that the waitress in the lobby blamed for the theft.

The banker bribed the journalist that the cops in the subway suspected of the break-in.

The nun hated the ballerina that the senator from the north nominated for the council.

The pilot trusted the architect that the judge in the city acquitted of the forgery.

All the tenants appreciate the locksmith that the tailor in the basement chose for the job.

Three brothers pitied the gardener that the attorney for the museum banned from the show.

The employer confronted the secretary that the accountant at the racetrack fired for gross insubordination.

The witness recognized the convict that the teller in the cafeteria accused of violent behavior.

The clown amused the boys that the actress in the mink drove to the stadium.

The hostess greeted the photographer that the swimmer with pale skin encountered at the meeting.

The janitor called the woman that the farmer in the store saved from the blaze.

The cabby contacted the millionaire that the mailman on the scooter struck on the head.

Few parents knew the sculptor that the professor of African geography appointed to the committee.

The optometrist aided the victim that the barber in the airport hurt in the fight.

The chef envied the writer that the soprano with blue eyes followed all over town.

The announcer interviewed the duchess that the painter without a passport defrauded of the treasure.

Many artists admired the poet that the priest from the mountain visited at the penitentiary.

The bride identified the gangster that the carpenter at the barbecue attacked with a knife.

The dentist treated the soldier that the athlete with a beard punched in the tavern.

The bartender criticized the cowboy that the trucker from the factory assaulted with a rifle.

The lifeguard rescued the dog that the hobo with a rock forced off a cliff.

The students cheered the doctor that the firemen in the parade applauded for tremendous bravery.

The warden released the junkie that the sailor in the desert forgave for grand larceny.

The boxer heckled the comedian that the referee with striped pants invited to the club.

The librarian comforted the jockey that the outlaw at the funeral threatened with a stick.

The butler summoned the zoologist that the sheriff with strong arms arrested for extreme cruelty.

The king punished the cobbler that the ambassador on the patio caught with the jewels.

A bee stung the musician that the usher with the radio reprimanded for public drunkenness.

Author Note

This research was supported by NIDCD grant R01-DC01240 and AFOSR grant 90-0246 (jointly funded by NSF) to Gail McKoon and by NIMH grants HD MH44640 and MH00871 to Roger Ratcliff. We thank Dave Swinney for extensive discussions of this work.

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

Table 1

Results of Experiments 1 - 5 Response Times and Error Rates							
Test Position							
		1		2		3	
		RT	Error %	RT	Error %	RT	Error %
Visual							
Simple-E 1	Assoc.	775	2				
	Control	798	3				
Simple-E 2	Assoc.			972	2	926	3
	Control			977	3	938	6
Auditory							
Simple-E 3	Assoc.	730	1	738	1		
	Control	759	1	770	0		
Simple-E 4	Assoc.			776	3	760	2
	Control			770	2	765	2
Cmplx-E 5	Assoc.			760	1	774	1
	Control			771	3	762	1

Table 2

Examples of Sentences with Test Words and Test Positions

A Complex Sentence:

The instructor held the skier $_1$ that the waitress in the lobby $_2$ blamed $_3$ for the theft.

A Simple Sentence:

Somebody held the skier $_1$ that Doctor Hillcroft $_2$ blamed $_3$ for the theft.

Associate Test Word: snow

Table 3					
Results of Experiments 6 - 9 Response Times and Error Rates					
		Test Position			
		2		**3**	
		RT	Error %	RT	Error %
Visual					
Simple-E 6	Assoc.	872	4	871	1
	Control	922	6	854	6
Cmplx-E 7	Assoc.	845	5	827	3
	Control	873	9	881	6
Auditory					
Simple-E 8	Assoc.	760	3	753	4
	Control	816	8	832	5
Cmplx-E 9	Assoc.	753	3	742	2
	Control	819	9	784	7

Table 4				
Response Times and Error Rates for Filler Test Items				
	Words		Nonwords	
	RT	Error %	RT	Error %
Exp 1	826	8	818	6
Exp 2	999	8	1051	5
Exp 3	829	3	899	4
Exp 4	813	6	852	5
Exp 5	779	3	850	5
Exp 6	905	7	952	6
Exp 7	908	6	990	7
Exp 8	823	4	868	5
Exp 9	793	2	865	5

JOURNAL OF MEMORY AND LANGUAGE 32, 000–000 (1993)

Syntactic Prominence Effects on Discourse Processes

GAIL MCKOON, ROGER RATCLIFF, AND GREGORY WARD

Northwestern University

AND

RICHARD SPROAT

AT&T Bell Laboratories

We propose that the meaning of a text is determined in part by syntactic structures that affect the relative prominence given to the concepts in the text. This proposal was tested in four experiments; the data showed that concepts placed in syntactically prominent positions have increased accessibility in short-term memory during reading and also increased accessibility later in long-term memory. We speculate on how such effects might be understood in terms of current theories of text processing and memory retrieval.

It is often assumed that little or no syntactic information is represented in long-term memory for discourse; once syntactic information has served its purpose of organizing different pieces of information into their relative roles of subject and object, pronoun and antecedent, given and new, and so on, it is quickly forgotten. The generally accepted rule is that memory for the verbatim surface forms of sentences lasts only a few seconds. In contemporary psycholinguistics, this assumption had its roots in demonstrations by Sachs (1967; see also Jarvell, 1971; Caplan, 1972) that only the meaning of sentences is remembered, and the assumption has been incorporated into models of memory for text (cf. Anderson & Bower, 1973; Kintsch, 1974; Kintsch & Van Dijk, 1978). The assumption is still current, as evidenced by the absence of discussion of syntactic structures in recent theoretical work on discourse processes (cf. Kintsch, 1988; McKoon & Ratcliff, 1992a). Despite the fact that syntactic information has been intensively studied within the context of comprehension for single sentences (cf. Boland, Tanenhaus, & Garnsey, 1990; Fodor, 1989; Fodor, in press; Frazier & Rayner, 1982; McKoon, Ratcliff, & Ward, 1993; Rayner & Morris, 1991), its possible role in controlling the semantic interpretation of larger discourse units has received little attention. In this article, we attempt to begin to fill this gap by investigating the role of syntax in determining the relative prominence, or salience, of different parts of a discourse.

Despite the wide acceptance of the idea that syntactic information is not remembered, there have been several empirical demonstrations to the contrary. Keenan (1975) and Anderson (1974) showed relatively long-term memory for the exact wording of sentences read in an experimental situation, and Keenan, MacWhinney, and Mayhew (1977) and Kintsch and Bates (1977) showed such memory for spoken discourse from more natural situations. Begg and Wickelgren (1974) found that syntactic information was not forgotten at a faster

This research was supported by NIDCD Grant R01-DC01240 and AFOSR Grant 90-0246 (jointly funded by NSF) to Gail McKoon and by NIMH Grants HD MH44640 and MH00871 to Roger Ratcliff. We thank Beth Levin for discussions of this work. Address correspondence and reprint requests to Gail McKoon, Psychology Department, Northwestern University, Evanston, IL 60208.

0749-596X/93 $5.00

V32 #4 93@ PAGE: 2 SESS: 3 OUTPUT: Thu May 27 22:40:41 1993
p/jml/05808a/1

rate than semantic information. However, perhaps surprisingly, none of these demonstrations changed the prevailing theoretical view. The reason for this may lie in the (sometimes implicit) belief that memory for surface information resides in a different form or kind of representation than memory for meaning. Putting verbatim surface information in a different kind of memory makes it plausible that it can, on rare occasions like the studies just mentioned, last longer than the usual few seconds, but still have no influence on meaning. This notion of a different kind of memory for surface form was suggested by Kolers (1976; Kolers & Roediger, 1984), who proposed that the procedures with which information is acquired are remembered not as objects in memory but rather are evidenced in facilitation when those same procedures are reexecuted at a later time. The notion of a different kind of memory for surface form is also part of Kintsch's models (van Dijk & Kintsch, 1983; Kintsch, Welsch, Schmalhofer, & Zimny, 1990); in these models, surface information is encoded into a different level of representation from other kinds of discourse information. In this article, we do not take issue with the view that surface information is represented separately. What we do claim is that, in addition to whatever separate memory may exist for surface information, there are also direct effects of syntactic surface information on the representation of meaning.

The generally accepted role of syntactic information is to connect pieces of information together in their syntactically specified roles. Consider the sentences *The student had to clean up his apartment. He crammed his closet with boxes.* Syntactic processes would identify *the student* as subject of the verbs *clean* and *cram*, *student* as the referent of *he*, and perhaps, for the second sentence, *he* as old information and *crammed his closet with boxes* as new information (cf. Chafe, 1976; Clark, 1977). Such connections control meaning in only a minimal way, and they are not represented in the long-term memory representation of a text in most current theories. The same propositions would appear in the long-term memory representation for a variety of different surface structures. For example, the representation of the propositions (*clean, student, apartment*) and (*cram, student, closet, boxes*) would be the same, whether the sentences had been stated as above or as *The apartment had to be cleaned up by the student. He crammed boxes into his closet. We propose in this article that surface form is not always lost in this fashion, but instead can be preserved in the meaning of a text.*

Before proceeding, it should be noted that there is already one, often overlooked, way in which the surface form of sentences in a discourse has been taken to affect memory for meaning in the manner we have in mind. Many researchers use Kintsch's (1974) propositional scheme for representing discourse information, and in that scheme, propositions are ordered in terms of importance relative to a topic proposition. The choice of topic proposition is heavily influenced by surface form aspects of the text: the proposition is usually taken from the main clause of the first sentence in the text, and it usually represents the main verb of that clause and its arguments. Surface form affects the choice of the topic proposition, and that choice, in turn, affects the overall organizational meaning of the other propositions in the text. In short, surface form points to the most salient proposition in the text. What we test in the experiments described below is whether surface form also makes other aspects of the text (that are not the topic proposition) more or less salient.

The proposal that surface syntactic structure interacts with discourse meaning is based in part on current work in linguistics, where the "information packaging" functions of syntactic constructions have been widely studied (Chafe, 1974, 1976; Givon, 1976; Kuno, 1986; Prince, 1978; Wilson & Sperber, 1979; Ward, 1985). In every lan-

JOBNAME: @ V32 #4 93@ PAGE: 3 SESS: 3 OUTPUT: Thu May 27 22:40:41 1993
/xy85/disk4/tsp/jml/05808a/1

guage, speakers have choices about how to convey or package information, and it is a central tenet of studies in functional syntax that these choices are not random. Different syntactic constructions have different discourse functions, and knowing which constructions are appropriate or felicitous or most useful in a given context constitutes part of a speaker's general linguistic competence.

One of the functions often claimed in linguistics for syntactic constructions, the one that is relevant to the research described here, is to vary the relative "status" of the concepts in a discourse. There have been at least two suggestions about how syntax might accomplish this function: within a proposition, differences in relative status might be due to the linking of the arguments of a verb to different syntactic positions, and across propositions, differences in relative status might be due to the assignment of concepts to "foregrounded" versus "backgrounded" syntactic positions.

Within a proposition, the arguments of a verb can be assigned to several different syntactic positions, including subject, direct object, and indirect object. It has been pointed out that an argument may be understood to be more affected by the verb it is it placed in one syntactic position rather than another (cf. Rappaport, Laughren, & Levin; 1987). For example, consider the following two sentences:

1. Bees are swarming in the garden.
2. The garden is swarming with bees.

When *garden* is in the subject position, it is understood to be more affected than when it is in an object position; in other words, it is more likely that the whole garden is swarming with bees with sentence 2 than with sentence 1. Consistent with this intuition, the clause *but most of the garden has no bees in it* is odd when added to the end of sentence 2 but less so when added to sentence 1 (examples from Anderson, 1971). Similarly, in sentences 3 and 4, the entity *wall* is more affected as a direct object than as an indirect object: it is more likely that the whole wall is covered with paint with sentence 4 than with sentence 3. We hypothesize that the more affected a discourse entity is by the action of the verb, as indicated by its syntactic position relative to the verb, the more prominent or salient will be its position in the discourse model. This hypothesis is based on the assumption that, all other things being equal, more affected entities are more central to the meaning of the discourse. Sentence 2 is more likely to be part of a discourse about the garden than sentence 1, and sentence 4 is more likely to be part of a discourse about the wall than sentence 3. It must be stressed that other discourse considerations may override affectedness. In a discourse about insects, we might want to use sentence 1, even though the more affected interpretation of sentence 2 was intended and we would have to continue the sentence with *they fill every corner*. Nonetheless we propose that, in general, entities in positions associated with greater affectedness are more salient.

3. John smeared paint on the wall.
4. John smeared the wall with paint.

Different syntactic positions are also associated with different degrees of prominence when considered in the context of discourse units larger than a single proposition. Pragmatically, a speaker or writer can chose whether to place some specific piece of information in the foreground of a discourse or the background, and the choice is manifested by syntactic structure. Notions of foregrounding have been discussed by many linguists, using a variety of terms to describe distinctions in prominence. Examples most directly related to our research come from Wilson and Sperber (1979). They propose that the syntactic positions of propositions order them in terms of importance, and that the more important a proposition, the more relevant it is to the discourse as a whole. For example, the proposition *admire, I, Bergstrom* is said to have more importance pragmatically in sentence 6 than in sentence 5, and therefore

JOBNAME: @ V32 #4 93@ PAGE: 4 SESS: 3 OUTPUT: Thu May 27 22:40:41 1993
/xy85/disk4:tsp/jml/05808a/1

the proposition is more relevant to its discourse context if it is expressed in sentence 6 instead of sentence 5 (examples from Wilson and Sperber, 1979, p. 305).

5. I have invited Bergstrom, who I admire, to give the opening address.

6. I admire Bergstrom, and I have invited him to give the opening address.

Similarly, Wilson and Sperber point out the reduction in importance associated with a proposition being expressed in a modifying phrase instead of a main clause, as in sentences 7 and 8 where *boring, book* is expressed either as a clause or a modifier.

7. This book is boring, and it is expensive.

8. This boring book is expensive.

The goal of the research described in this article was to test the psychological hypotheses implicit in these linguistic claims. We thought that a reader might use the syntactic position in which a discourse entity is expressed to guide processing for that entity during comprehension. An argument expressed in a more affected position relative to its verb would be perceived as more salient by the reader than an argument in a less affected position, and a proposition in a more important syntactic position would give greater salience to its arguments than a proposition in a less important syntactic position. We hypothesized further that, during reading, more salient entities would be more likely to remain in short-term memory longer for more processing than other entities, and that because of this extra processing, they would be more accessible in the long-term memory representation of the discourse. Experiments 1 through 4 tested these hypotheses.

Experiment 1

George is having second thoughts about his new job.

His critical boss is demanding. or *His demanding boss is critical.*

George is thinking of quitting.

The first sentence of this short discourse introduces George. The second sentence is made up of three propositions: (*his, boss*), (*critical, boss*), and (*demanding, boss*). For the latter two propositions, there is a choice about how to represent them syntactically. Both could be main clauses, or one or the other could be modifying phrases. In the two versions that we used for experiment 1, one modifier was given a main clause position (a predicate modifier) and the other was mentioned as a prenominal modifier. In the first case (. . . *boss is demanding*), *demanding* was given the more prominent syntactic position and in the second case (. . . *demanding boss* . . .), it was given the less prominent syntactic position. We hypothesized that the increased prominence for *demanding* as a predicate modifier would lead to more processing during reading, and therefore more accessibility in short-term memory and/or a longer period of time in short-term memory.

We tested this hypothesis by presenting subjects with short texts like the George paragraph to read. Immediately after each text, a test word was given for recognition. Subjects were instructed to indicate as quickly and accurately as possible whether the test word had or had not appeared in the text. For the George text, *demanding* was tested after the third sentence, and we expected that responses to it would be faster and/or more accurate if the text had mentioned *demanding* in the predicate modifier position as opposed to the prenominal modifier position.

Method

Materials. Each of 24 experimental texts had two versions, with two modifiers switched between the predicate and the prenominal positions in each version, as shown by example above. Each text began with a lead-in sentence (mean length, 7.9 words) and ended with a third sentence (mean length, 7.5 words). The middle sentence was always five words in length: a possessive pronoun or article, followed by a modifier, followed by a noun, followed by a form of the verb *to be*, followed by a mod-

JOBNAME: @ V32 #4 93@ PAGE: 5 SESS: 3 OUTPUT: Thu May 27 22:40:41 1993
/xy85/disk4/tsp/jml/05808a/1

ifier. The two modifiers were both used as test words for the experimental texts. The texts were always displayed in three lines on the CRT screen.

There were two sets of filler texts, each text with one test word. One set of 44 texts averaged 52 words and six lines as presented on the CRT screen; for these texts, 9 had positive test words and 33 had negative test words. The other set of 24 fillers averaged 67 words and five lines on the CRT screen; the test word for each of these was positive.

Procedure. For all four experiments described, all stimuli were presented on a CRT screen, and all responses were collected on the CRT's keyboard. The CRT was controlled by a real-time microcomputer system.

Experiment 1 began with a short list of lexical decision test items, used to give subjects practice with the response keys. After this, six practice filler paragraphs were presented and then the remaining filler paragraphs and the modifier paragraphs were presented in random order. Each paragraph began with an instruction to *Press the space bar* on the CRT keyboard when ready to begin reading. Subjects read the paragraphs one line at a time, pressing the space bar to advance from each line to the next. After the last line, the paragraph was erased from the CRT screen and a single test word was presented. Subjects were instructed to respond as quickly and accurately as possible, pressing the ?/ key if the word had been in the paragraph just read and pressing the z key if it had not. For 44 of the filler texts, a true/false test statement followed the test word. Subjects were instructed to read each paragraph carefully so that they would be able to respond correctly on a true/false test. If the response on the true/false test was incorrect, the word *ERROR* was displayed for 2000 ms. After the test word (and the true/false test if there was one) and a 1000-ms pause, the instruction to press the space bar for the next paragraph was displayed.

Design and subjects. For each of the modifier texts, either the first or the second of the two modifiers was tested (which was designated first and which second was decided arbitrarily), and either the first or the second modifier was presented in the predicate position (the other modifier was presented in the prenominal position). Crossing these two variables resulted in four conditions, which were crossed with groups of subjects (21 per group) and sets of paragraphs (six per set). All cells of the Latin square were not equally represented across subjects (because of constraints on the design of an unrelated experiment involving one of the sets of fillers) so paragraphs were paired for analyses of results (making 12 pairs). A different random order of presentation of the paragraphs was used for each second subject. The 84 subjects participated in the experiment for credit in an Introductory Psychology class.

Results

For all the experiments, means were calculated for each subject and each item in each condition; these means were analyzed by analyses of variance across both subjects and items, $p < .05$.

As predicted, responses were faster and more accurate when the modifier (e.g., *demanding*) was presented in the predicate position (*his critical boss was demanding*), 978 ms and 4% errors, than when it was presented in the prenominal position (*his demanding boss was critical*), 1036 ms and 5% errors. The difference in response times was significant, $F1(1,83) = 11.5$ and $F2(1,11) = 6.0$. One of the two test words (which was labeled first and which was labeled second was arbitrarily designated when the paragraphs were written) had slower response times than the other, by 46 ms. This difference was significant, $F1(1,83) = 4.7$ and $F2(1,11) = 5.1$. However, the predicate position was facilitated over the prenominal position for both test words: the interaction between test word and modifier position was not significant,

Fs < 1.0. The standard error of the mean response times was 10.1 ms. No differences in error rates reached significance, all Fs < 2.4.

Reading times for the sentences containing the modifiers and reading times for the sentences that followed the modifier sentences (the sentences that immediately preceded the test word) did not differ significantly across experimental conditions. The mean reading time for the modifier sentences was 1784 ms (standard error of the mean was 19.0 ms) and the mean reading time for the final sentences was 1739 ms (standard error of the mean was 14.3 ms).

For filler test words, mean response time for correct positive responses was 1255 ms (21% errors) and for correct negative responses, 1083 ms (2% errors). For true test sentences, correct responses averaged 2102 ms (10% errors), and for false test sentences, correct responses averaged 2160 ms (12% errors).

Experiment 2

In Experiment 1, the predicted result was obtained: a modifier presented in a predicate position was more accessible after an intervening sentence than a modifier presented in a prenominal position. This result is consistent with our hypothesis that different syntactic positions are associated with differing degrees of prominence in a discourse, and that these differing degrees of prominence have consequences for how a reader comprehends the discourse. In particular, the result of Experiment 1 suggests that more prominent discourse entities are more accessible in short-term memory during reading or remain longer in short-term memory than less prominent entities.

There is one alternative explanation of the result of Experiment 1 that immediately presents itself, and that is that the predicate modifier is associated with faster response times because it is more recent relative to the test point than the prenominal modifier. In the George paragraph, the prenominal modifier is eight words back from the test point and the predicate modifier is only six words back. However, this alternative would predict that the difference between predicate and prenominal modifiers would appear only in a short-term memory test, not in a long-term memory test. In contrast, our hypothesis that the predicate modifier receives more processing because of its increased salience suggests that the difference should appear on both short-term and long-term memory tests.

We have proposed that discourse entities assigned to different syntactic positions receive different amounts of processing during reading. Most theories of short-term memory assume that the more a concept is processed in short-term memory and the longer it remains in short-term memory, the more likely it is that the concept is encoded into long-term memory (cf. Gillund & Shiffrin, 1984). However, is it not clear whether and how this assumption extends to a concept presented as part of a discourse. While the result of Experiment 1 suggests that a more prominent syntactic position gives more accessibility in short-term memory, it is not clear whether this increased accessibility represents the kind of processing that would increase the probability of representation in long-term memory. As mentioned in the introduction above, it has long been thought that syntactic information is *not* part of long-term memory for discourse

The purpose of Experiment 2 was to test whether a concept associated with a syntactically more prominent position in its discourse was more accessible in the long-term memory representation of the discourse than a concept associated with a less prominent syntactic position. The same texts were used as in Experiment 1, each with two modifiers that could be switched from prenominal to predicate position. Subjects were given a series of study-test lists. For the study phase of each list, they read a number of short paragraphs (all unrelated to each other). For the test phase, they were given a list of single words; for each

JOBNAME: @ V32 #4 93@ PAGE: 7 SESS: 3 OUTPUT: Thu May 27 22:40:41 1993
/xy85/disk4/tsp/jml/05808a/1

word, they were asked to decide, as quickly and accurately as possible, whether it had appeared in any of the paragraphs they had just read. We predicted that responses to a word that had been read in the more prominent predicate position would be faster and/or more accurate than responses to a word from the less prominent prenominal position.

Method

Materials. The modifier texts were the same as those used in Experiment 1, each text with the same two test words. There were 46 filler texts. One set of 32 fillers had a mean length of 49.2 words (averaging 6.2 lines on the CRT screen), and the other set of 14 fillers had a mean length of 29.1 words (always three lines). For each filler text, there were four test words that had appeared in the text. Negative test words were chosen from a pool of 966 words that did not appear in any text.

Procedure. Experiment 2 began with a short list of lexical decision test items, used to give subjects practice with the response keys. After this practice, there were seven study-test list sequences. For the first study list, 10 filler texts were presented. The remaining six study lists each contained four of the modifier texts, four of the longer fillers, and two of the shorter fillers, all presented in random order except that the modifier texts were never in the first or the last two positions of the study list. Each test list was made up of 64 test words, 32 positive words from texts in the immediately preceding study list and 32 negative test words. Except for the first test list, the 32 positive test words included the two modifiers from each modifier text and 4 test words from each filler text in the study list. For each of the modifier texts, one of the modifiers was tested at some point in the test list after the 20th position, and the other modifier was tested at least 10 positions later in the test list. The test position immediately preceding each modifier was filled by a positive test word from one of the filler texts. Otherwise, the positions of test words were chosen randomly.

In designing this experiment, we debated whether the reading time for each text should be controlled by the experimenter or by the subject. Control by the experimenter reduces variability across subjects and items, but control by the subject allows the subject to read at the right rate for whatever level of comprehension the subject adopts as his or her goal. Moreover, reading rate is affected by the degree of accuracy needed for reasonable performance on the test list. Informing subjects each time they make an error increases accuracy, and making feedback aversive (by presenting an error message for a long amount of time, e.g., 2000 ms) should increase accuracy even more. Over the three long-term memory experiments presented in this article, we tried three different combinations of reading time control and accuracy feedback. In Experiment 2, reading time was controlled by the experimenter, and errors were indicated by a 2000-ms error message.

Each study list began with an instruction to press the space bar of the CRT keyboard to initiate the list. Then the texts were presented one at a time, for 10 s for filler texts and for 6 s for modifier texts, with a 1-s blank interval between each text. After the 10th text, a row of asterisks was presented for 2 s to signal the beginning of the test list. Then the test words were presented one at a time. A test word remained on the CRT screen until the subject pressed a response key on the keyboard (?/ for positive responses, z for negative responses). If the response was correct, the next test word appeared after a 50-ms blank interval. If the response was not correct, the word *ERROR* was presented for 2000 ms. Subjects were instructed to respond quickly and accurately.

Design and subjects. For each modifier text, one of the two modifier words was tested first in the test list, and it was studied either in the predicate or the prenominal position. Crossing these two variables re-

sulted in four conditions, all presented as the first test word from their text in the test list. Whichever modifier was not tested first was tested later in the test list, resulting in the same four conditions. For example, for the text about George above, *critical* was tested first in two conditions (studied as predicate and studied as prenominal) and *demanding* was tested first in two conditions (studied as predicate and studied as prenominal). The four conditions for each test word were crossed with four sets of texts and four groups of subjects. Order of presentation of materials was random (except for the constraints mentioned above), different for each second subject. The 28 subjects participated in the experiment for credit in an Introductory Psychology class.

The design of Experiment 2 used both modifiers as test words, but only one of them could be the first to access the representation of the text in long-term memory. In other research, the results obtained at a second test position have been shown to be affected by the first test. Dell, Ratcliff, and McKoon (1981) found that evidence of text structure disappeared at a second test: at that point, all test words from a text had about the same response times and error rates. Thus, for Experiment 2, we expected the first test position to show the effect of syntactic salience, but did not know whether the effect would still be obtained at the second test position.

Results

The prediction was that responses for modifier test words would be facilitated when the modifiers had appeared in their texts in the predicate position relative to the prenominal position. This facilitation was obtained for both test positions: 837 ms vs 903 ms (20% errors in each case) for modifiers tested first in the test list and 863 ms vs 891 ms (16% errors vs 21% errors) for modifiers tested second in the test list. The effect was somewhat smaller for one of the test words than the other, although which was designated the first and which the second had been decided randomly.

Analyses of variance on response times showed the main effect of predicate versus prenominal significant, $F1(1,27) = 6.9$ and $F2(1,46) = 7.1$. The interaction between predicate/prenominal and test word approached significance with items as the random variable, $F2(1,46) = 2.0$, and was significant with subjects as the random variable, $F1(1,27) = 4.9$. Both test words showed facilitation of predicate over prenominal sentence position when they were tested first in the test list; for first test positions, the interaction between test word and predicate/prenominal was not significant when those responses alone were analyzed (F's < 2.2). Why the predicate/prenominal effect diminished for one of the test words in the second test position is not clear (see the discussion of Dell et al., 1981, above). Other response time effects in the experiment were not significant (F's < 1.1), except that the effect of test position in the items analysis approached significance, $F(1,46) = 2.7$. The standard error of the response time means was 24 ms. For error rates, none of the main effects or interactions approached significance. Mean response time for positive fillers was 862 ms (24% errors) and mean response time for negative fillers was 976 ms (30% errors).

Experiment 3

Experiments 1 and 2 were designed to test whether the prominence associated with a modifier in a predicate position led to increased accessibility immediately after a discourse was read and whether it led to increased accessibility in the long-term memory representation of the discourse. Both effects were obtained. In Experiment 1, increased syntactic prominence was confounded with recency, but recency should affect only the test of short-term memory. Because the prominence effect was also obtained in the test of long-term memory, recency is probably not the explanation of the result from Experiment 1. Instead, we attribute the results of both Experiments 1 and 2 to syntactically determined salience.

However, the syntactic prominence of

JOBNAME: @ V32 #4 93@ PAGE: 9 SESS: 3 OUTPUT: Thu May 27 22:40:41 1993
/xy85/disk4/tsp/jml/05808a/1

the predicate position was confounded with another, simple variable: the modifier in the predicate position was always the last word of its sentence. The results of Experiments 1 and 2 may reflect, not syntactic prominence, but instead prominence associated with the last word of a sentence as compared with other words in the middle parts of a sentence. In Experiment 3, we eliminated this confound by adding adjunct phrases to the ends of the modifier sentences. The sentences about George became:

George is having second thoughts about his new job.

His critical boss is demanding at times. or *His demanding boss is critical at times.*

With the adjunct phrase added, neither the predicate nor the prenominal modifier appears at the end of the second sentence. Experiment 3 was also designed to generalize the results of Experiment 2 by changes in procedure: The reading time for the texts was controlled by the subjects, not the experimenter, and less emphasis was placed on the accuracy of responses in the test list.

Method

Materials. The same materials were used as in the preceding experiments except that an adjunct phrase was added to the end of each second sentence of the direct object–indirect object texts, as shown above for the George text. The number of words in the adjunct phrases varied from two to four.

For both the modifier texts and the filler texts, only the first two sentences of each text were used in this experiment. Subjects had found Experiment 2 very difficult, and we thought that reducing the length of the texts would make it easier. There was a pool of 66 filler paragraphs, each with two lines as displayed on the CRT screen, averaging 20 words in length. There were two positive test words for each paragraph. Negative test words were drawn from a pool of words that did not appear in any text, the same pool as in Experiment 2.

Procedure, design, and subjects. The procedure and design were almost the same as those for Experiment 2; there were only the following differences: The study lists each contained four of the modifier texts and eight filler texts. Each test list was made up of 40 test words, 20 positive and 20 negative. For each of the modifier texts, one of the modifiers was placed at some point in the test list after the eighth position, and the other modifier was placed at least eight positions later. Subjects controlled the reading time for each text by pressing the space bar when they had finished reading each text. There was a 1-s blank interval after each text. In the test list, if a response was not correct, the word *ERROR* was presented for 500 ms (as compared to 2000 ms in Experiment 2). The 24 subjects participated in the experiment for credit in an Introductory Psychology class.

Results

It was predicted that response times for a modifier test word would be faster when the modifier had been presented in the predicate position, even though the predicate position was not the last word of its sentence. This is the result that obtained, but only for the first test position. Because the results were different at the two test positions, we analyzed them separately.

At the first test position, response times for predicate modifiers averaged 733 ms (4% errors) and response times for prenominal modifiers averaged 780 ms (5% errors). This difference was significant, $F1(1,23) = 4.7$ and $F2(1,20) = 7.3$. The effect of which of the two words was tested and the interaction of predicate/prenominal and test word were not significant, F's < 2.2. The standard error of the response time means was 21.2 ms. There were no significant effects on error rates, F's < 1.0.

At the second test position, the standard error of the response times means, 31 ms, was much greater than at the first test position. This larger standard error may have contributed to the failure to find an effect of predicate versus prenominal study position in the second test position. Response times

for predicate modifiers averaged 786 ms (5% errors) and response times for prenominal modifiers averaged 792 ms (7% errors). In the subjects' analysis, one test word was responded to more quickly than the other, $F1(1,23) = 4.7$, but the effect was not significant in the items' analysis, $F2(1,20) = 2.2$. The main effect of predicate/prenominal position and the interaction of predicate/prenominal and test word were not significant, F's less than 2.5.

The mean reading time for the two-sentence modifier texts was 4916.5 ms, with a standard error of 194 ms. The mean response time for positive filler test words was 798 ms (8% errors) and for negative test words, it was 1066 ms (59% errors). Note that subjects had a strong bias to respond *yes*, which led to fast *yes* responses and a high error rate for negative test words. Nevertheless, for the first test position, the predicate/prenominal variable still had a significant effect.

EXPERIMENT 4

Experiments 1 through 3 show the effect of syntax on the relative accessibilities of different propositions. The proposition that George's boss is demanding can be made more or less accessible by moving it from one syntactic position (main clause predicate) to another (prenominal modifying phrase). Experiment 4 examined a second syntactic effect, the relative salience associated with the different syntactic positions to which the arguments of a verb can be assigned.

The librarian was furious when she got to work today.

Somebody had inserted some magazines inside some newspapers late last night.

or

The librarian was furious when she got to work today.

Somebody had inserted some newspapers inside some magazines late last night.

In this text, the proposition with the verb *insert* has three arguments: *somebody*, *magazines*, and *newspapers*. In one version, *magazines* is linked to the direct object position, and in the other, it is linked to the indirect object position. In the introduction to this article, we reviewed the linguistic notion that an entity in the direct object position is taken to be more affected by the verb, and we suggested that more affected entities were associated with greater prominence. Greater prominence, in turn, we hypothesized to be associated with greater accessibility in the mental representation of a text.

In Experiment 4, we used texts like the one above about the librarian. Subjects were given a series of study-test lists, as in Experiments 2 and 3, and the direct and indirect objects (*magazines* and *newspapers*) were presented for recognition in the test lists. We predicted faster and/or more accurate responses for the objects when they had appeared in the direct object position than in the indirect object position. For the librarian text, *magazines* would have faster and/or more accurate responses with the first version of the second sentence than the second version. Each of the object sentences ended with an adjunct phrase so that the indirect object was never the final word of its sentence.

Method

Materials. There were 28 paragraphs each with two objects that could be switched between the direct object and the object of preposition positions. Each paragraph began with a lead-in sentence (these averaged 8.75 words) and then continued with a sentence containing the two objects (averaging 10.71 words). This sentence had the form: subject noun phrase, verb, object noun phrase, prepositional phrase, adjunct phrase. The two objects were used as test words. These paragraphs were displayed in two lines on the CRT screen. The same filler paragraphs and pool of negative test words were used as in Experiment 3.

Procedure, design, and subjects. The experiment differed from Experiment 2 only in the following respects: Each of seven study lists contained four of the objects texts and eight filler texts. Each test list was

made up of 40 test words, 20 positive words from texts in the immediately preceding study list and 20 negative words that had not appeared in any studied text. For each of the object texts, one of the objects was tested at some point in the test list after the eighth position, and the other object was tested at least eight positions later in the test list. Subjects controlled the reading time for each text by pressing the space bar when they had finished reading each text. There was a 1-s blank interval between each text. If a response to a test word was not correct, the word *ERROR* was presented for 2000 ms, as in Experiment 2. The 32 subjects participated in the experiment for credit in an Introductory Psychology class.

Results

As predicted, responses for object test words were faster when the object had been presented in its text as a direct object than when it had been the object of a prepositional phrase. The facilitation for the direct object was apparent when the object was tested at the first test position in the test list: response times were 679 ms (7% errors) versus 704 ms (6% errors); and when it was tested in the second test position: 661 ms (5% errors) versus 683 ms (4% errors). The amount of facilitation was significant, $F1(1,31) = 6.3$ and $F2(1,27) = 4.6$. The amount of facilitation did not interact either with test position or with which of the two object words was tested, F's < 1.3. Responses for the second test position were faster than for the first, approaching significance, $F1(1,31) = 3.1$ and $F2(1,27) = 3.6$, and the interaction of test position and test word was significant, $F1(1,31) = 5.4$ and $F2(1,27) = 4.1$ (although which test word was designated first vs second had been decided randomly). Standard error of the response time means was 18.8 ms. The only significant effect for error rates was that there were more errors in the first test position, $F1(1,31) = 4.2$ and $F2(1,27) = 4.3$.

Reading times for the two-sentence object texts averaged 5104 ms with a standard error of the mean of 89.8. Responses on positive filler test words averaged 728 ms (6% errors), and responses on negative filler test words averaged 974 ms (49% errors).

General Discussion

The experiments presented in this article were designed from a theoretical view of text processing by which syntactic information is assumed to influence the relative salience of different pieces of text information during reading, and in so doing, helps to determine how much attention is given to different pieces of information. More attention for some concept or proposition translates, we assume, into more processing for a longer period of time in short-term memory.

The experiments presented here test the first and most immediate consequences of this theoretical view. The parts of a text that are expressed in more salient syntactic positions should be more available immediately after they are read, and they should be more accessible in the long-term memory representation of the text. In the first three experiments, we manipulated whether a proposition was placed in a syntactic position of greater prominence—a main clause—or lesser prominence—a modifying phrase. The modifier in the more prominent position was more available immediately after reading, and it was also more accessible in long-term memory. In Experiment 4, we manipulated whether an argument of a verb was placed in the direct object position or an indirect object position, and, as predicted, arguments in the direct object position were more accessible. Like the results of Experiments 1 through 3, this result points to the role of syntax in guiding discourse processing. It also provides experimental evidence to support the linguistic claims about the different degrees of affectedness associated with different syntactic positions for the arguments of a verb.

While differences in accessibility are the most immediate consequences of syntactic variables, the most important conse-

JOBNAME: @ V32 #4 93@ PAGE: 12 SESS: 3 OUTPUT: Thu May 27 22:40:41 1993
/xy85/disk4/tsp/jml/05808a/1

quences may be those that result more indirectly from the extra short-term memory processing given to more prominent pieces of information. Extra processing may affect how the text information is organized and what information is included in the final representation of meaning that is eventually constructed for the text. How this would be accomplished is easy to speculate about (see below), given current models of text processing. But, first, we should consider the sizes of the effects in our experiments.

We need to consider whether the results of our experiments are an example, to put it metaphorically, of the cup being half full or half empty. So far, we have emphasized that the experiments did in fact produce the results that were predicted. However, the effects were small. Across the three long-term memory experiments, the response time differences between syntactically more and less prominent test words were 66, 47, and 25 ms on a baseline of 700–900 ms (for first test positions in Experiments 2, 3, and 4, respectively). Are these effects big enough that a large theoretical structure can be built upon them? Of course, the answer is that we don't know. However, certainly when we speculate theoretically about syntax in discourse processing, the size of the effects should constrain our thinking.

A theory about the role syntax might play in discourse processing can be constructed out of two kinds of already existing models: Kintsch's model (1988) for the processing of propositions and the compound cue models for memory access (Dosher & Rosedale, 1989; Ratcliff & McKoon, 1988; McKoon & Ratcliff, 1992b; Ratcliff & McKoon, 1993). First, consider Kintsch's model for how propositions are processed through short-term memory and encoded into long-term memory. Givon (in press) has proposed that "grammatical devices" are signals that trigger mental operations; he views grammatical signals as "mental processing instructions." This idea can be made concrete in Kintsch's model in order to show how syntactic prominence could come to influence the organization of the propositions of a text. In the model, propositions are processed in cycles. On each cycle, some number of propositions is input to the processing system, where they are connected to each other by argument repetition (i.e., any two propositions that share a common argument are connected to each other). The only connections that are made (without searches of long-term memory) are those between propositions that are in short-term memory at the same time. At the end of a cycle, all but a small subset of the propositions in short-term memory are transferred to long-term memory, and a new cycle with new input propositions begins. Currently, the model chooses which propositions to keep in short-term memory from one cycle to the next according to how closely they are connected to the original topic of the text and how recently they were mentioned in the text. However, it would be straightforward to change the model so that concepts in more prominent syntactic positions were preferentially maintained in short-term memory from one cycle to the next. Preferential maintenance would then allow them to be connected to propositions in the next input cycle, creating connections that would not otherwise be formed. Thus, simply holding syntactically salient information longer in short-term memory (through extra processing cycles) could create an organization of the propositions that would be influenced by syntactic salience. Holding salient information longer would also predict the results of our experiments: a more salient concept would be more accessible a sentence after it was mentioned than a less salient concept (Experiment 1), and a more salient concept would be more strongly represented in long-term memory (Experiments 2, 3, and 4) because it would have had more time to accumulate strength of encoding into long-term memory and/or more time to build its strength of connections to other encoded items (cf. Gillund & Shiffrin, 1984).

JOBNAME: @ V32 #4 93@ PAGE: 13 SESS: 3 OUTPUT: Thu May 27 22:40:41 1993
/xy85/disk4/tsp/jml/05808a/1

It is not only plausible that the organization of the propositions in the final representation of a text would be affected by holding syntactically prominent propositions over from one cycle to the next, but also consistent with other current results. Kintsch (1992) has simulated the effects of adding syntactic preference rules to his model, and the final organization produced by the model does, in fact, change when the rules are added. There is also one empirical finding that is consistent with the notion that syntactic salience affects how propositions are connected together. McKoon, Ward, Ratcliff, and Sproat (in press; see also Ward, Sproat, & McKoon, 1991) examined syntactic salience and pronominal reference with texts from which 1 and 2 below are taken:

1. . . . *lately he's taken up deer hunting. He thinks that they are really exciting to track.*

2. . . . *lately he's taken up hunting deer. He thinks that they are really exciting to track.*

In the second sentences of both examples, the pronoun *they* is intended to refer to *deer*. In the first sentence of 1, *deer* is placed in a modifier position and in the first sentence of 2 it is the object of the verb *hunting*. As indicated by the results of the experiments above, the modifier position should be less prominent and so should make *deer* less salient. In terms of cycles of propositions through short-term memory, decreased salience translates into lower probability of staying in short-term memory. So if a cycle ends after the first sentence of these examples, *deer* will be less likely to be in short-term memory for the beginning of the second sentence in example 1 than in example 2. As a result, understanding the referent of *they* will be more difficult in the first example than the second. This prediction was confirmed by McKoon et al.'s experiments (in press): reading times for the second sentences were longer for the first example than the second, consistent with pronoun resolution taking more time in the first example than the second (see McKoon et al. for experiments that rule out a number of alternative explanations for this result).

The plausibility of the idea that syntactic prominence contributes to preferential maintenance of propositions in short-term memory, as well as the results of Kintsch's (1992) simulations and McKoon et al.'s experiments, all point to the effects of syntactic variables on the long-term memory organization of text information. However, the organization of the propositions given by a text is not the only part of text processing that might be influenced by preferential maintenance in short-term memory. Preferential maintenance might also allow propositions and concepts to be combined in short-term memory in ways that they otherwise might not be, and therefore allow them to form cues for memory retrieval that would not otherwise be formed. Compound cue models of memory retrieval (Dosher & Rosedale, 1989; Ratcliff & McKoon, 1988, 1993) based on the global memory models (e.g., Gillund & Shiffrin, 1984; Hintzman, 1988; Murdock, 1982) claim that a familiar relation between two or more concepts is recognized if and only if the concepts are in short-term memory at the same time. Being in short-term memory at the same time means that the concepts form a compound cue with which they can jointly access memory. For example, the familiar relation between *green* and *grass* would be apparent if they were near enough together in a text that they could be in short-term memory at the same time (see Foss & Speer, 1991, for a discussion similar to this one). In traditional lexical decision priming experiments, words like *green* and *grass* are presented in lists of single words, and the facilitation given by *green* to *grass* is observed only if *grass* immediately follows *green* or they are separated by only one or two other items (McNamara, 1992; Ratcliff, Hockley, & McKoon, 1985; Ratcliff & McKoon, 1978; 1988, 1993). This indicates that, for a list of single items, the compound

cue for memory retrieval contains only two or three of the most recent words. But if the words are not just a list of unrelated concepts but instead form a text, then the compounds for memory retrieval will almost certainly be different. They may contain concepts, semantic propositions, the verbatim words of the text, and so on (see Ratcliff & McKoon, 1988), and which of these are held from one processing cycle to the next will not be determined only by recency, but also by how closely a concept or proposition is connected to the text's topic and, we suggest, by how prominent the concept or proposition is in the syntactic structure of the text. If *green* is placed in a syntactically prominent enough position, it may still be in short-term memory when *grass* is read, even if *grass* appears many words later in the text. The relation between *green* and *grass* that was thus made apparent could potentially change how the text was understood, and so change the encoded meaning of the text.

The syntactic effects on text processing that we have demonstrated in the experiments reported here are small. Concepts linked to syntactically more prominent positions were more accessible in both short-term and long-term memory tests, but not dramatically so. In this discussion, we have speculated that even these small effects might have powerful consequences for the organization and content of the mental representation of discourse. Syntactic "mental processing instructions" (Givon, in press) might, for some pieces of information, mean a little more time spent in short-term memory, and allow a little extra processing, and whether that means a lot for comprehension of a text as a whole is a subject for further research.

References

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

ANDERSON, J. R., & BOWER, G. H. (1973). *Human associative memory*. Washington, DC: Winston.

ANDERSON, S. R. (1971). On the role of deep structure in semantic interpretation. *Foundations of Linguistics*, 7, 387–396.

BEGG, I., & WICKELGREN, W. (1974). Retention functions for syntactic and lexical vs. semantic information in sentence recognition memory. *Memory & Cognition*, 2, 353–359.

BOLAND, J. E., TANENHAUS, M. K., & GARNSEY, S. M. (1990). Evidence for the immediate use of verb control information in sentence processing. *Journal of Memory and Language*, 29, 413–432.

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

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

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

CLARK, H. H. (1977). Inferences in comprehension. In D. LaBerge & S. J. Samuels (Eds.), *Basic processes in reading: Perception and comprehension*. (pp. 243–264). Hillsdale, NJ: Erlbaum.

DELL, G. S., RATCLIFF, R., & MCKOON, G. (1981). Study and test repetition effects in item recognition priming. *American Journal of Psychology*, 94, 497–511.

DOSHER, B. A., & ROSEDALE, G. (1989). Integrated retrieval cues as a mechanism for priming in retrieval from memory. *Journal of Experimental Psychology: General*, 2, 191–211.

FODOR, J. D. (1989). Empty categories in sentence processing. *Language and Cognitive Processes*, 4, 155–209.

FODOR, J. D. (in press). Processing empty categories: A question of visibility. *Language and Cognitive Processes*.

FOSS, D., & SPEER, S. (1991). Global and local context effects in sentence processing. In R. Hoffman & D. Palermo (Eds.), *Cognition and the symbolic processes: Applied and ecological perspectives*, Hillsdale, NJ: Erlbaum.

FRAZIER, L., & RAYNER, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14, 178–210.

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

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

GIVON, T. (in press). The grammar of referential coherence as mental processing instructions. *Cognitive Science*.

HINTZMAN, D. (1988). Judgments of frequency and

recognition memory in a multiple-trace memory model. *Psychological Review*, 95, 528–551.

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

KEENAN, J. M. (1975). *The role of episodic information in the assessment of semantic memory representation for sentences*. Unpublished doctoral dissertation, University of Colorado, Boulder.

KEENAN, J. M., MACWHINNEY, B., & MAYHEW, D. (1977). Pragmatics in memory: A study of natural conversation. *Journal of Verbal Learning and Verbal Behavior*, 16, 549–560.

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

KINTSCH, W. (1988). The role of knowledge in discourse comprehension: A construction–integration model. *Psychological Review*, 95, 163–182.

KINTSCH, W. (1992). How readers construct situation models for stories: The role of syntactic cues and causal inferences. In A. F. Healy, S. Kosslyn, & R. M. Shiffrin (Eds.), *Essays in honor of William K. Estes*. Hillsdale, NJ: Erlbaum.

KINTSCH, W., & BATES, E. (1977). Recognition memory for statements from a classroom lecture. *Journal of Experimental Psychology: Human Learning and Memory*, 3, 150–159.

KINTSCH, W., & VAN DIJK, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85, 363–394.

KINTSCH, W., WELSCH, D., SCHMALHOFER, F., & ZIMNY, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language*, 29, 133–159.

KOLERS, P. A. (1976). Remembering a year later. *Journal of Experimental Psychology: Human Learning and Memory*, 2, 554–565.

KOLERS, P. A., & ROEDIGER, H. L. (1984). Procedures of mind. *Journal of Verbal Learning and Verbal Behavior*, 23, 425–449.

KUNO, S. (1986). *Functional syntax: Anaphora, discourse, and empathy*. Chicago, IL: Chicago Univ. Press.

MCKOON, G., & RATCLIFF, R. (1992a). Inference during reading. *Psychological Review*, 99, 440–466.

MCKOON, G., & RATCLIFF, R. (1992b). Spreading activation versus compound cue accounts of priming: Mediated priming revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 1155–1172.

MCKOON, G., RATCLIFF, R., & WARD, G. (1993). *Testing theories of reading: An empirical investigation of the on-line lexical decision task*. Manuscript submitted for publication.

MCKOON, G., WARD, G., RATCLIFF, R., & SPROAT, R. (in press). Morphosyntactic and pragmatic factors affecting the accessibility of discourse entities. *Journal of Memory and Language*.

MCNAMARA, T. P. (1992). Theories of Priming: I. Associative distance and lag. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 1173–1190.

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

PRINCE, E. (1978). A comparison of *wh*-clefts and *it*-clefts in discourse. *Language*, 54, 883–906.

RAPPAPORT, M., LAUGHREN, M., & LEVIN, B. (1987). *Levels of lexical representation*. Lexicon project working papers, MIT.

RATCLIFF, R., HOCKLEY, W. E., & MCKOON, G. (1985). Components of activation: Repetition and priming effects in lexical decision and recognition. *Journal of Experimental Psychology: General*, 114, 435–450.

RATCLIFF, R., & MCKOON, G. (1978). Priming in item recognition: Evidence for the propositional structure of sentences. *Journal of Verbal Learning and Verbal Behavior*, 17, 403–417.

RATCLIFF, R., & MCKOON, G. (1988). A retrieval theory of priming in memory. *Psychological Review*, 95, 385–408.

RATCLIFF, R., & MCKOON, G. (1993). *Retrieving information from memory: Spreading activation theories versus compound cue theories*. Manuscript submitted for publication.

RAYNER, K., & MORRIS, R. K. (1991). Comprehension processes in reading ambiguous sentences: Reflections from eye movements. In G. Simpson (Ed.), *Understanding word and sentence*. Amsterdam: North Holland Press.

SACHS, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics*, 2, 437–442.

VAN DIJK, T. A., & KINTSCH, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

WARD, G. L. (1985). *The semantics and pragmatics of preposing*. Philadelphia: Univ. of Pennsylvania dissertation. Reprinted in 1988 New York: Garland.

WARD, G., SPROAT, R., & MCKOON, G. (1991). A pragmatic analysis of so-called anaphoric islands. *Language*, 67, 439–474.

WILSON, D., & SPERBER, D. (1979). Ordered entailments: An alternative to presuppositional theories. *Syntax and Semantics*, 11, 299–323.

(Received August 5, 1992)

(Revision received February 1, 1993)

Journal of Experimental Psychology:
Learning, Memory, and Cognition
1992, Vol. 18, No. 6, 1155-1172

0278-7393/92/$3.00

Spreading Activation Versus Compound Cue Accounts of Priming: Mediated Priming Revisited

Gail McKoon and Roger Ratcliff
Northwestern University

Spreading activation theories and compound cue theories have both been proposed as accounts of priming phenomena. According to spreading activation theories, the amount of activation that spreads between a prime and a target should be a function of the number of mediating links between the prime and target in a semantic network and the strengths of those links. The amount of activation should determine the amount of facilitation given by a prime to a target in lexical decision. To predict the amount of facilitation, it is necessary to measure the associative links between prime and target in memory. Free-association production probability has been the variable chosen in previous research for this measurement. However, in 3 experiments, the authors show priming effects that free-association production probabilities cannot easily predict. Instead, they argue that amount of priming depends on the familiarity of the prime and target as a compound, where the compound is formed by the simultaneous presence of the prime and target in short-term memory as a test item.

An important function of memory is to provide the information necessary for an integrated understanding of the various objects that we encounter. People, words, and objects do not occur in isolation; rather, they occur in some larger context, and memory must provide the means of integrating the individual parts into the unified context. Memory processes use multiple cues to focus on some relevant subset of the vast amount of information in memory. For example, *housewives* in the context of *children* evokes a different set of information than *housewives* in the context of *careers*, or *housewives* in the context of *linoleum* (Light & Carter-Sobell, 1970; Tulving & Thomson, 1973). Currently, two classes of theories have been proposed to explain how focusing is accomplished: spreading activation theories and compound cue theories. In this article, we show that one set of published data (McNamara & Altarriba, 1988), claimed to be consistent only with spreading activation theories, can also be accommodated by compound cue theories.

Spreading activation is assumed to work within a semantic memory network. The network consists of a set of interconnected nodes, with each node representing a concept. Nodes are connected to each other if they are related by prior association (*baby–mother*), if they have been recently studied together (*baby–concrete* in the sentence *The baby hit the concrete*), or if they share semantic features. When a concept is presented to the system, activation of the node representing the concept is increased, and activation spreads through the network, temporarily increasing the activation of nearby concepts. The amount of activation given to nearby concepts is a function of the distance between them and the input concept, or the relative strengths of the links between them and the input, or both. It is this spread of activation that leads to focusing on information relevant to the input. This process also accounts for the phenomenon of priming, whereby presentation of one item—a prime—facilitates responses to a subsequent, related item—the target.

Compound cue theories have recently been proposed by Ratcliff and McKoon (1988) and Dosher and Rosedale (1989). The mechanism by which focusing is said to occur in a compound cue theory is very different from that proposed by spreading activation. There is no temporary activation of information in the long-term memory system. Instead, items presented to the system are assumed to join together in short-term memory to form a compound cue. This compound cue is assumed to have some degree of familiarity, where familiarity is determined by the strengths of associations between the compound in short-term memory and items in long-term memory. The familiarity value is assessed by direct access to a composite long-term memory or by parallel comparisons to all items in long-term memory (depending on specific global memory model implementation). In the compound cue view, focusing is accomplished by means of a matching process that matches compounds formed from items that co-occur in short-term memory against all the items in long-term memory. Priming phenomena are consistent with compound cue theories because a response to the second of two items in a compound will be facilitated by a high familiarity value for the compound. What determines the value of familiarity depends on the task. For recognition, the global memory models spell out in detail how familiarity is computed from factors involved at encoding (i.e., the probability that features

This research was supported by National Science Foundation (NSF) Grant 85-16350, National Institute of Deafness and Other Communicative Diseases Grant R01-DC01240, and Air Force Office of Scientific Research Grant 90-0246 (jointly funded by NSF) to Gail McKoon and by National Institute of Mental Health Grants HD MH44640 and MH00871 to Roger Ratcliff. We thank Tim McNamara for extensive discussions of this work and for providing the lists of stimuli used in Experiment 1. We also thank Mark Seidenberg for words used in the naming latency experiment. For Experiment 3, Ken Church was extremely generous with his ideas and his time in providing us with a large list of possible stimuli and their associated statistics, and we are very grateful.

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

of an item are encoded or that strength of the item is built up). In lexical decision, familiarity would be based on other factors such as preexperimental familiarity, frequency, learned associations (McKoon & Ratcliff, 1979, 1989), and semantic relatedness or association.

The compound cue mechanism can be implemented in a number of current memory models (Gillund & Shiffrin, 1984; Grossberg & Stone, 1986; Hintzman, 1986; Murdock, 1982). The key to all the implementations is a boost in the familiarity value for a compound when items in the compound are mutually associated in long-term memory. For example, in an implementation of Hintzman's or Murdock's models, associated pairs of items (for two-item compounds) are stored in a single vector or convolution of two vectors, respectively (see Ratcliff & McKoon, 1988). If a prime–target probe matches a stored pair, the value of match will be much larger than if the probe pair partly matches different pairs (e.g., if A–B is stored, then the probe A–B will have a high degree of match; the probes A–C and D–B will have much lower degrees of match). In Hintzman's model, this is because the degree of match involves a cubing operation, and in Murdock's model, a partial match (A–B with A–C) of a convolution is no better than a match between unrelated pairs. The Gillund–Shiffrin model differs from Hintzman's and Murdock's models in that the degree of match for a compound depends both on direct associations in memory between the two words in the compound and on associations between the two words and one intermediate concept (but only such two-step associations, not more than two). Multiplication of the strength of association of the words in the compound with their mutually associated concepts in memory gives the nonlinearity required to boost the match value.

Because priming phenomena have been such a major source of evidence for the spreading activation mechanism, they have provided the grounds for confrontation between spreading activation and compound cue theories. Ratcliff and McKoon (1988) summarized a number of priming effects and their explanations in terms of each class of theory. For example, they showed that both spreading activation and compound cue theories can account for automatic and strategic priming processes, empirical characteristics of the temporal onset of priming, effects of neutral primes, forward and backward priming effects, and priming of ambiguous words. More telling were comparisons between the theories' accounts of the decay function for priming effects and of the range of priming effects.

Decay of priming refers to the finding that, as other test items intervene between prime and target, the amount of facilitation on the target is reduced. According to compound cue theories, decay must occur rapidly because the effect of an earlier prime must be small and must get smaller as the prime is less likely to be included in the compound and weighted less in calculating familiarity. Thus, for the compound cue mechanism, decay is a function of items intervening between prime and target in short-term memory. Spreading activation, on the other hand, is not affected by the contents of short-term memory (but see ACT*; Anderson, 1983). Activation decays as a function of time, and the rate is a free parameter, constrained only post hoc by empirical data. Ratcliff and McKoon (1988) tested these two views of decay against each other. In their experiments, the time delay between an associated prime and target was held constant, and the variable was whether a third, unrelated item intervened between them. By the spreading activation hypothesis, the intervening item should have had no effect on the level of activation of the target, and so no effect on the amount of priming from the prime to the target. But, in fact, the intervening item did reduce the priming effect, as predicted by a compound cue mechanism in which the intervening item would "bump" the prime out of the compound in short-term memory.

The range of priming is defined as the number of concepts across which priming should occur. For example, consider a story that is made up of a number of propositions connected in a linear fashion such that each proposition is directly connected only to the proposition that occurs temporally before it and the proposition that occurs temporally after it (Ratcliff & McKoon, 1988). According to spreading activation theories, input of a concept from one of the propositions should give rise to activation spreading from the input concept through the temporal chain to concepts in the other propositions. The amount of activation at any one proposition will be a function of its distance from the input concept (see Ratcliff & McKoon, 1981, for discussion of the temporal dynamics of this process). The maximum distance at which there will still be significant amounts of activation is not determined by any intrinsic assumption of the spreading activation theories but instead is a post hoc parameter set to account for available data. In contrast, for the compound cue mechanism, the range of priming effects is completely constrained by the architectures of the models in which the mechanism is implemented. In the Gillund–Shiffrin implementation (1984), priming between two concepts can occur only if they are directly connected to each other or if they are separated by no more than one intervening concept. In implementations with Hintzman's model (1986) or with Murdock's model (1982), the two concepts must be directly connected. When Ratcliff and McKoon (1988) tested the range of priming, they found results in accord with the compound cue mechanism. Using concepts from linearly structured stories, they found a strong priming effect when the prime and target concepts were directly connected or separated by only one concept. But priming effects were at a minimum when the prime and target were separated by only four other concepts, and the priming effect was no larger for four intervening concepts than for six.

Both the decay of priming and range of priming functions provide tests that could have potentially falsified the compound cue theories. But empirical results did not falsify these theories; results were exactly as predicted by the compound cue mechanism. However, the results can also be explained by spreading activation theories as long as parameters of those theories are set to accommodate the data. Thus, although compound cue theory has been subjected to more stringent tests than spreading activation, both the compound cue and spreading activation mechanisms are still viable hypotheses.

The purpose of this article is to address another empirical test of the range of priming, a test that has been claimed to

show support for spreading activation theories over compound cue theories. The finding has been labeled "mediated priming." A mediated prime–target pair is a pair of words assumed to be connected in memory not directly but only via a third concept. Priming would be said to occur for a mediated pair if the response to the target were facilitated by the prime (where priming is usually measured in lexical decision response times). Mediated priming is claimed to be problematic for (some) compound cue theories because these theories predict that facilitation will occur only when the relation between prime and target is direct, not when it is mediated. In this article, we challenge this claim by arguing that mediated primes and targets are actually directly (although weakly) related.

In previous research designed to support spreading activation theories, mediated priming effects have been predicted from free-association production probabilities. The assumption has been that the amount of facilitation given by a prime to a target can be predicted by the probability that the prime will produce the target (directly or indirectly) in free association. This assumption is explicit in the experimental work of de Groot (1983), Balota and Lorch (1986), and McNamara and Altarriba (1988). For example, if *animal* is produced as a free associate of *deer* with a high probability, then *animal* would be said to be directly associated to *deer*, and *deer* should facilitate responses to *animal*. For indirect associations, a prime is said to be connected to a target via a mediator if the mediator is produced as an associate of the prime, the target is produced as an associate of the mediator, and the target is *not* produced as an associate of the prime. *Deer* and *vegetable* would be said to be mediated if *deer* produced *animal* in free association and *animal* produced *vegetable*, but *deer* did not produce *vegetable*. By spreading activation views, the prime of a mediated pair (*deer*) should facilitate a lexical decision on the target (*vegetable*) via activation spreading among the prime, mediator, and target (although the amount of facilitation would be reduced because the prime and target are not directly connected). Reliance on free association to predict priming effects was stated explicitly by Balota and Lorch (1986): "If the mediated target does not occur across associates given either within a subject or across subjects, then it is highly unlikely that there is a direct association from the mediated prime to the mediated target" (p. 338).

We take this logic (or definition) one step further. If a target does not occur across associates to the prime, and it does not occur across associates of associates of the prime, then it is highly unlikely that there is a mediated association between the prime and target. And if there is no direct or mediated association, then according to spreading activation theories, there should be no facilitation from prime to target. It is critical to note that Balota and Lorch's statement is the only statement we have been able to find that provides an explicit empirical method for determining mediation. No method other than free association has been suggested for finding out whether pairs are mediated or not (except intuition).

We show that, in fact, there is facilitation for pairs of words that fulfill the conditions of no direct or mediated associations. Two conclusions can follow from this demonstration. Either spreading activation accounts of priming are wrong, or free association does not provide an infallible index of associative links in memory. If free association does not provide an infallible index, then it may be that all pairs of words that exhibit priming are actually directly connected in memory (with various degrees of strength), and contrary to previous claims, findings of mediated priming are fully consistent with compound cue theories because they are actually demonstrations of direct priming.

We took as the starting point for our experiments nonmediated prime–target pairs—pairs for which we thought the prime and target should be weakly and directly associated but for which the target would not be produced in free association either as a response to the prime or as a response to any associate of the prime. For these pairs, we used as primes words that were primes in Balota and Lorch's materials. *Deer–grain* is an example. *Grain* is not strongly associated to *deer*; *grain* is not produced as a response to *deer* in free association. But *deer* and *grain* are likely to be (weakly) directly associated because grain is something deer can eat. From the compound cue theories, we predicted that weakly and directly associated pairs of words would show small but significant priming effects. The priming effects depend on the weak direct association in long-term memory that is cued by the presence of both words of the pair in the compound formed in short-term memory. It is the simultaneity of their presence in short-term memory that gives rise to a high value of familiarity. From the reasoning used in previous tests of mediated priming (e.g., Balota & Lorch, 1986), these nonmediated pairs should not exhibit priming because free association shows no connection between the prime and target.

In the first experiment, we used pairs of two types. The pairs of the first type (previously used by McNamara & Altarriba, 1988) had mediating concepts through which activation could hypothetically spread among prime, mediator, and target; *deer–vegetable* with the mediator *animal* is an example. We label these pairs the McNamara–Altarriba pairs. Pairs of the second type, for example, *deer–grain*, did not have mediators through which activation could spread (according to free-association productions); we label these the McKoon–Ratcliff pairs. We measured the facilitation given by the prime of each pair to the target, using lexical decision as the response task. If the spread of activation is measured by free association, then according to spreading activation theories, there should be facilitation only for pairs with mediators, not for pairs without mediators. But for the compound cue theories, the existence of a mediator is irrelevant to the lexical decision response; facilitation should depend only on the familiarity of the pair of words as a compound, and if the familiarity of the two types of pairs is equal, then the amount of facilitation should be equal. (Note that by "familiarity" we mean the theoretical construct postulated by the compound cue theories, which is not necessarily the same as the empirical "familiarity" that is sometimes measured by subjects' ratings.)

Results were consistent with the compound cue view—there was facilitation for both types of pairs and about the same amount of facilitation. In the second experiment, a different and larger set of nonmediated pairs was used, and

again there was significant facilitation. These first two experiments showed that facilitation effects are not predicted by free association. The goal of the third experiment was to determine whether facilitation effects might be predicted by another variable, the frequency with which the two words of a pair co-occur in natural language.

In the final section of this article, we discuss how free-association production probabilities fail to predict priming effects and what other variables might be used to predict priming effects.

Experiment 1

Experiment 1 used two sets of materials, the McNamara–Altarriba mediated pairs, previously developed by Balota and Lorch (1986) and McNamara and Altarriba (1988), and the McKoon–Ratcliff nonmediated pairs. Balota and Lorch collected free-association data in order to determine, for each pair, that the target was produced as an associate of an associate of the prime but that the target was not produced as a direct associate of the prime. Balota and Lorch showed that the primes of these pairs facilitated naming responses to the targets, and McNamara and Altarriba showed that the primes facilitated lexical decisions to the targets. Facilitation was measured against a control condition in which primes and targets were randomly re-paired to give an unrelated prime for each target. For these pairs, we expected to replicate McNamara and Altarriba's finding of a small but significant priming effect in lexical decision.

The McKoon–Ratcliff pairs were made up of a prime from a pair used by Balota and Lorch (1986) and McNamara and Altarriba (1988), and a new target. The new target was a word we thought to be weakly and directly related to the prime but not produced directly as an associate of the prime in free association nor as an associate of an associate of the prime. If spreading activation is measured by free-association responses, then spreading activation theories predict either that priming will be reduced for these pairs relative to the McNamara–Altarriba pairs, or that there will be no significant priming. Compound cue theories predict that the amount of priming will reflect the familiarity of the prime–target pairs. If the familiarity for the McKoon–Ratcliff pairs is as high as the familiarity for the McNamara–Altarriba pairs, then the amount of priming will be the same for the two kinds of pairs.

McNamara and Altarriba (1988) showed that priming in lexical decision with their pairs can be obtained only under certain experimental conditions. Their data indicated that the relation between the prime and the target of a mediated pair should not be obscured by the relations between much more highly associated primes and targets. Our goal with the McNamara–Altarriba pairs was simply to replicate the priming previously obtained by McNamara and Altarriba so that we could compare it to priming with the McKoon–Ratcliff pairs. Therefore, we replicated McNamara and Altarriba's experimental design exactly (McNamara & Altarriba, 1988, Experiment 2, mediated-only condition), and in particular, there were no highly associated primes and targets in our experiment.

In presenting Experiment 1, we first describe the results for lexical decision priming, showing that small but significant amounts of priming are found for both the McNamara–Altarriba and McKoon–Ratcliff pairs. Then we describe a number of follow-up analyses of the two sets of pairs, in which we compare them using free-association production statistics and ratings of prime–target relatedness. Among all the follow-up analyses, the only difference between the two kinds of pairs is that the McNamara–Altarriba pairs have mediating concepts. Hence, we argue that there are no confounding variables that might provide spreading activation theories with the means to discount nonmediated priming.

Method

Subjects. The subjects in the lexical decision experiment were 88 students from an introductory psychology course, participating in the experiment for credit in the course. The experiment described here, about 10 min in length, preceded another experiment of about 30 min that is not relevant to this article. One group of 44 students was tested with the McNamara–Altarriba pairs. We used the exact lists of stimuli used by McNamara and Altarriba. The second group of 44 students was tested with the McKoon–Ratcliff pairs that we generated.[1]

Materials. For the group of subjects who were tested with the McNamara–Altarriba pairs, the materials were exactly the same as those used by McNamara and Altarriba, and a complete description is given in McNamara and Altarriba (1988, Experiment 2). These materials included words of the 48 triples from Balota and Lorch (1986) and 48 nonwords.

For the group of subjects who were tested with McKoon–Ratcliff pairs, the materials included the new nonmediated pairs, filler words, and nonwords. The new pairs were constructed from the 48 triples used by McNamara and Altarriba, where each triple was made up of a prime, a mediator, and a target (e.g., *cat, mouse, cheese*). The two words in the constructed pair were the original prime (*cat*) and a new word to be used as target (*meat*). The new target was chosen to share meaning with the prime in somewhat the same way as the old target did, but we intended that there would be no direct mediator between the prime and the new target. For *cat*, for example, we could think of no highly associated mediator that would lead to *meat*, but we thought that the overlap in meaning was about the same because *meat* and *cheese* are both things that animals eat. We constructed pairs like this for 20 of the 48 triples, as follows: *lion–spots, beach–bag, deer–grain, nurse–teacher, war–noisy, eyes–taste, soap–eat, cat–meat, rough–cotton, ceiling–drapes, hard–wool, navy–gun, moon–cold, flower–root, window–roof, school–go, birthday–pudding, oyster–*

[1] Our first effort to replicate McNamara and Altarriba's (1988) findings was not successful, and so it is important to describe details of our procedure exactly and completely. When we failed to replicate, we used test lists that we constructed from the Balota and Lorch (1986) materials rather than McNamara and Altarriba's lists, the experiment was conducted in the winter and spring quarters, the experimenter was sometimes an undergraduate work-study student, and many subjects were participating in their second or third reaction-time experiment in our laboratory. When we succeeded in replicating, we used McNamara and Altarriba's lists, the experiment was conducted in the fall quarter with almost all subjects freshmen, the experimenter was a recent graduate and so older than the subjects, and all subjects were participating in their first reaction-time experiment in our laboratory. We believe that the difference between succeeding and failing to replicate was due to reduction in variance as a result of using motivated, serious subjects.

bracelet, lemon–salty, summer–rain. The filler words for the subjects who were tested with the McKoon–Ratcliff pairs were chosen from triples that were not used to form the McKoon–Ratcliff pairs, and the nonwords were chosen from those used in the McNamara and Altarriba lists.

Procedure. All test items were presented on a cathode ray tube (CRT) screen and responses were collected on the CRT keyboard. Stimulus presentation and response recording were controlled by a real-time computer system.

The experiment began with 30 word–nonword test items for practice. Then the 120 test items of the experiment proper were presented. To begin the practice items, and before the first and the 61st test items, the instruction *Press the space bar when ready* was displayed on the CRT screen. When the space bar was pressed, the test items were displayed one at a time. Each test item remained on the screen until a response key was pressed, then the test item was erased, and if the response was correct, the next test item appeared after a 100-ms pause. If the response was not correct, the word *ERROR* was displayed for 1,500 ms followed by a pause of 1,000 ms before the next test item. Subjects were instructed to press the ?/ key on the keyboard to respond "word" and the Z key to respond "nonword." They were instructed to respond as quickly and accurately as possible. This procedure is the same as that used by McNamara and Altarriba.

For the subjects with McNamara–Altarriba pairs, the test lists were those constructed by McNamara and Altarriba to have no directly related test pairs; all related pairs of words were related through a mediator and not directly (see McNamara & Altarriba, 1988, Experiment 2). A complete description of the test lists is given in McNamara and Altarriba (1988). To summarize, the lists contained 12 related pairs (e.g., *cat–cheese*), 12 control pairs (unrelated words), 24 nonword–word pairs, and 24 word–nonword pairs. The words of each pair were presented one immediately after the other in the test list, and thus the pairings were not apparent to subjects in any obvious way.

The test lists for the McKoon–Ratcliff pairs were constructed in the following way: The first 60 test items comprised 5 experimental targets immediately preceded in the test list by their related words (e.g., *cat–meat*), 5 targets immediately preceded by a control word (e.g., *sky–meat*), 10 filler words followed directly by nonwords, and 10 filler words preceded directly by nonwords. These 30 pairs were placed in the test positions in random order. The second 60 test items were arranged in the same manner.

Design. Assignment to the two groups, one receiving McNamara–Altarriba pairs and one McKoon–Ratcliff pairs, was random according to arrival time at the lab, except that the number of subjects in each group was kept approximately equal. For the group of subjects who received McKoon–Ratcliff pairs, there were two experimental conditions: The target was preceded in the test list either by its related prime or by a control word. The control word was a prime for some other target. The experimental conditions were crossed with sets of pairs (10 per set) and groups of subjects. For the groups of subjects who received the McNamara–Altarriba pairs, the design was somewhat more complicated (see McNamara and Altarriba, 1988) but could be treated in the same way as for the McKoon–Ratcliff pairs, with each target preceded by its related prime or a control word (the control word was a prime for some other target).

Results

Means were calculated for each subject and each item, and means of these means are shown in Table 1. Analyses of variance were performed on these means, with both subjects and items as the random variables, and $p < .05$ was used throughout. One of the McKoon–Ratcliff pairs was deleted from the analyses for reasons given in the Materials Analyses section. However, the pattern of results (and the significance of the effects) did not change whether or not this item was included.

Table 1
Response Times (RTs in Milliseconds) and Error Rates (ER in Percentages) for Targets From Experiment 1

	Mediated pairs		Nonmediated pairs	
Condition	RT	ER	RT	ER
Related	570	3	562	2
Control	584	5	575	6
Word filler	575	2	574	2
Nonword filler	702	13	707	9

As can be seen in the table, the amount of facilitation given by a related word to its target is 13 ms with the McKoon–Ratcliff nonmediated pairs and 14 ms with the McNamara–Altarriba mediated pairs, in both cases remarkably close to the 14 ms of facilitation obtained by McNamara and Altarriba (1988, Experiment 2, mediated-only). Analyses of variance showed the amount of facilitation significant, $F_1(1, 86) = 5.3$ with subjects as the random variable, and F_2 (1, 38) = 4.1 with items as the random variable. The *F*s for the main effect of the two groups of subjects (one group for the McNamara–Altarriba pairs and one for the McKoon–Ratcliff pairs) and the *F*s for the interaction of the two variables were less than 1. The standard error of the response time means was 4.3 ms. For error rates, all *F*s were less than 1. These analyses included only the 20 of the McNamara–Altarriba pairs that had the same prime as the McKoon–Ratcliff pairs.

Materials analyses. The results of Experiment 1 suggest that an associated prime can facilitate the lexical decision on a target when, by looking at free-association production probabilities, it appears that the two words are neither strongly directly associated nor associated through a mediator. As previously argued, it is difficult to account for this result with standard spreading activation models if we assume that priming is predicted by free-association production probabilities. Free association is the only method of determining connections between concepts that has been offered as a predictor variable with which to account for priming effects with spreading activation. Without free association, it is not clear how spreading activation theories can predict when facilitation should and should not occur. However, several questions can be raised about the McKoon–Ratcliff pairs of words that were generated for Experiment 1. In this section, we address these questions.

First, it might be the case that the prime and target for the McKoon–Ratcliff pairs were more strongly associated than the prime and target for the mediated pairs, or that, despite our intentions, there actually were mediators for the McKoon–Ratcliff pairs. To rule out these possibilities, we asked subjects to generate free associations to the primes, using the same procedure that was originally used by Balota and Lorch (1986) for the mediated triples.

Two questionnaires were constructed, one for the prime word (e.g., *cat*) of 10 of the McKoon–Ratcliff pairs used in

Experiment 1 and one for the prime word of the other 10 pairs. Ninety subjects were each given one of the questionnaires and asked to write down eight associates for each prime, and in addition, they were asked to try not to generate the associates from their own responses but rather to generate associates from the prime words directly. On the questionnaires, each prime was presented on one line, eight blank lines followed, then the next prime and eight blank lines, and so on.

The responses on the questionnaires were scored in four ways. For the original McNamara–Altarriba mediated triples, we searched for the mediators and the targets, and for the McKoon–Ratcliff pairs, we searched for the targets and any possible mediators. For example, for the prime *lion*, we searched for *tiger, stripes, spots*, and any possible mediator between *lion* and *spots*, such as *leopard*.

For the McNamara–Altarriba mediated triples, the mediator should be given frequently (Balota & Lorch, 1986), and this is what we found. Out of 900 possible chances (10 primes per subject for 90 subjects), the mediator was given as a response 402 times (45%). For these triples, Balota and Lorch found that targets were never given as responses to the primes. However, in our questionnaires, 1 of 45 subjects gave *cheese* in response to *cat*, 3 gave *carpet* in response to *ceiling*, 2 gave *necklace* in response to *oyster*, and 2 gave *sweet* in response to *lemon*; this amounts to 0.8%.

For the 20 McKoon–Ratcliff pairs, 1 of 45 subjects gave the target as a response to the prime for each of four primes (*lemon, flower, moon*, and *war*). This pattern of a few targets generated as associates closely matches the pattern for the McNamara–Altarriba targets. However, for one of our pairs (*navy–gun*), the target was given by 6 of 45 subjects. This item was the one eliminated from analyses of the response time data.

In searching the responses to the primes for the McKoon–Ratcliff pairs, we looked for responses that could have been possible mediators between a prime and its target (e.g., a mediator between *deer* and *grain*). We found only one such response, *leopard* as a mediator between *lion* and *spots*, given by only one subject. We also tabulated the data to obtain the four most frequently given responses for each prime word (after first eliminating responses that were the targets or the mediators for the mediated targets). Questionnaires were constructed with the four responses for each of 10 of the primes (40 words in all). Twenty subjects were asked to give four associates to each of these 40 words. Of the 3,200 responses ($20 \times 4 \times 40 = 3{,}200$), only two were the McKoon–Ratcliff targets for the original prime word. It appears, therefore, that free association does not produce any mediators between the McKoon–Ratcliff prime and target that could account for significant priming effects.

Another possible problem with the McKoon–Ratcliff pairs might be that the McKoon–Ratcliff target was a high associate of the McNamara–Altarriba target. In other words, for the prime *cat* with the mediated target *cheese, meat* might be an associate of *cheese*. If this were the case, then the reason for the facilitation of responses to *meat* might be activation spreading through the original mediator and the original McNamara–Altarriba target to the McKoon–Ratcliff target. To check this possibility, we used another set of questionnaires with the McNamara–Altarriba target as the word to which associates were given, and we counted the number of times the McKoon–Ratcliff target was given as an associate. For 19 subjects who each generated four associates to the McNamara–Altarriba target, only 4% of the time was the McKoon–Ratcliff target given. Elimination of the five items that accounted for most of the generated McKoon–Ratcliff targets from the analyses of the lexical decision priming data still showed significant amounts of facilitation for the McKoon–Ratcliff as well as for the McNamara–Altarriba pairs (and no interaction between amount of facilitation and type of pair).

Another way to compare the McKoon–Ratcliff prime–target pairs to the McNamara–Altarriba prime–target pairs is to ask subjects to rate "how related" are the two words of a pair. It is possible that empirical relatedness ratings might reflect the theoretical construct of familiarity used in compound cue theories. Thus, it is possible that relatedness ratings might predict the amount of facilitation on target responses. To check this possibility, we constructed another set of questionnaires with pairs of words for subjects to rate (on a scale of 1 to 7, with 7 being *most highly related*). There were two questionnaires, each with 10 of the McKoon–Ratcliff pairs, 10 of the McNamara–Altarriba pairs, 15 pairs of highly associated words such as *thin–fat* (taken from the highly associated pairs used by McKoon & Ratcliff, 1979), and 15 pairs of words for which there was no obvious relation (e.g., *games–round*). Twenty subjects were tested with each of the questionnaires. The mean rating for the McKoon–Ratcliff pairs was 3.16; for the McNamara–Altarriba pairs, 2.61; for the high associates, 3.5; and for the unrelated words, 1.1. Analysis of variance showed the difference between ratings on the McKoon–Ratcliff pairs and the McNamara–Altarriba pairs marginally significant, $F_2(1, 19) = 3.7$, but the difference was due to only four of the pairs. Eliminating these pairs from the analysis led to means of 2.69 for the McKoon–Ratcliff pairs and 2.65 for the McNamara–Altarriba pairs, and to an F_2 value less than 1. Eliminating these four pairs from the analyses of the lexical decision response times did not change the pattern of results; the amount of facilitation for the McKoon–Ratcliff pairs was still 14 ms, and the effect was still (marginally) significant. We also calculated the correlation between the mean rating for each word pair and the mean amount of facilitation for that pair from Experiment 1. For the McKoon–Ratcliff pairs, we found $r = -.14$, and for the McNamara–Altarriba pairs, $r = -.044$, both nonsignificant.

The relatedness ratings show that the lexical decision results for the McNamara–Altarriba and McKoon–Ratcliff pairs cannot be explained as due, in some way, to differences in relatedness for the two kinds of pairs. Other conclusions that might be drawn about the ratings are more tenuous. Within the groups of items, the ratings did not correlate with lexical decision response times. But this would probably not be true in general; larger differences in ratings (which might be obtained by including strong direct associates in the experiment) would certainly lead to positive correlations between ratings and response times. It is also not possible to draw a general conclusion about the relation between relatedness ratings and the theoretical construct of familiarity that is part of the compound cue theories. Familiarity is hypothesized to drive the processes involved in fast, automatic decisions like lexical

decisions. Relatedness ratings are not fast and automatic but based on slower assessments, and so they probably do not reflect exactly the same information that enters into lexical decisions (see Ratcliff & McKoon, 1982, 1989).

Naming latency. With the original McNamara–Altarriba pairs used by Balota and Lorch (1986) and McNamara and Altarriba (1988), facilitation was obtained between prime and target in both lexical decision and naming latency. Therefore, we checked whether the McKoon–Ratcliff pairs also showed facilitation in naming latency.

In this experiment, words were presented in pairs. Subjects were instructed to read the first word of the pair and then pronounce aloud the second word of the pair. The first word was displayed for 250 ms on a CRT screen and then erased from the screen, and the second word was displayed until the subject pronounced it. The subject then pressed a key to indicate whether the pronunciation had been correct. Then, after a 1,000-ms pause, the first word of the next pair was presented.

There were 15 pairs for practice. Then the 20 McKoon–Ratcliff targets with their primes plus 40 filler targets and primes were presented in random order. The McKoon–Ratcliff targets were presented either with their related primes or with a prime for some other target. Half of the words used as filler primes and targets were words used in the original McNamara–Altarriba pairs, and half were words known to have slow naming latencies from previous data (they were chosen from the 10% slowest from a corpus of about 3,000 words). Half of each kind of filler were primes and half were targets. No word was used more than once in the experiment. The subjects were 36 undergraduates from the same population as in Experiment 1.

The results showed that the McKoon–Ratcliff primes did facilitate naming latency for their targets, by 12 ms (515 ms vs. 527 ms). This difference was significant with subjects as the random variable, $F_1(1, 35) = 9.1$, and with items as the random variable, $F_2(1, 18) = 7.5$, with a standard error of 3.0 ms.

Considerable discussion of priming effects has involved the naming task. However, the compound cue models do not address priming phenomena in naming because of the differences in processing. In the view of these models, naming requires retrieval of a specific test item from one of a large number of verbal items in order for a response to be given, whereas lexical decision requires deciding the degree of familiarity of a test item. Empirically, priming in naming latency has been found for the McNamara–Altarriba pairs (Balota & Lorch, 1986), and the data presented here show that priming can also be found for the McKoon–Ratcliff pairs and that it is of about the same magnitude (Balota & Lorch found an effect of 16 ms). Thus, we have addressed the empirical issue, but theoretical interpretation must wait for a comprehensive model of naming and lexical representation (see Ratcliff & McKoon, 1992a, for further discussion on this point).

Discussion

The result of Experiment 1 is straightforward. The amount of facilitation given by a prime to its target did not depend on the existence in free-association productions of a mediating concept to relate the prime to the target. For prime–target pairs with mediators (as defined by free-association production probabilities), there was 14 ms of facilitation; for prime–target pairs without such mediators, there was 13 ms of facilitation. In previous tests of priming by spreading activation theorists, the amount of facilitation has been said to be predictable from free-association responses: The amount of facilitation should be greater when there is a mediating concept between prime and target than when there is not. For the prime–target pairs in Experiment 1, the probability that a mediator would be given in free association for the McNamara–Altarriba pairs was .45, whereas it was only .008 for the McKoon–Ratcliff pairs. If priming is to be predicted from free association, this large difference should be reflected in the amount of facilitation in the lexical decision task, but it was not.

If free-association production probabilities cannot in general be used to predict priming effects, then they are almost certainly not a direct reflection of associative links in memory. If this is the case, then there is no basis on which to claim that the primes and targets of mediated pairs are not directly connected to each other. It may be that they *are* directly connected, but by links that are not used in free association. If they are directly connected, then finding priming for them is fully consistent with compound cue theories. Thus, the phenomenon of mediated priming is not evidence against these theories.

Experiment 2

The goal of the second experiment was to extend the generality of the nonmediated priming result to a new and larger set of prime–target pairs. The McKoon–Ratcliff targets used in Experiment 1 were generated by intuition, and it was desirable to find pairs that we ourselves had not constructed. In addition, we extended generality by using a slightly different procedure. Instead of requiring a lexical decision response to both primes and targets, as was done in Experiment 1 and in McNamara and Altarriba's Experiment 2, the procedure in our Experiment 2 followed McNamara and Altarriba's Experiment 1 in requiring a response only to the target. The prime was presented 200 ms in advance of the target, and subjects were asked to read it but to make no response to it.

New nonmediated priming pairs were obtained from the words of sentences used by Duffy, Henderson, and Morris (1989). Their sentences (originally used by Stanovich & West, 1981) contained a subject noun and an object noun that were weakly associated. Examples include *climber–summit*, *gardener–trowel*, and *skier–avalanche*. We hypothesized that these words were weakly and directly associated, so that there would be significant priming between them when they were presented as prime and target.

Duffy et al. (1989) did not test for priming between the words in these pairs. However, they did test for priming with whole sentences, including articles and verbs. The prime in their experiments was a phrase made up of the words of a sentence up to the final object noun; these words included the subject noun, a verb, articles, and sometimes an auxiliary verb. The final object noun was presented as a target. In one

condition, the sentence formed by the priming phrase and the target object had relatively high familiarity, for example, *The climber reached the – summit.* In a second condition, the sentence formed by the priming phrase and the target object had relatively less familiarity, for example, *The climber watched the – summit.* As Duffy et al. point out, responses to the target noun should be inhibited in the second condition relative to the first, and this is the result they obtained. However, there is no way to determine from this result what would happen if the subject noun alone were presented as the prime (*climber* alone instead of *The climber watched the*). With only the two words, subject noun and object noun as prime and target, they would both certainly be in short-term memory and enter the compound with which memory was probed. But with a whole sentence, it is less certain that the subject noun and object noun would both be part of the compound. In addition, even if the whole sentence does form the compound, we have no a priori way of determining the relative familiarities of the subject–object compound (*climber–summit*) and the phrase–object compound (*The climber watched the summit*). Duffy et al. do provide another condition for comparison, a phrase prime that used a different subject word (e.g., *The people watched the* for the target *summit*). But there is still no way to use this condition to determine priming for the subject–object pair. Again, this is because there is no way to determine the relative familiarities of the different compounds. The familiarities of the two phrase–object compounds (*The climber watched the summit* and *The people watched the summit*) may not be significantly different. In summary, there are no data from Duffy et al.'s experiments upon which to base our prediction that there would be priming for the subject–object pairs from their sentences. Our prediction was based on our intuition that the pairs had some familiarity greater than the familiarity of randomly paired words.

If the subject–object pairs do have familiarity greater than that of randomly paired words, then compound cue theories predict a significant priming effect between the subject as prime and the object as target. The prediction from spreading activation theory depends on whether there is a mediator such that activation can spread among prime, mediator, and target. The only way suggested to determine the existence of such a mediator has been free association. If free-association responses map memory, and if they do not produce a mediator, then either there should be no facilitation from prime to target, or at least the amount of facilitation should be reduced relative to pairs for which there are such mediators (such as the McNamara–Altarriba pairs in our Experiment 1).

Method

Materials. The 44 word pairs were chosen from the sentences used by Duffy et al. (1989). The cue word of each pair was the subject of one of the sentences used by Duffy et al., and the target word was the object of the sentence. Some examples are *wine–decanter, mortician–cadaver, politician–constituency,* and *accountant–ledger.* The complete set of sentences is given in Duffy et al. There were also a pool of 480 words used as fillers and a pool of 600 nonwords.

Procedure. The test items were presented on a CRT screen, and responses were collected on the CRT's keyboard. Test items were presented as prime–target pairs. Each pair was preceded by a warning signal (a row of pluses) displayed for 400 ms; then, on the next line, the prime was displayed for 200 ms; and then, on the next line, the target was displayed. The target remained on the screen until a response key was pressed (? for "word," Z for "nonword"). If the response was correct, the warning signal for the next item was displayed after a pause of 700 ms. If the response was an error, the word *ERROR* was displayed for 1,500 ms before a blank interval of 1,000 ms followed by the next warning signal.

The experiment began with 15 practice test items. After that, the items were divided into four blocks. Each block began with an instruction to press the space bar on the keyboard to initiate the block. Each block included 5 or 6 of the experimental targets with their related primes, 6 or 5 of the experimental targets with unrelated primes, 40 pairs for which the prime and target were unrelated words, and 40 pairs for which the prime was a word and the target was a nonword. These pairs were arranged in random order, except that the experimental targets could not occur in the first four positions in the block. Assignment of items to blocks was also random. No word or nonword was presented more than once in the experiment.

Design and subjects. The experimental targets were presented either with their related primes or with unrelated primes. The unrelated primes were the related primes for other targets. This variable was crossed with two sets of items (22 per set) and two sets of subjects. There were 38 subjects, participating in the experiment for credit in an introductory psychology course.

Results

Means were calculated for each subject and each item in each condition. The main result was that responses to targets were faster with a related prime than with an unrelated prime, 643 ms (11% errors) versus 667 ms (12% errors), $F_1(1, 37) = 5.3$ and $F_2(1, 43) = 9.9$. The standard error of the response time means was 7 ms. There were no significant differences in error rates. Mean response time on filler words was 587 ms (5% errors), and mean response time on nonwords was 698 ms (10% errors). Responses to the experimental targets were slower and less accurate than responses to the fillers, we assume because the targets occur with lower frequency in the language.

We checked free associations and relatedness ratings for these pairs of words as we did for the pairs used in Experiment 1. Twenty-five subjects rated how related the 44 pairs were; the correlation between the ratings and facilitation was $r = -.135$. Thirty-nine subjects were each given 22 of the cues and asked to generate eight free associates to each one. Only 0.3% of the time did subjects give a target word as a response, less than for the McKoon–Ratcliff pairs and McNamara–Altarriba pairs used in Experiment 1. (In tabulating the data, we counted synonyms of targets as well as actual targets.) We searched the responses to each prime for words that could serve as mediators—words to which the target might be produced as a free associate—but there were almost no possible mediators. This finding is easiest to document with examples. For the primes of the first five pairs, the three most frequently given free associates were as follows: for the prime *wine–red, white, glass*; for the prime *mortician—death, coffin, black;* for the prime *politician—campaign, corrupt, speech*; for the prime *accountant—money, taxes, numbers*; for the prime *general—army, war, stars.* The targets for these five

primes were *decanter, cadaver, constituency, ledger,* and *strategy*. None of the associates given to the primes seems likely to give a target in free association, and therefore none seems likely to serve as a mediator.

Discussion

The nonmediated pairs of Experiment 2 showed a priming effect just as the nonmediated pairs of Experiment 1 did. Experiment 2 used a larger and different set of pairs than Experiment 1, and a slightly different procedure, and so provides generality for nonmediated priming.

The primes and targets in Experiment 2 were the subjects and objects of sentences used by Duffy et al. (1989). The result that these pairs show priming suggests a new interpretation of Duffy et al.'s data. They argued that a subject did not prime its related object, and they based this argument on their finding that a phrase prime containing the subject did not prime the object, relative to a neutral control condition. However, from the compound cue point of view, the absence of a priming effect with a phrase does not necessarily predict the absence of priming with a single word. A phrase prime is not the same as a single word prime, even if the phrase prime adds only what could be seen as "neutral" information to the single word. In the example *The climber watched the summit,* the addition of the seemingly neutral information *The... watched the* to the subject *climber* may change the familiarity of the resulting compound. Whereas *climber-summit* may have enough familiarity to give priming relative to a neutral control, a climber watching a summit may not. The effect of neutral information on priming has been documented before. O'Seaghdha (1989) placed function words between primes and their highly associated targets. If the function words were syntactically well formed, then priming effects were larger than if the function words were not syntactically well formed (e.g., *author of this book* vs. *author the and book*). In both cases, the function words were neutral information, but the form of the neutral information significantly affected priming.

Experiment 3

For Experiments 1 and 2, the pairs for which association was weak and direct were chosen on the basis of intuition. The pair *accountant–ledger* sounded good to us in a way that *wine–ledger* did not. There was no independent measure of the familiarity of the pairs. Priming was clearly not predicted by free-association production probabilities.

The purpose of Experiment 3 was to examine an alternative measure of weak association. In the compound cue theories, priming depends on familiarity, as defined in the global memory models. If the notion of familiarity is taken literally, then what is needed is a measure of the frequency with which the subjects in our experiments have encountered or processed a compound in past experience. Of course, there is no such measure, but what is available as the beginning of an approximation is a measure of frequency of occurrence in large samples of written language.

Church and Hanks (1989) have developed a measure they label an *association ratio*, defined for two words x and y as the mutual information (unidirectional) between the two words, $\log_2 [P(x, y)/P(x)P(y)]$. For a sample of language, this ratio compares the probability of observing the words x and y together (joint probability) with the probability of observing each of the words independently. If the two words are likely to co-occur in the sample, then their joint probability will be larger than the product of their independent probabilities, and the value of the ratio will be larger than 1. The probabilities are estimated from samples of the Associated Press (AP) newswire (several million words). The independent probabilities for x and y are estimated by counting the number of times x and y occur in the sample and normalizing by the number of words in the sample. The joint probability of x and y is estimated by counting the number of times that x is followed by y in a window of w consecutive words. If the value of the association ratio for a pair of words is larger than 1, then the words co-occur more often than would be expected by chance. Whether they co-occur significantly more often can be estimated with a t statistic (Church & Hanks, 1989).

For Experiment 3, we chose target words that we know to have highly associated primes (from published norms). For each target, we chose two additional prime words that co-occurred in a six-word window more often than would be expected by chance. The association ratios were based on statistics from a corpus of 6 million words from the AP newswire. We used word pairs for which the association ratio had a high t value and pairs for which the ratio had a low t value. It should be stressed that the corpus on which the t values were based was not large enough to make us confident about the relative sizes of the t values. To provide reliability and generality, it would be necessary to compute the t values from other corpora and for much larger corpus sizes. However, we thought it useful to include both the high and low t values to determine whether there was a priming effect for both or only for the high t-value pairs, and to leave reliability of the split into high and low t values until larger corpora become available.

For each target word used in the experiment, there were four different priming conditions. One prime was a word from which the target would be produced in free association with a high probability. For example, the target *baby* is produced in response to the prime *child* with a high probability (according to free-association norms). The second and third primes for a target were the words that formed pairs with either high or low t values. For the target *baby*, the association ratio for the pair *hospital–baby* had a high t value, and the association ratio for the pair *room–baby* had a low t value. The fourth prime for a target was unrelated to the target; it was a randomly chosen low t value prime for some other target.

The high and low t value primes were chosen so that they would be unlikely to elicit their targets or mediators to their targets in free association. However, the probability of production in free association could not be kept as low as for the nonmediated pairs that were used in Experiments 1 and 2. This was because there were three constraints on the pairs that had to be simultaneously met. First, the targets had to be

words for which a highly related associate prime was available from free-association production norms. Second, the targets had to be words that occurred frequently enough in the AP newswire corpus to provide meaningful association ratios. Third, the targets had to have primes that had significant t values (and that gave the targets with low probability in free association). For the 40 targets that met these constraints, the probability that the high t value primes elicited the targets in free association was .04 (up from .004 for the nonmediated pairs in Experiment 1), and the probability that the high t value primes elicited mediators was estimated to be .12 (up from .0025 in Experiment 1).

Method

Materials. Forty target words were chosen such that each had three prime words. For one prime, the target was highly related, as measured by free-association data (from standard norms). For the second and third primes, the target co-occurred more often than would be expected by chance within a window of six words in the AP newswire corpus. For the second prime, the t statistic averaged 6.56, and for the third prime, it averaged 1.73. There were primes for which the t value was higher, but we did not use primes or synonyms of primes that were associated to the targets in the free-association norms. The 40 sets of words are given in the Appendix. It should be noted, first, that the high and low t value primes reflect their origin in the AP newswire corpus, and second, that these primes represent several kinds of associations with their targets. In addition to the primes and targets, there were a pool of 309 words to be used as fillers and a pool of 600 nonwords.

Procedure. Stimuli were presented on a CRT screen, and responses were collected on the CRT's keyboard. The test items included highly associated prime–target pairs. Previous research (McNamara & Altarriba, 1988) suggests that including such pairs in the experiment may lead subjects to adopt strategies that result in the absence of priming for weakly associated pairs. However, McNamara and Altarriba suggested that these strategies can be avoided if responses are required to both the prime and the target. Hence, we used this procedure (similar to the procedure used in Experiment 1). Lexical decision responses were made to both prime and target test items. Test items were presented one at a time, with each item displayed until a response key was pressed. If the response was correct, the next item was displayed after a 100-ms blank interval. If the response was not correct, the word *ERROR* was displayed for 1,500 ms, followed by a 1,000-ms blank interval before the next test item.

The test list was divided into a practice list of 30 items, followed by 10 sublists of 36 items. Each sublist was made up of 4 target words, each preceded in the list by the prime word appropriate to its experimental condition, 16 filler words, and 12 nonwords. Except that the experimental targets could not occur in the first four test positions, the test items were randomly ordered. No test item occurred in the experiment more than once.

Design. There were four experimental conditions. The target word was preceded in the test list by the prime highly related in free-association norms, by the prime related by a high value of the t statistic, by the prime related by a low value of the t statistic, or by an unrelated word. The unrelated primes were chosen from the low t-value primes for other targets. The four conditions were combined with four sets of items and four groups of subjects in a Latin square design. There were 52 subjects serving in the experiment for credit in an introductory psychology course.

Results

Means were calculated for each subject and each item in each condition. Over the four conditions, there were significant differences in the response time means, $F_1(3, 153) = 6.5$ and $F_2(3, 117) = 7.5$, with a standard error of 7.5 ms. The fastest response times occurred with the prime highly related by free-association norms, 500 ms (0.8% errors), and the slowest times with the unrelated prime, 549 ms (1% errors). As predicted, the prime related by a high value of the t statistic speeded responses to a mean of 528 ms (2% errors). This mean was significantly different from the unrelated mean, $F_1(1, 153) = 3.9$ and $F_2(1, 117) = 4.3$. The prime related by the low value of the t statistic speeded responses somewhat, 532 ms (1% errors), but not significantly so, $F_1(1, 153) = 2.6$ and $F_2(1, 117) = 2.8$. For filler words, the response time mean was 571 ms (2% errors), and for nonwords, 712 ms (8% errors).

As in the preceding experiments, we collected ratings of the relatedness of the prime and target words. The mean of the ratings for the low t statistic prime with the target was 3.9, the mean for the high t-statistic prime with the target was 4.9, and the mean for the free-association prime was 5.9 (calculated over 64 subjects, who each rated all of the 40 targets, one third with each of the three primes). The correlation between amount of facilitation of response times and relatedness rating was .26 for the low t-statistic primes, and $-.11$ for the high t-statistic primes. Free-association responses (four responses for each prime word) were collected from 12 subjects for 35 of the 40 items used in the experiment. The probabilities with which targets and mediators to targets were produced were given in the introduction section.

Discussion

Experiment 3 shows that co-occurrence statistics calculated from large corpora have potential applicability as predictors of priming effects. While the corpus we used was relatively small, we anticipate the availability of larger corpora and further research with them. Meanwhile, we point to co-occurrence statistics as variables that fit naturally with the compound cue theory point of view.

General Discussion

We have previously claimed that compound cue theories of priming can explain at least as much data as spreading activation theories and that therefore compound cue theories provide an important alternative view (Ratcliff & McKoon, 1988; Dosher & Rosedale, 1989). Compound cue theories can explain the many kinds of priming effects outlined in this article. They also inherit all the properties of the global memory models on which they are based and so are embodied in a framework that can account for a range of other kinds of data such as recognition, recall, frequency judgments, categorization, and so on.

Mediated Priming?

Recently, the compound cue approach has been criticized for its inability to account for mediated priming (McNamara & Altarriba, 1988). In this article, we argue that what has been called mediated priming for a prime and target is instead priming resulting from weak *direct* associations between prime and target—priming that is fully consistent with compound cue theories.

The crux of the argument is how to decide whether a prime and target are directly related or related only through a mediator. Previous investigations of mediated priming have depended on free-association production probabilities to determine that a particular prime and target are not related directly but that they are related through a mediator. However, Experiments 1 and 2 indicate that free association does not adequately explain priming. In Experiment 1, for example, production probabilities differed dramatically from the mediated pairs used by McNamara and Altarriba (1988) to the new, nonmediated pairs that we generated. The probability of a mediator appearing in free association was .45 for the McNamara–Altarriba pairs, whereas it was estimated to be only .008 for the McKoon–Ratcliff pairs. But the facilitation in response time was almost identical for the two sets of pairs (13 ms and 14 ms).

If free-association production probabilities cannot by used to distinguish whether a prime and target are directly related or related only through a mediator, then one possibility is to simply abandon free association as a predictor variable for priming. This course of action carries with it two important consequences. First, it leaves compound cue theories free of criticism based on mediated priming; mediated priming can be said to be priming between directly related weak associates. Second, abandoning free association would mean that spreading activation theories lose the only way they have had to predict priming effects from network distance. In previous studies, the only variable that has been used to distinguish direct from mediated priming has been free-association production probabilities. Without free association, spreading activation theories will need to find some new (noncircular) way of predicting priming.

In contrast, compound cue theories do not need free association as a predictor of priming. In fact, from the point of view of these theories, free association would not necessarily correspond exactly to priming because the cue to the memory system is different in the two cases. The cue in priming includes both the prime and target, whereas the cue in free association does not include the target. Instead of free association, compound cue theories find a natural predictor variable in co-occurrence statistics. Although the co-occurrence statistics used in Experiment 3 were based on only a small corpus and the results of the experiment are somewhat tentative, we expect that this approach will be a fruitful one in the future. Compound cue theories can also make use of semantic relationships among words. Fischler (1977) selected pairs of words for which the target was never given as a free-association response to the prime and for which there was very low probability that the same words were given in response to both the prime and target. Fischler found that the amount of priming for these pairs was as large as the amount of priming for pairs that were strongly directly associated according to free-association production probabilities. Semantic relatedness correlated positively with the size of the priming effect, but free-association production probabilities correlated negatively with priming (see also the replication by Seidenberg, Waters, Sanders, & Langer, 1984). Although recent work (McKoon & Ratcliff, 1992; Shelton & Martin, 1992) suggests the need for more research into semantic priming effects,[2] semantic relatedness and co-occurrence statistics are variables consistent with compound cue theories as predictors of priming effects. In sum, abandoning free association as a variable to predict priming is not problematic for compound cue theories but has serious consequences for spreading activation theories.

One response that spreading activation theorists can make is to try to salvage free association. McNamara (1992) attempts to do exactly this by finding potential mediators for the McKoon–Ratcliff pairs and validating them with free-association production probabilities. However, as will be detailed subsequently, these new mediators have different characteristics from the original mediators for the McNamara–Altarriba pairs. Unlike the mediators for the McNamara–Altarriba pairs, the new mediators are not among the highest-probability associates produced from their primes.

To generate the new mediators for the McKoon–Ratcliff pairs, McNamara (1992) thought up potential mediators himself and then tested these potential mediators in free association. For example, consider the McKoon–Ratcliff pair *flower–root*. In the free-association data collected for Experiment 1, subjects did not give any responses to *flower* that in turn would lead to *root*. But McNamara thought that *plant* would be a potential mediator. To show that it was, he collected free-association responses to all three words, the prime, the potential mediator, and the target. He found that the probability that *plant* was produced in response to the prime *flower* was very low (.08), consistent with the free-association data from Experiment 1. But he also found that the probabilities with which the prime and target were produced from the mediator were high (both *flower* and *root* were frequently given as responses to *plant*). Using his method, McNamara (1992, Appendix C) was able to find pathways (connected links for which the free-association production probabilities were larger than zero) among prime, target, and one or more mediators for all but one of the McKoon–Ratcliff pairs.

There are two problems with the use of these production probabilities to predict priming. The first concerns how the probabilities should be measured, and the second concerns how they should be averaged across items. When McNamara (1992) examined his potential new mediators for the McKoon–Ratcliff pairs, he calculated the probability that a me-

[2] Shelton and Martin (1992) failed to find priming in lexical decision for a set of semantically related word pairs (e.g., *spider–ant*). However, using the same set of pairs, McKoon and Ratcliff (1992) did find a significant priming effect. Experiments that attempt to resolve this discrepancy in results are currently in progress.

diator was given in response to the prime by counting responses from all output positions, that is, from all the responses that subjects produced during 1 min. The probabilities reported for Experiment 1 were also based on all eight responses that subjects produced. However, according to earlier work in free association, a better measure is the first-production probability, that is, the probability that a word is produced as the first response to its prime (Keppel & Strand, 1970; Postman, 1970). The earlier researchers were attempting to measure strength of association, and they argued that (instructions to the contrary) responses later in the sequence are likely to be generated not just from the prime but from the prime plus the additional context of the other responses, in chains or other sorts of combinations of prime plus responses (see also Cramer, 1968). In the data from Experiment 1, one subject in response to *beach* produced *sand, water, ball, swimming,* and *umbrellas,* things that might be encountered at the beach, followed by *California, ocean, sea.* This example indicates that later responses may not be independent of earlier responses and that the later responses can be contaminated by earlier responses. Thus, following the earlier work, we would claim that first-production probabilities, not production probabilities calculated over all output positions, should be used in comparing different sets of items and in efforts to model free association and priming processes.

Figure 1 provides examples of differences between the old mediators for the McNamara–Altarriba pairs and the new mediators found by McNamara for the McKoon–Ratcliff pairs. The data are based on the free-association responses collected for Experiment 1, for which subjects were asked to generate eight free associates for each prime. First, the McKoon–Ratcliff pairs were divided into two sets. The first set is made up of the McKoon–Ratcliff pairs for which McNamara found one new mediator for a two-step chain (e.g., for the McKoon–Ratcliff pair *flower–root,* he found the mediator *plant* to give the chain *flower–plant–root*). The second set is composed of pairs for which he found two new mediators for a three-step chain (e.g., for the pair *deer–grain,* he found the chain *deer–animal–farm–grain*).

Figure 1 gives the probabilities with which mediators were given as responses to the primes. For example, for the prime *flower,* the figure shows probabilities of production for the new mediator *plant* that would hypothetically mediate between *flower* and the McKoon–Ratcliff target *root.* For the three-step chains, the figure shows probabilities for the first mediator in the chain. The figure also shows probabilities of production for the old mediators that would hypothetically mediate between the prime and the McNamara–Altarriba target (e.g., *flower–rose–thorn*). In each of these cases, two measures of production probability are given. One is based only on responses that were the first produced to the prime, and the other is based on all eight responses that were produced. For example, for the prime *flower,* the response *plant* might never be produced as any subject's first response, and so its probability of first production would be zero. But *plant* still might be produced quite frequently in later positions in subjects' lists of responses.

Figure 1 shows that the old and new mediators can differ on both measures. Consider first the two-step items. The old

Free-Association Data (Experiment 1)

Two-step chains

	Prime	→ Mediator	→ MR Target
	flower	*plant*	*root*
Prob. from all responses	.176 (.081)		
Prob. from first response	.053 (.019)		

	Prime	→ Mediator	→ MA Target
	flower	*rose*	*thorn*
Prob. from all responses	.423		
Prob. from first response	.180		

Three-step chains

	Prime	→ Mediator	→ Mediator	→ MR Target
	deer	*animal*	*farm*	*grain*
Prob. from all responses	.336 (.207)			
Prob. from first response	.114 (.022)			

Figure 1. Probabilities of free-association responses to primes for the two-step McKoon–Ratcliff (MR) pairs (top panel); the McNamara–Altarriba (MA; 1988) pairs (middle panel); and the three-step MR pairs (bottom panel). (The numbers in parentheses are the probabilities for pairs that did not include a MA mediator.)

mediators for the McNamara–Altarriba pairs appear among all responses with a high probability (.423), whereas the new mediators for the McKoon–Ratcliff pairs appear among all responses with a lower probability (.176). The probabilities of the mediators being produced as first responses show a greater difference: .180 versus .053. For the three-step items, the differences are not as large. Calculated over all responses, the probabilities are .423 versus .336; and over first productions only, .180 versus .114. For some of the items, the first mediator in the chain constructed by McNamara for the McKoon–Ratcliff pairs was the same word as the mediator for the old McNamara–Altarriba pairs. If we consider only those new McKoon–Ratcliff mediators that were not the same as for the McNamara–Altarriba pairs, then the differences between the new McKoon–Ratcliff mediators and the old McNamara–Altarriba mediators are much larger: .423 versus .081 and .207, and .180 versus .019 and .022.

The probabilities for the old mediators for the McNamara–Altarriba pairs and the new mediators for the McKoon–Ratcliff pairs in Figure 1 show quite different patterns. However, this is not the only problem in comparing the two kinds of mediators. There is also a problem with averaging. Suppose that for some of the two-step chains, the production probabilities were from prime to mediator, .1, and from mediator to target, .8; and that for other two-step chains, the probabil-

ities were the opposite: .8 and .1. Then the average prime-to-mediator probability would be .45, the same as the average mediator-to-target probability. This kind of averaging produces a potential problem for most spreading activation models. The amount of priming from prime to target will be predicted to be much larger if the prediction is based on averages than if it is based on the component probabilities from which the averages were calculated. For example, in the first case, using the components, .1 of the activation from the prime would be passed to the mediator and .8 of that would be passed to the target, that is, .08 would be passed to the target. But using the averages, .45 times .45 would be passed to the target, that is, .20, over twice as much as if the components were used. Inspection of the McKoon–Ratcliff pairs in McNamara (1992, Appendix C) shows that 15 out of 18 cases have one probability in the chain twice as large as another, and 13 out of 18 have one probability three times as large as another. In contrast, for the McNamara–Altarriba pairs, the prime-to-mediator probabilities include few very small values: the probability for most of the items is about the same as the average shown in Figure 1.

The analysis shown in Figure 1 is incomplete; it shows data only for free associations from the prime word to the mediators, not associations back to the primes or from the mediators to and from other mediators or the targets. Nevertheless, the mediators proposed by McNamara (1992) to link the McKoon–Ratcliff primes to their targets clearly pattern differently than the mediators proposed to link the McNamara–Altarriba pairs to their targets. The averages are different, as shown in Figure 1, and these averages are based on different distributions of probabilities across items. McNamara argues that these differences are not important when all the production probabilities for all the links among prime, mediators, and target are placed into a model such as ACT*; even given the differences, ACT* could predict equivalent amounts of priming for the two sets of pairs. However, the modeling has not yet been done, and so this remains an open question (see Ratcliff & McKoon, 1992a).

In summary, the ability of spreading activation models to use free-association production probabilities to explain the priming effects obtained in Experiment 1 appears to us to be an open question. Free-association production probabilities, as they have been defined in previous research, cannot predict the equality of priming for the McKoon–Ratcliff and the McNamara–Altarriba pairs. The new mediators suggested by McNamara (1992) may work, but a specific model such as ACT* has not been tested against the data. Moreover, questions remain about which measure of production probability is most appropriate for modeling, and how probabilities should be averaged across items.

So far, we have considered whether spreading activation models could be made consistent with both the priming and free-association data of Experiment 1. At this point, it seems reasonable to ask whether compound cue models can predict priming effects directly from free-association data. But is it reasonable?

Compound cue models, as we have mentioned, are intended to describe the processes by which cues focus on subsets of information in memory. The whole point of considering the prime and target as a compound is to focus on exactly those associations that make the appearance of the prime and target together in short-term memory more or less familiar. These might not be the same associations that come into focus when the prime is presented alone, in the context of a free-association experiment (Ratcliff & McKoon, 1992b). And if they are not the same associations, then predicting effects of one set of associations (based on the prime–target compound) from a different set of associations (based on a prime–free-association-context compound) will likely fail.

McNamara (1992) shows such a failure. He uses the compound cue theory as implemented in SAM (Gillund & Shiffrin, 1984; Ratcliff & McKoon, 1988). To apply SAM to the free-association production and priming data, connection strengths are set to produce familiarity values that fit the priming data. But once these strengths are set, McNamara shows that they are not consistent with free-association data. That is, if they are set strong enough to give the right amount of priming, then they also predict much higher probabilities of free-association production than are actually obtained in data. Thus, SAM cannot jointly accommodate priming effects and free-association production probabilities. But unlike ACT*, it is not necessarily desirable for SAM to do this; in SAM, different contexts (free association vs. prime–target pairs) may focus on different associations in memory.

Failure of models to predict both free association and priming should not be surprising. There are a number of norms that give frequencies of first-associate production (e.g., Postman & Keppel, 1970). These norms show that sometimes the first associate is given by as many as 70% of the subjects and the second most likely associate by only 4%, and other associates are even less likely. If priming effects were linearly related to production probability, then the priming effect for the most frequent associate would be 15–20 times that of the priming effect for the next most frequent. What would be surprising would be if only the most frequent associate ever gave priming, or if the priming effect for that associate were 20 times larger than for the next most frequent associate.

One clear conclusion to be drawn from this discussion is that there is currently no good account of the relation between free association and priming effects. The conclusion to be drawn about priming theories is less clear. If spreading activation theories can no longer depend on free association to predict priming effects, then these theories will have to find new predictor variables (or rely on intuition). Compound cue theories, on the other hand, already have other predictor variables (co-occurrence statistics, semantic relationships), but these variables are not yet well understood.

Lag Effects

Priming in lexical decision is usually studied when the target is presented immediately after the prime. But priming can also occur when the prime and target are separated in the test list by an unrelated item (Joordens & Besner, 1992; McNamara, 1992; Ratcliff, Hockley, & McKoon, 1985; Ratcliff & McKoon, 1978). This result implies that the compound with which memory is accessed might sometimes contain three test items, not just two. In the discussion that follows,

we label the three items preprime, prime, and target, where they are respectively the first, second, and third items presented in a successive triple (embedded in a long sequence of single-item trials).

It should be noted that priming from the preprime item is problematic for ACT*. In ACT*, activation arises from information that is currently being presented to the system. For ACT* to predict priming from preprime to target (as in the sequence *hammer–vase–nail*), both the prime and preprime items would have to be sources of activation. Given the parameters of lag experiments, the preprime would have to stay active for about 1,000–1,300 ms (depending on assumptions about when the prime starts to decay as a source of activation and when the decision process begins on the target). However, assuming that the preprime is active for this amount of time is problematic in light of other data. Ratcliff and McKoon (1988, Experiment 2) examined target–prime–target sequences (e.g., *dog–floor–cat*) and found that if the intervening prime was a word, then priming from the previous target to the current target was eliminated. If the previous target had been active for 1,000–1,300 ms, then priming should not have been eliminated. So, while keeping a preprime item active for 1,000–1,300 ms may allow ACT* to predict some lag effects, it leads to problems with other lag effects.

For compound cue models, if the compound contains three test items, then the relative amounts of priming for all the possible combinations of three items should be predictable. Consider, for example, the preprime, prime, and target sequence *hammer–vase–nail*. If the compound contains all three of these items, then the familiarity of *hammer–nail* should facilitate responses to *nail*, but the facilitation would be less than if the sequence were *vase–hammer–nail*. The reduction in amount of facilitation would come from placing less weight on the preprime than on the prime and less weight on the prime than on the target in the calculation of familiarity. There would also be facilitation for the target *vase* in the sequence *hammer–nail–vase* because of the association of *hammer* and *nail*, but the facilitation would be even smaller, again because of lower weights on the preprime and prime than on the target. Contrary to this last prediction, McNamara (1992) did not find facilitation for a target when the preprime and prime were related to each other but not to the target, and he uses this finding to argue against compound cue theory.

The problem with McNamara's (1992) argument is that it depends on the relative weights of the preprime, prime, and target. If the weights of the preprime and prime combined equal the weight of the target, and the weight on the preprime is greater than half of the prime weight, then McNamara is right—the amount of priming on the target should be large enough to observe empirically. But these are unreasonable assumptions. If the preprime and prime weights combined equal the weight of the target, then if the two items preceding the target are nonwords, the error rate on the target word would be 50%. More reasonably, the preprime and prime combined should be given less than half the total weight, and similarly, the preprime should have less than half the weight of the prime. Under these assumptions, the predicted amount of facilitation is too small to detect empirically.

Table 2 shows familiarity values calculated from the SAM model for preprime, prime, target triples for different values of weights and strengths of associations. In the table, U stands for a word unrelated to any other word in its triple, and R stands for words related to each other. For example, the triple *hammer–vase–nail* is represented as RUR. For the calculations, we assumed that the strength connecting a word presented as a cue to its own image in memory (e.g., *nail* to *nail*) was high and also that the strength connecting a word to a related image (e.g., *nail* to *hammer*) was high; these values were both set to 1.0 in the first column of Table 2. All other strengths were set to the same lower value (e.g., .2 in Column 1; see Ratcliff & McKoon, 1988, Table 1).

Consider the familiarity values in the first column of the table, where the target is given a little more weight than the prime and preprime combined (.6 vs. .3 vs. .1). When the prime is related to the target (URR), the value of familiarity for the target is much larger than when neither the prime nor the preprime is related to it (UUU); the familiarity values are 3.86 versus 3.45, an increment in familiarity due to priming of 0.41. However, in the condition which McNamara claimed a problem for compound cue theories, in which the preprime and prime are related to each other but not to the target (RRU), there is only a small amount of facilitation, 3.50 versus 3.45, an increment of only 0.05. This predicted amount of priming in familiarity for the RRU condition is only about 13% of the amount for the URR condition, and it would not be observable empirically (assuming roughly linear mapping from familiarity to reaction time). If URR gave 30 ms of priming, then RRU would give about 4 ms, which would be too small to observe empirically. At the same time, the facilitation for the RUR condition is about 30% of the UUU condition, which is detectable (though this is less facilitation than was obtained empirically by McNamara, 1992). In contrast, using McNamara's weights (.2, .3, and .5, so that half

Table 2
Familiarity of Various Preprime, Prime, and Target Relations

	Weights				
Triple	.1, .3, .6[a]	.14, .29, .57[a]	.14, .29, .57[b]	.2, .3, .5[a]	.1, .2, .7[a]
UUU	3.45	3.41	26.77	3.34	3.58
RRU	3.50	3.47	26.93	3.44	3.61
RUR	3.57	3.56	27.14	3.53	3.73
URR	3.86	3.77	27.60	3.64	3.90

Note. U = words unrelated to any other word in its triple; R = words related to each other.
[a] Strengths = 1 and .2. [b] Strengths = 5 and .2.

the total weight is on the preprime and prime; see column 4), priming in the RRU condition is 30% of priming in the URR condition, an amount of priming that would be observable empirically.

Further examples are given in the other columns of Table 2. With the weights in the second column of Table 2, the target gets twice the weight of the prime, which gets twice the weight of the preprime. In the fifth column, the target is weighted most heavily, showing priming in the RUR condition but little chance of detecting priming in the RRU condition. Again, it would be difficult to observe any priming in RRU with these values of weights (facilitation between 10% and 15% of URR), but priming of RUR would be observable (facilitation of about 50% of URR). The third column shows that results are similar if much higher strength values are used. In sum, Table 2 shows that if the preprime and prime combined have as much weight (or more) than the target, there should be an observable priming effect for RRU triples, but if the target has only half the weight or less, the effect will be too small to be observed.

McNamara (1992) also considers a second kind of triple, in which the preprime can be a nonword. He argues that compound cue theories cannot account for the effects of a nonword preprime, whereas spreading activation theories can. To understand this argument, it is important to understand what the two classes of theory predict, and why.

Consider a preprime, prime, target sequence in which the preprime can be either a nonword or a word completely unrelated to the prime or target. For spreading activation theories, activation will not spread from a nonword to the prime or target, and activation from a completely unrelated word will not spread to the prime or target. Therefore, responses to the target will not be affected by whether the preprime is a nonword or an unrelated word.

But the data show otherwise; a nonword preprime slows response times to the target (it slows response times equally for targets related to their primes and targets unrelated to their primes). This finding would seem to contradict the spreading activation prediction, but McNamara argues that the slow-down comes from some other processes than spreading activation. He labels these processes "sequential effects," as they have previously been called in the literature (Falmagne, 1965; Laming, 1968; Remington, 1969), and requires that they be explained in the standard way, by whatever reaction time model is appended to spreading activation models.

Compound cue theories could give two different accounts for the effects of nonword preprimes. The first is the same as for spreading activation theories. Sequential effects could be attributed to an appended reaction time model in which nonwords slow responses by changing response criteria. The second is more interesting and comprehensive. We have suggested (Ratcliff & McKoon, 1988) that sequential effects are not due to some separate process but are instead the result of compounding. So a nonword preprime will slow responses to a target because the familiarity value for a compound that includes a nonword will be low—lower than for a compound that includes an unrelated word preprime. This follows from the assumption that associations between nonwords and words are lower than associations between unrelated words. How much lower is a theoretical question and will depend on the weight given to the preprime compared with those for the prime and target. It may be that the difference in the priming effect for word and nonword preprime will be predicted to be small while at the same time an overall slowdown is predicted.

A nonword preprime will reduce the size of the priming effect for a related prime and target, because the values of prime–target familiarity are multiplied with the values of all combinations of preprime with prime and target, and these values are smaller for a nonword preprime than for a word preprime. However, how much the size of the priming effect is reduced depends on the relative weights given the preprime, prime, and target. It may be that the reduction in priming effect is small and unobservable compared to how much the nonword preprime slows responses overall. Moreover, the smaller priming effect will be measured against the slower overall baseline due to the nonword prime. A smaller priming effect against a slower baseline may appear to be the same size in milliseconds as a larger priming effect against a faster baseline. For example, a 30-ms priming effect on a baseline of 500 ms may, given current reaction time models (see Ratcliff, 1978), be equivalent to a 50-ms priming effect on a baseline of 700 ms. Unfortunately, there are currently no data to show exactly what these baseline effects might be for priming in lexical decision.

The assumption that compounding rather than an appended reaction time model accounts for sequential effects in reaction time has a precedent in the reaction time literature. This notion of compounding is similar to the linear model proposed for sequential effects in choice reaction time (e.g., Laming, 1973, Secs. 11.6–11.7). In the linear model, the subjective probability of a particular event is a continuous variable and depends on the previous sequence of stimuli; reaction time depends on this subjective probability. This assumption is similar to the notion that the compound cue tested at any point is a weighted average of prior items. In choice reaction time, it is clear from empirical data that there is a rapid decay of the influence of earlier items. For example, Laming (1968, Figure 8.11) shows that the effect of prior items in a sequence is roughly exponentially decaying as a function of position back in the sequence and that the effect has roughly dissipated by a lag of 2. Thus, the linear model is consistent with the lag effects observed in lexical decision priming studies.

In summary, the effects of a nonword preprime do not allow a clear discrimination between the compound cue and spreading activation models. To test compound cue models for these effects, we would need a model of how baseline changes affect the amount of priming. For spreading activation models, the appeal to sequential process would need some theoretical support from a specific reaction time model.

Conclusion

1. Whether the small priming effects obtained for weakly associated pairs such as *deer–vegetable* are problematic for spreading activation or compound cue theories turns on the issue of how these priming effects are to be predicted. We

have shown that they cannot be easily predicted from free-association production probabilities by any current model. Spreading activation theorists need to demonstrate how free association and priming effects can be jointly modeled, or they will need to find a new predictor variable that makes sense in the context of their theories. Compound cue theorists need more research to further document co-occurrence statistics and semantic relationships as predictor variables in the context of their theories.

2. Compound cue theories can accommodate priming effects over triples of three sequentially presented words, but their success in doing so depends on the weights given to the preprime, prime, and target in the calculation of familiarity for the response to the target. With the reasonable assumption that words are given significantly less and less weight as they increase in the distance with which they precede the target, SAM (Gillund & Shiffrin, 1984) can account for data presented by McNamara (1992).

3. When the preprime that precedes a prime and target is a nonword, responses to the target slow down (McNamara, 1992). Both spreading activation and compound cue theories can account for this finding. Spreading activation theories attribute the slow-down to sequential effects in whatever reaction time model would be appended to the spreading activation memory retrieval model (McNamara, 1992). Compound cue theories could use the same appended reaction time model explanation, or they could assume that the nonword, with its very low familiarity value, was combined with the prime and target.

Spreading activation was first proposed as a general retrieval mechanism by which the memory system could focus on a contextually relevant subset of all the information in memory and by which long pathways of connected information could be retrieved. The activation of items input to the system and items connected to them is intended to provide a focusing process, giving information that can be evaluated by subsequent decision processes or recycled to generate activation of additional information for recall processes. This spread of activation over distance from input information is the primary function of spreading activation. If spreading activation does not serve this function, then its utility is substantially diminished. Both the data reported here and earlier data (Balota & Lorch, 1986; de Groot, 1983; Ratcliff & McKoon, 1988) indicate that activation does not spread over any significant distance.

In contrast, compound cue theories use information in short-term memory to focus on appropriate subsets of information in long-term memory. The information in short-term memory is assumed to form a compound with which long-term memory is probed. The familiarity of the compound determines recognition decisions, and the compound is also used to generate retrieved information for recall tasks. Distance between concepts in memory is represented by the strengths of their mutual associations. In lexical decision, large priming effects reflect a high degree of familiarity of a compound (e.g., *baby–child*), and smaller priming effects reflect lower degrees of familiarity (e.g., *hospital–child*). The presence or absence of mediating concepts is irrelevant for the compound cue theories, because only directly associated pairs (or pairs with one mutually associated item in the Gillund–Shiffrin implementation, 1984) will produce an increment to familiarity in the models.

The compound cue theories and the results of the experiments reported in this article suggest that there are large numbers of weak direct associations in memory. The ubiquity of these associations is consistent with the way we were able to measure them in Experiment 3. Many pairs of words must co-occur more often than would be expected by chance, and identifying them is a matter of finding large enough and diverse enough databases. Experiment 3 provides the beginning of such an effort, using only a relatively small database from a relatively restricted source (the AP newswire). But even with this restricted database, over 300 words co-occur with words like *war* and *school* more often than would be expected by chance.

The compound cue view emphasizes that a word is understood in the context in which it is encountered (i.e., the information that co-occurs with it in short-term memory). In computational linguistics, this view has been summarized by the theme, "You shall know a word by the company it keeps" (Firth, 1957; cited by Church & Hanks, 1989). Hanks (1987) has pointed out that we can understand *bank* by its context *river, swim, boat* or *money, account, savings*. Similarly, we can know *housewife* by the different contexts *linoleum, baby*, or *career*. It should not be surprising that our long-term knowledge contains all of these different associations or that, in context, they are all familiar.

References

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Balota, D. A., & Lorch, R. F. (1986). Depth of automatic spreading activation: Mediated priming effects in pronunciation but not in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 336–345.

Church, K., & Hanks, P. (1989). Word association norms, mutual information, and lexicography. In *Proceedings of the 23rd Annual Meeting of the Association for Computational Linguistics*. Vancouver, British Columbia, Canada: Association for Computational Linguistics.

Cramer, P. (1968). Mediated priming of polysemous stimuli. *Journal of Experimental Psychology, 78*, 137–144.

de Groot, A. M. B. (1983). The range of automatic spreading activation in word priming. *Journal of Verbal Learning and Verbal Behavior, 22*, 417–436.

Dosher, B. A., & Rosedale, G. (1989). Integrated retrieval cues as a mechanism for priming in retrieval from memory. *Journal of Experimental Psychology: General, 2*, 191–211.

Duffy, S. A., Henderson, J. M., & Morris, R. K. (1989). Semantic facilitation of lexical access during sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 791–801.

Falmagne, J. C. (1965). Stochastic models for choice reaction time with applications to experimental results. *Journal of Mathematical Psychology, 12*, 77–124.

Firth, J. (1957). A synopsis of linguistic theory 1930–1955. In F. Palmer (Ed.), *Selected papers of J. R. Firth*. London: Longman.

Fischler, I. (1977). Semantic facilitation without association in a lexical decision task. *Memory & Cognition, 5*, 335–339.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both

recognition and recall. *Psychological Review, 91,* 1–67.
Grossberg, S., & Stone, G. (1986). Neural dynamics of word recognition and recall: Attentional priming, learning, and resonance. *Psychological Review, 93,* 46–74.
Hanks, P. (1987). Definitions and explanations. In J. Sinclair (Ed.), *Looking up: An account of the COBUILD project in lexical computing.* London: Collins.
Hintzman, D. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93,* 411–428.
Joordens, S., & Besner, D. (1992). Priming effects that span an intervening unrelated word: Implications for models of memory representation and retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 483–491.
Keppell, G., & Strand, B. (1970). Free-association responses to the primary purposes and other responses selected from the Palermo-Jenkins norms. In L. Postman & G. Keppel (Eds.), *Norms of word association* (pp. 177–240). San Diego, CA: Academic Press.
Laming, D. R. J. (1968). *Information theory of choice reaction time.* New York: Wiley.
Laming, D. R. J. (1973). *Mathematical psychology.* New York: Academic Press.
Light, L. L., & Carter-Sobell, L. (1970). Effects of changed semantic context on recognition memory. *Journal of Verbal Learning and Verbal Behavior, 9,* 1–11.
McKoon, G., & Ratcliff, R. (1979). Priming in episodic and semantic memory. *Journal of Verbal Learning and Verbal Behavior, 18,* 463–480.
McKoon, G., & Ratcliff, R. (1989). Assessing the occurrence of elaborative inference with recognition: Compatibility checking vs. compound cue theory. *Journal of Memory and Language, 28,* 547–563.
McKoon G., & Ratcliff, R. (1992). [Lexical decisions]. Unpublished raw data.
McNamara, T. P. (1992). Theories of priming. I. Associative distance and lag. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 1173–1190.
McNamara, T. P., & Altarriba, J. (1988). Depth of spreading activation revisited: Semantic mediated priming occurs in lexical decisions. *Journal of Memory and Language, 27,* 545–559.
Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review, 89,* 609–626.
O'Seaghdha, P. G. (1989). The dependence of lexical relatedness effects on syntactic connectedness. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 73–87.
Postman, L. (1970). The California norms: Association as a function of word frequency. In L. Postman & G. Keppel (Eds.). *Norms of word association* (pp. 241–320). San Diego, CA: Academic Press.
Postman, L., & Keppel, G. (1970). Norms of word associations. San Diego, CA: Academic Press.
Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85,* 59–108.
Ratcliff, R., Hockley, W. E., & McKoon, G. (1985). Components of activation: Repetition and priming effects in lexical decision and recognition. *Journal of Experimental Psychology: General, 114,* 435–450.
Ratcliff, R., & McKoon, G. (1978). Priming in item recognition: Evidence for the propositional structure of sentences. *Journal of Verbal Learning and Verbal Behavior, 17,* 403–417.
Ratcliff, R., & McKoon, G. (1981). Does activation really spread? *Psychological Review, 88,* 454–462.
Ratcliff, R., & McKoon, G. (1982). Speed and accuracy in the processing of false statements about semantic information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 16–36.
Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review, 95,* 385–408.
Ratcliff, R., & McKoon, G. (1989). Similarity information versus relational information: Differences in the time course of retrieval. *Cognitive Psychology, 21,* 139–155.
Ratcliff, R., & McKoon, G. (1992a). *Compound cue versus spreading activation accounts of priming.* Unpublished manuscript.
Ratcliff, R., & McKoon, G. (1992b). *Context effects in free association and lexical decision.* Unpublished manuscript.
Remington, R. J. (1969). Analysis of sequential effects in choice reaction times. *Journal of Experimental Psychology, 82,* 250–257.
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.
Shelton, J. R., & Martin, R. C. (1992). How semantic is automatic priming? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 1190–1209.
Stanovich, K. E., & West, R. F. (1981). The effect of sentence context on ongoing recognition: Tests of a two-process theory. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 658–672.
Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80,* 352–373.

(Appendix follows on next page)

Appendix
Materials Used in Experiment 3

Highly related free-association prime, high *t*-value prime, low *t*-value prime: target.

1. child, hospital, room: baby
2. children, young, father: kids
3. blade, kitchen, putty: knife
4. blue, night, fireworks: sky
5. brain, heat, radio: wave
6. ceiling, convention, manufacturer: floor
7. city, residents, flames: town
8. doctor, army, public: nurse
9. earth, earthquake, stake: ground
10. grow, power, growers: plant
11. foot, textile, workman: shoe
12. arm, left, amputation: leg
13. bake, piece, candles: cake
14. boy, death, love: girl
15. cars, fire, sound: trucks
16. country, newspapers, conscience: nation
17. crust, apple, cream: pie
18. memory, doubt, image: mind
19. green, acres, plane: grass
20. finger, cash, guard: hand
21. heal, bullet, blood: wound
22. house, vacation, morning: home
23. man, police, affair: woman
24. numbers, calls, protest: letters
25. play, war, season: games
26. priest, separation, mainstream: church
27. lamp, sales, glass: light
28. bed, hours, days: sleep
29. stomach, emergency, flowers: food
30. ocean, air, holes: water
31. door, bedroom, rain: window
32. justice, state, welfare: law
33. leaf, family, branch: tree
34. moon, movie, female: stars
35. music, theme, show: song
36. people, cheering, candidate: crowd
37. porthole, passenger, transport: ship
38. sickness, public, package: health
39. soldier, officer, protest: army
40. tobacco, black, passenger: smoke

Received June 14, 1991
Revision received April 13, 1992
Accepted April 20, 1992 ■

Carr Appointed Editor of the *Journal of Experimental Psychology: Human Perception and Performance*, 1994–1999

The Publications and Communications Board of the American Psychological Association announces the appointment of Thomas H. Carr, PhD, Michigan State University, as editor of the *Journal of Experimental Psychology: Human Perception and Performance* for a 6-year term beginning in 1994. As of December 15, 1992, manuscripts should be directed to

Thomas H. Carr, PhD
Department of Psychology
Michigan State University
East Lansing, Michigan 48824

Manuscript submission patterns for *JEP: Human Perception and Performance* make the precise date of completion of the 1993 volume uncertain. The current editor, James E. Cutting, PhD, will receive and consider manuscripts until December 14, 1992. Should the 1993 volume be completed before that date, manuscripts will be redirected to Dr. Carr for consideration in the 1994 volume.

Journal of Experimental Psychology:
Learning, Memory, and Cognition
1993, Vol. 19, No. 5, 1-13

0278-7393/93/$3.00

Discourse Models, Pronoun Resolution, and the Implicit Causality of Verbs

Gail McKoon, Steven B. Greene, and Roger Ratcliff

Some interpersonal verbs, such as *admire* and *amaze*, describe an action or property of one person (the reactor) that is necessarily a response to an action or property of another (the initiator). We hypothesized that these verbs make the initiator relatively more accessible in a comprehender's discourse model and that this change in relative accessibility aids identification of the referent of a pronoun in a subsequent *because* clause. We predicted that, as a result, subjects would be faster to recognize a character's name after a *because* clause that uses a pronoun to refer to that character than after one that refers to some other character. Four experiments confirmed this prediction. Three further experiments demonstrated the importance of the verb's causal structure and of the presence of the connective *because* to this result.

The use of psychological methods to study linguistic phenomena offers the possibility of simultaneous progress on issues in both fields. At least as far back as early empirical investigations of the derivational theory of linguistic complexity (e.g., Fodor & Bever, 1965; Fodor, Garrett, & Bever, 1968; Miller, 1962), psychologists have sought empirical evidence for hypotheses put forth by their colleagues in linguistics. The finding of such evidence both supports the linguistic hypotheses and allows the construction of models of underlying psychological processes that presumably rely on linguistic regularities.

In what follows, we describe the use of psychological methods to study the processes of pronoun resolution during comprehension of linguistic stimuli of special interest. These stimuli are of special interest because they employ verbs from a class exhibiting "implicit causality" (Garvey & Caramazza, 1974). We specify the nature of this implicit causality in greater detail later; for now, some illustrations will make this property clear. Consider the sentence frame "Mathilda amazed Jonathan because. . . ." When asked to complete a sentence frame of this form, subjects show great regularity in choosing to say something about Mathilda rather than about Jonathan. Note that either type of continuation is possible, for example, "because she displayed such refined talent" or "because he had never seen a fire-eater before." Garvey and Caramazza identified this type of implicit causality as NP_1 causality because the bias is to continue the sentence by saying something about the surface subject. Some verbs exhibit NP_2 causality instead, such as in "Felix admired Alexandra because. . . ," which most subjects will complete by describing a property of Alexandra's ("because she aced the accounting exam") rather than a property of Felix's ("because he was always in desperate need of a role model"). A number of verbs exhibit NP_1 causality; a number of others exhibit NP_2 causality. We discuss later the characteristics of these two groups of verbs.

Psychologists studying language have long been interested in how information conveyed by the main verb of a sentence contributes to the sentence's grammatical structure (e.g., Healy & Miller, 1971). More recently, their attention has focused on the particular issue of the implicit causality of verbs, which has been studied using a variety of tasks (Au, 1986; Brown & Fish, 1983; Caramazza, Grober, Garvey, & Yates, 1977; Ehrlich, 1980; Hoffman & Tchir, 1990; Hudson, Tanenhaus, & Dell, 1986). However, there has to date been no systematic, empirical demonstration that implicit causality is understood except under conditions in which subjects have been asked to engage in some explicit strategy; for example, they may be asked to generate a continuation for the sentence or to identify the antecedent of a pronoun by speaking it aloud. Whether implicit causality is understood in the absence of such specific strategies is still an open question. Ideally, we would like an empirical demonstration that implicit causality has an effect on comprehension, plus some method for measuring that effect. One promising place to look for an effect of implicit causality is in the processes that identify an argument of a verb as the referent for a subsequent pronoun because there is a widely accepted technique for studying these processes: comparing the accessibility of referents and nonreferents after pronouns are read (Chang, 1980; Corbett & Chang, 1983; Dell, McKoon, & Ratcliff, 1983; Gernsbacher, 1989; MacDonald & MacWhinney, 1990; McKoon & Ratcliff, 1980, 1984).

A demonstration of effects of a verb's implicit causality on pronoun resolution would be especially interesting in light of the difficulty of finding evidence of pronoun resolution in other contexts. Recently, Greene, McKoon, and Ratcliff

Gail McKoon, Department of Psychology, Northwestern University; Steven B. Greene, Department of Psychology, Princeton University; Roger Ratcliff, Department of Psychology, Northwestern University.

This research was supported by National Institute of Deafness and other Communicative Disorders Grant R01-DC01240 and Air Force Office of Scientific Research Grant 90-0246 (jointly funded by the National Science Foundation) to Gail McKoon and by National Institute of Mental Health Grants HD MH44640 and MH00871 to Roger Ratcliff.

We thank Beth Levin for discussions of this work.

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

(1992) proposed a framework in which to study pronoun processing. According to the Greene et al. framework, comprehenders construct a discourse model that represents the entities and events evoked by a discourse and the relationships among them (see Grosz, 1981; Grosz, Joshi, & Weinstein, 1983; Grosz & Sidner, 1986; McKoon, Ratcliff, Ward, & Sproat, in press; McKoon, Ward, Ratcliff, & Sproat, 1993; Sidner, 1983a, 1983b; Ward, Sproat, & McKoon, 1991; Webber, 1983). Each entity in the discourse model has some degree of accessibility relative to all other entities. The initial degree of accessibility of an entity is determined by the syntactic, semantic, and pragmatic means by which it is introduced, and its accessibility changes as comprehension of various syntactic and semantic structures alters the relationships represented in the model. The accessibility of an entity in a discourse model is therefore determined not only by the manner in which it is introduced into the discourse but also by subsequent references to it.

In this framework, the job a pronoun performs is seen not as a trigger that initiates a serial search for an antecedent (see Matthews & Chodorow, 1988) but as a cue to identify the discourse entity that best matches the semantic and grammatical features of the pronoun (see also Gernsbacher, 1989). Specifically, the identification of a referent for a pronoun is first attempted by a fast, automatic process that depends on the accessibility of the intended referent in the discourse model. This process matches the features of the pronoun in parallel against those of all entities in the discourse model. If one entity matches sufficiently well and better than all other entities, it is identified as the most likely referent of the pronoun. On the other hand, if either no referent matches sufficiently or more than one referent matches equally well, the comprehender may optionally engage in further, strategic, processing to identify the referent. A series of experiments by Greene et al. in which subjects read short (three-sentence) texts describing two equally salient characters found evidence of successful pronoun resolution only when subjects had extrinsic motivation to keep track of the characters and generous time in which to do so. In the absence of these factors, no evidence of pronoun resolution was found. The pronoun-as-cue framework explains this result: Because the two entities were equally salient, neither matched the pronoun sufficiently better than the other to be uniquely identified as its likely referent. On the basis of this evidence, Greene et al. argued that the processes responsible for pronoun resolution in previous psychological experiments (e.g., Chang, 1980; Corbett & Chang, 1983; Gernsbacher, 1989) may have been optional, strategic processes and not a mandatory component of comprehension.

In contrast to typical experimental materials that describe two characters who are equally in the focus of attention, natural discourse commonly uses a pronoun to refer to a discourse entity that is already highly salient, relative to other entities (Brennan, 1989; Chafe, 1974; Ehrlich, 1980; Fletcher, 1984; Greene et al., 1992; see also Givon, 1976). The occurrence of a pronoun usually indicates to the comprehender that the discourse is still centered on the previously salient entity or entities (Greene et al., 1992; Grosz et al., 1983). Numerous syntactic, semantic, and pragmatic devices can be used to establish one discourse entity as the current focus of attention and, therefore, as likely to be referred to subsequently (Gernsbacher, 1990; Gernsbacher & Shroyer, 1989; Grosz, 1981; McKoon, Ratcliff, Ward, & Sproat, in press; McKoon, Ward, Ratcliff, & Sproat, 1993; Sidner, 1983b; Ward et al., 1991). An utterance containing a verb exhibiting implicit causality may have the effect of establishing the verb's more prominent argument as the current focus of attention (Hudson et al., 1986). In terms of the pronoun-as-cue framework, these verbs may alter the relative accessibilities of their arguments in a discourse model. That change in accessibility may be sufficient to ensure that the fast, automatic process of pronoun resolution can provide one of them as the likely referent of a subsequent pronoun. If that is the case, then we may be able to find evidence of successful pronoun resolution even when the experimental procedures employed do not encourage subjects to engage in strategic processing.

Before turning to the empirical evidence, we examine in greater detail why some verbs exhibit the implicit causality that we hypothesize to privilege one possible referent over the other in a discourse model framework. Garvey and Caramazza (1974) coined the term *implicit causality* to describe a property of transitive verbs that relate two nouns referring to human or animate beings in such a way that "[o]ne or the other of the noun phrases is implicated as the assumed locus of the underlying cause of the action or attitude" (p. 460). Garvey and Caramazza argued that implicit causality is part of the semantics of the verb root: Some verbs, such as *confess*, *telephone*, and *approach*, assign the cause of the event to the subject noun phrase (NP_1), whereas others, such as *fear*, *praise*, and *admire*, assign the cause to the object noun phrase (NP_2). By examining subjects' completions of sentence frames such as "The prisoner confessed to the guard because he. . ," these researchers established that, when asked to do so, English speakers reliably attribute causality to NP_1 for some verbs and to NP_2 for other verbs.

A subsequent experiment (Caramazza et al., 1977) showed that subjects were faster to name the antecedent for a pronoun after reading a sentence containing a verb exhibiting implicit causality if that pronoun was consistent with the causality than if it was not. For example, when asked to identify the referent for *he*, subjects responded "Jimmy" faster after reading "Jimmy confessed to Mary because he wanted forgiveness" than they responded "Michael" after reading "Cathy confessed to Michael because he offered forgiveness."

Garvey and Caramazza (1974) identified the "locus of the underlying cause" as the relevant factor in determining a verb's implicit causality, but they stopped short of a full explanation of why that factor is critical and how one determines this locus. Following Au (1986; also Osgood, 1970), we discuss interpersonal verbs in terms of which of their arguments initiates a state of affairs and which one reacts to it. We use the term *interpersonal verbs* to refer to those verbs that describe a relationship between two people that has an essential psychological component: At least one of the people

must have some mental representation of the other. The implicit causality of a verb is toward the argument that initiates an action or evokes a response. As noted earlier, the subject of *confess* initiates the action: We confess for things we ourselves have done. In contrast, the subject of *thank* is reacting to a state of affairs brought about by the object: We thank others for things they have done. In one case, the grammatical subject is the initiator, and the object is the reactor; in the other, the object is the initiator, and the subject is the reactor. Note that the reactor may very well carry out some action, as in *thank*, as well as in *correct* and *congratulate*; the key is that the action is necessarily in response to an initiating state or action of someone else. Often the reactor's action is a speech act, but it need not be, as in *help*.

Levin's (in press) recent discussion of English verb classes supports the initiating–reacting distinction. Levin, summarizing earlier work in linguistics, classifies verbs of psychological states ("psych-verbs"), such as *amaze* and *admire*, into two categories, depending on whether the experiencer of some emotional reaction is the surface subject or object. She also describes another category, "judgment verbs," such as *congratulate*, *reproach*, and *scold*, which are like the *admire* psych-verbs in that the *admire* verbs "relate to a particular feeling which someone may have in reaction to something, [and] the judgment verbs relate to a judgment or opinion which someone may have in reaction to something" (p. 175). Thus, both the *admire* verbs and the judgment verbs indicate that the surface subject is experiencing some reaction at the initiation of the surface object. Levin's analysis of judgment verbs is reminiscent of Fillmore's (1971) analysis of the same verbs as presupposing responsibility on the part of the argument filling the role he labeled "defendant," generally the surface object.

The initiating–reacting distinction intuitively matches our understanding of implicit causality. Subjects' completions of *because* clauses reveal what aspect of the verb's meaning subjects believe requires a causal explanation. The initiating of a state of affairs typically demands an explanation; the reaction is explained by the state of affairs itself. Thus, *because* clauses should typically explain the behavior of the initiator, not the reactor.

In summary, verbs that exhibit implicit causality are those whose arguments fill the roles of initiator and reactor. Some property or action of the initiator causes a response by the reactor; this response may simply be an emotion (*admire*) or a perception (*notice*), or it may include an action (*thank*). A *because* clause will naturally then explain what property or action of the initiator provoked the response by the reactor. However, as Garvey and Caramazza (1974) first noted, it is, of course, possible for *because* clauses to offer an explanation in terms of a property or action of the reactor, as in "Cathy confessed to Michael because he offered forgiveness." In such an instance, in which the *because* clause is inconsistent with the implicit causality of the verb, the analysis requires an additional step. A property or action of the initiator still causes a response by the reactor, but the nature of the explanation offered by the *because* clause is different. In this case, the *because* clause explains what property or action of the reactor made the initiator's property effective or the initiator's action possible.

Although our analysis of implicit causality is compatible with current linguistic discussions of the argument-taking properties of verbs, it differs somewhat from that found in previous psychological work (e.g., Brown & Fish, 1983). Researchers since Garvey and Caramazza's original work have sometimes replaced their atheoretical NP_1/NP_2 classification scheme with one that distinguishes between "state verbs," which describe a situation in which one person (the stimulus) induces a psychological state in another (the experiencer), and action verbs, which describe a situation in which one person (the agent) instigates an action directed at another (the patient) (Brown and Fish, 1983). According to Brown and Fish's analysis, state verbs will exhibit implicit causality for NP_1 or NP_2, depending on which noun phrase refers to the stimulus. Action verbs, in contrast, should always exhibit implicit causality for NP_1, the agent, according to this analysis. However, Au (1986) found that although some action verbs, such as *cheat* and *flatter*, exhibit implicit agent causality, others, such as *correct* and *praise*, exhibit implicit patient causality. Au instead resurrected an earlier analysis of causal attribution, that of Osgood (1970), to explain the implicit causality of action verbs, while retaining the Brown and Fish analysis of state verbs.

Our conclusion is that the state–action distinction is superfluous to understanding implicit causality. Implicit causality has been found to be a property of some, but not all, verbs in both categories. Therefore, classifying a verb as belonging to either category tells little about whether that verb will exhibit implicit causality, and further, classifying a verb as an action verb tells nothing about which way the causality will go. No matter whether a verb is categorized as action or state, its semantics still must be further analyzed to predict its implicit causality. So for the purposes of the research described in this article, both state and action verbs are analyzed solely in terms of the initiating and reacting roles of their arguments to predict implicit causality.

Experiments 1–4

These experiments examine pronoun resolution in a *because* clause that follows a verb exhibiting implicit causality. Table 1 shows examples of the texts that were used in the experiments. Consider the first example in Table 1; in the third sentence, *infuriate* is a verb for which the subject—in this case, James—is the initiator. The subject does something or has some property that brings about a reaction by the object; in this case, the reaction is an emotion. The example shows two possible continuations of the third sentence: In the first, the *because* clause is consistent with the implicit causality of *infuriate*; in the other, it is inconsistent. Given our analysis of verbs exhibiting implicit causality and the pronoun-as-cue processing hypothesis, we can suggest how the two alternative continuations of the final sentence might be understood during comprehension. As a verb exhibiting implicit causality, *infuriate* makes the initiator, James, relatively more accessible than other entities in the discourse

Table 1
Examples of Experimental Texts

Verb category	Item
Subject initiating	James and Debbie were working on a political campaign together.
	They were both planning on pursuing careers in politics.
	James infuriated Debbie because
	(a) he leaked important information to the press.
	(b) she had to write all the speeches.
Object initiating	The boss had been giving Diane and Sam a hard time lately.
	Finally the two of them decided to do something about it.
	Diane valued Sam because
	(a) he always knew how to negotiate.
	(b) she never knew how to negotiate.

model of the text. In the first continuation, "he leaked important information to the press," the pronoun is intended to refer to James. When it is matched as a cue against the entities in the discourse model, the most accessible entity, James, is identified as the most likely referent. The gender of the pronoun is consistent with James as the referent, and perhaps more importantly, the information in the continuation is consistent with the implicit causality structure of the verb; it explains what state of affairs James created. The several factors of increased accessibility in the discourse model, gender agreement, and appropriateness of the continuation for the verb's causality all conspire toward identification of James as the referent for the pronoun.

In contrast, consider the second continuation, "she had to write all the speeches." The most accessible referent is still the initiator, James, but now the gender of the pronoun does not match. Moreover, the content of the continuation is inconsistent with the verb's implicit causality. The predicate explains what Debbie had to do in response to the state of affairs created by James, not what James himself did. Because of these mismatches, the initiator should be discarded as a potential referent. The remaining two possibilities are that pronoun resolution may fail, leaving the pronoun reference unresolved, or that the other, intended, referent—Debbie—may be selected.

The situation is similar for verbs for which the object is the initiator, like *value*, in the second example in Table 1. The object of *value* does something or has some property that brings about a reaction by the subject. Thus, *value* makes its object relatively more accessible in a discourse model. In the first continuation, "he always knew how to negotiate," which is consistent with the implicit causality of *value*, the pronoun is intended to refer to Sam, and the continuation explains what property of Sam's prompted Diane's reaction. So, when the pronoun is matched against the discourse model, Sam is identified as the most likely referent, and the matching gender and consistent continuation confirm this selection.

Once again, in the other continuation, "she never knew how to negotiate," the pronoun mismatches the most accessible entity on gender, and the information in the continuation is inconsistent with the causality implicit in the verb. The continuation explains what property of Diane's allowed her to appreciate the property of Sam's, and only indirectly what property Sam possessed. As with the inconsistent continuation of the subject-initiating verb *infuriate*, pronoun resolution may fail, or the only other potential referent, the reactor, may be selected.

All of the experiments described here compare subjects' reaction times to recognize a character's name as having appeared in the current text when the test occurred after the two types of continuations: those in which a pronoun refers to the tested character and those in which a pronoun refers to the other character. The test always occurred at the end of the third sentence of three-sentence texts like those in Table 1. Following the reasoning just outlined, for the character that was the referent of the pronoun in the consistent continuation (e.g., James in the first example in Table 1), we anticipated that responses to that character's name would be facilitated when it was tested after the consistent continuation relative to the inconsistent continuation; that is, responses would be facilitated for the name when that character was the referent versus when it was not. We refer to this as a matching effect: Responses to a character's name are facilitated when that character matches the referent of the pronoun versus when it does not.

However, for the character intended as the referent in the inconsistent continuation, two outcomes are possible. In this case, the processes of pronoun resolution may leave the reference unresolved, resulting in no matching effect but perhaps overall facilitation for the initiator because of its initial greater accessibility. Or, if the pronoun resolution process does not fail but instead selects the other character, the reactor, as the referent for the pronoun, we would again expect facilitation for the character referred to by the pronoun, in this case, the reactor. We would therefore expect a matching effect such that responses are facilitated when the character whose name is presented for recognition matches the referent of the pronoun in the continuation.

Experiments 1 and 2 examine subject-initiating verbs, like *infuriate*, and Experiments 3 and 4 examine object-initiating verbs, like *value*. These experiments were designed to examine pronoun resolution under conditions in which subjects read at approximately normal rates without adopting any special strategies. The materials were presented at a rate of about 250 ms/word, a rate that other research (e.g., Dell et al., 1983; Greene et al., 1992, Experiments 8 and 9; Just & Carpenter, 1980; Rayner, 1978) has shown to be reasonable for college students. Comprehension questions following the texts asked about a variety of information from the texts; they did not ask about specific kinds of information, such as which character carried out particular actions, so as not to induce subjects to adopt strategies specific to pronoun resolution (or any other task beyond that required by the experimental procedure directly). Finally, three times as many filler items as critical items were included in the experiments in order to reduce the

predictability of the type of item to be tested and the test locations.

Method

Materials. Twenty subject-initiating verbs and 20 object-initiating verbs were chosen from those used in previous research (Au, 1986; Brown & Fish, 1983). Because we selected only verbs that were subject or object initiating according to our analysis of implicit causality, we excluded some verbs, such as *telephone* and *hit*, that had been included in previous research. The subject-initiating verbs we selected were *aggravate, amaze, amuse, annoy, apologize, bore, charm, cheat, confess, deceive, disappoint, exasperate, fascinate, frighten, humiliate, infuriate, inspire, intimidate, scare,* and *surprise*. The object-initiating verbs were *assist, blame, comfort, congratulate, correct, detest, dread, envy, hate, help, jeer, notice, pacify, praise, reproach, scold, stare, thank, trust,* and *value*. The implicit causality of these verbs can be demonstrated by asking subjects to generate continuations of sentence fragments that present the verbs in the following frame: proper noun, verb (tense), proper noun, *because* (e.g., "James infuriated Debbie because _____"). Continuation data were collected for some of the 40 verbs used in our experiments by Au (1986), and we collected continuation data for the others. Overall, the mean percentage of subjects continuing a sentence fragment with a pronoun referring to the referent consistent with the causality of the verb was 89 for the subject-initiating verbs and 92 for the object-initiating verbs.

Each verb was used in the third sentence of a three-sentence text. The first sentence of each text introduced two characters, one male and the other female, and the third sentence mentioned these characters again by name. The second sentence referred to both of them by anaphora (usually *they*). For half of the texts, the first-mentioned character in both the first and third sentences was male, and for the other half, female. The critical verb was used in the first clause of the third sentence. The two clauses of the third sentence were always joined by *because*. There were two versions of the second clause of the third sentence: One version began with a pronoun matching the gender of the first character in the first clause and continued with information that made sense for that character in a causal role; the second version began with a pronoun matching the gender of the other character and continued with information that made sense for that character. An example of a text for a verb with each kind of implicit causality is shown in Table 1. The average length of the first and second sentences combined was 19.8 words, and the average length of the third sentence was 10.9 words. The average number of words between the first character's name in the first clause of the third sentence and the pronoun in the second clause was 3.2; the average number of words between the second character's name and the pronoun was 1 (*because*), and the average number of words between the pronoun and the end of the sentence was 5.7. There were two test words for each text, the two character names. There were also two test statements for each text, one true and one false. These tested a variety of kinds of information from the texts.

There were 60 filler texts used to provide different kinds of test words from the experimental texts. These texts were all three sentences long and averaged 33 words in length. Each text had 1 test word. Thirty-five of these test words had not appeared in any text (17 of these were proper names), and 25 had appeared in their text. Nineteen were tested in the first two sentences, and the remainder were tested in the third sentence. Each filler text had associated with it one true and one false test statement; as with the experimental texts, these were written to test a variety of kinds of information from the texts.

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

Each experiment began with 30 lexical decision test items. These items were included to give subjects practice with the response keys on the computer keyboard. After this practice, there were 20 filler texts, and then the remainder of the texts—20 experimental (20 subject-initiating texts in Experiments 1 and 2, and 20 object-initiating texts in Experiments 3 and 4) and 40 fillers—were presented in random order.

Each text began with the instruction to press the space bar on the keyboard to initiate the text. When the space bar was pressed, the text was presented, one word at a time. Each word was displayed in the same location on the CRT screen, and each was displayed for 170 ms plus 17 ms multiplied by the number of letters in the word. There was no pause between words. The last word of a sentence was displayed for an extra 200 ms unless it was immediately followed by a test word. When a test word was presented, it appeared in the same location as the text words; its letters were all in upper case (unlike the words of the text) and two asterisks were displayed immediately to its left and to its right. The test word remained on the screen until a response key was pressed (?/to indicate the word had appeared in the text, and z to indicate the word had not appeared in the text). In Experiments 1 and 3, after the response and a pause of 170 ms, the text continued or the PRESS SPACE BAR message for the true–false sentence was presented. In Experiments 2 and 4, if the response was slower than 1,100 ms, the message TOO SLOW! was displayed first for 500 ms. We used the response time feedback to encourage very fast responses, in order to be sure that the pattern of results obtained in Experiments 1 and 3 could be replicated under speed conditions, and so that we could be sure that decisions about the test words were not based on slow, strategic processes that began at the time of presentation of the test word. In all the experiments, each text was followed by a true–false test statement, and incorrect responses to this test statement were followed by an error message, the word ERROR, presented for 1,500 ms. Each text had a true and a false test statement; which one of these was presented was chosen randomly. For the test words, subjects were instructed to respond as quickly and accurately as possible. For the true–false test statements, they were told to aim for 100% accuracy.

Design and subjects. For all four experiments, there were two variables for the 20 experimental texts: The pronoun in the second clause of the third sentence matched in gender either the first or the second character in the first clause, and the test word was the name of either the first character or the second. Note that the consistent pronoun refers to the first character name for the subject-initiating verbs and to the second character name for the object-initiating verbs. For the experimental texts, the test word was always presented after the final word of the text. The four conditions formed by crossing the two variables were combined in a Latin square design with four sets of texts (5 per set) and four groups of subjects (5 in each group except for Experiment 2, in which there were 7 in each group). The subjects participated in the experiments for credit in an introductory psychology course at Northwestern University.

Results and Discussion

Means were calculated for each subject and each item in each condition, and means of these means are shown in Table 2. All response times longer than 2,000 ms were eliminated from the means and analyses. For Experiments 1 and 3, this was about 4% of the data, and for Experi-

Table 2
Results of Experiments 1–4: Response Times (RTs) and Error Rates

	Subject-initiating verbs			
	Experiment 1		Experiment 2	
	RT	% errors	RT	% errors
Test first character				
Consistent continuation (referent matches test)	1,005	5	776	7
Inconsistent continuation (referent does not match test)	1,083	0	780	5
Test second character				
Consistent continuation (referent does not match test)	1,130	2	835	6
Inconsistent continuation (referent matches test)	1,060	2	795	4
	Object-initiating verbs			
	Experiment 3		Experiment 4	
Test second character				
Consistent continuation (referent matches test)	993	2	733	5
Inconsistent continuation (referent does not match test)	974	1	764	4
Test first character				
Consistent continuation (referent does not match test)	1,008	9	784	12
Inconsistent continuation (referent matches test)	957	3	735	5

ments 2 and 4, this was less than 1% of the data. Response times for filler test words and true–false test statements are shown in Table 3 for all the experiments. Table 3 also shows the standard errors of the means for the experimental conditions of each experiment.

Examination of the data in Table 2 shows that the choice of pronoun used in the text had a strong effect on response times to the test words. Consider, for example, responses to the first character's name in Experiment 1. The first character was referred to by the pronoun in the consistent continuation, and responses for the first character's name were faster following the consistent continuation than the inconsistent continuation. In other words, responses to the test word were faster when the referent of the test word matched the referent of the pronoun than when it did not. A similar matching effect was obtained when the second character name was presented as a test word: When it matched the antecedent of the pronoun, responses were faster than when it did not match. We interpret the matching effect as showing that the subjects in these experiments understood which of the two characters in a text was the intended referent of the pronoun, in contrast to previous experiments in which they did not (Greene et al., 1992).

We had predicted the matching effect for the character in the initiator role: The causal structure of the verb should make this character more accessible in the discourse model, and the consistency of the information in the *because* clause with that character as the referent for the pronoun should

Table 3
Response Times (RTs) and Error Rates for Filler Test Words and True–False Test Sentences and Standard Errors of the Means

	Positive test words		Negative test words		True test sentences		False test sentences		
Experiment	RT	% errors	RT	% errors	RT	% errors	RT	% errors	SE_M
1	1,253	12	1,237	4	2,437	7	2,259	12	22
2	932	24	890	9	2,086	8	2,060	12	10
3	1,141	16	1,094	3	2,240	9	2,181	15	19
4	888	22	857	9	1,982	8	1,987	17	17
5	1,071	13	1,028	8	2,162	7	2,076	13	18
6	1,083	14	1,030	5	1,999	8	1,987	15	16
7	1,128	16	1,074	4	2,050	8	1,962	12	14

Note. Response times are in ms. Standard errors refer to the error in the means of the experimental conditions tested by analysis of variance.

Crop rules in CM

facilitate responses to that character's name as a test word. However, we were unsure about whether there would also be a matching effect for the character in the reactor role: A continuation that was inconsistent with the verb's causal structure would have to lead to a rejection of the most accessible possible referent (the initiator) and also lead to enough further processing to identify the reactor as the pronoun's referent. The fact that we did obtain the matching effect for the character in the reactor role indicates that this processing did occur. The failure of a *because* clause to be consistent with the causal structure of the verb, combined with the mismatch in gender between the pronoun and referent, is apparently sufficiently salient to invoke the extra processing required to identify the reactor as the referent.

One caveat about the interpretation of the pattern of data is in order. It should be clear that we have no measure of a neutral baseline for response times to our recognition tests of the characters' names following the texts. In the experiments in Greene et al., we used sentences like "Mary accidentally scratched John with a knife and then she dropped it on the counter." We measured the response time to a character's name both before and after the pronoun in the second clause of its sentence, so that we could examine the relative facilitation given by the pronoun to its referent versus a nonreferent. Whether any obtained facilitation was due to true facilitation for the referent or inhibition for the nonreferent is impossible to determine. Similarly, in the experiments reported here, we compared whether the response time to a character's name at the ends of the sentences changed as a function of whether the character matched the referent of the pronoun in the sentence, but whether that change was facilitation for a referent or inhibition for a nonreferent is impossible to say. Because we were concerned only with relative effects, this is not a serious problem. Our claim is only that the matching effect represents a relative change in the accessibilities of the referent versus the nonreferent.

The lack of a neutral baseline also makes it inappropriate to compare reaction time for one character's name as a test word to reaction time for another character's name as a test word. Because we have no a priori measure of the relative accessibility of the two characters, that comparison would give us no basis on which to conclude that the process of pronoun resolution differentially affected the accessibility of the two characters. The only comparison permitted by the present data concerns whether the consistent and inconsistent continuations differentially affect the accessibility of the same character; this is the comparison revealed in the matching effect.

The matching effect held for both subject-initiating verbs and object-initiating verbs, as well as for subjects who were pressed to respond quickly (by the TOO SLOW! message) and those who were not, with one exception. For the subject-initiating verbs tested with the TOO SLOW! message (Experiment 2), the test word referring to the referent of the consistent pronoun did not show a matching effect. In this one case, response times did not appear to slow significantly when the referent of the test word did not match the referent of the pronoun, and this result suggests that pronoun resolution may be somewhat less robust with subject-initiating verbs than with object-initiating verbs.

The matching effect in each experiment represents an interaction between the character name that was tested and the pronoun that was used in the sentence. The significance of the interactions was demonstrated by analyses of variance (ANOVAs) that treated subjects as the random variable (F_1) and analyses that treated items as the random variable (F_2). For Experiment 1, $F_1(1, 19) = 12.2$ and $F_2(1, 19) = 7.4$; for Experiment 2, $F_1(1, 27) = 5.8$ and $F_2(1, 19) = 5.8$; for Experiment 3, $F_1(1, 19) = 6.8$ and $F_2(1, 19) = 5.0$; and for Experiment 4, $F_1(1, 19) = 6.0$ and $F_2(1, 19) = 8.0$, all $ps < .05$. With one exception noted later, no other reaction time effects approached significance in either subjects or items analyses. Standard errors of the response time means are shown in Table 3 (for all experiments). Error rate differences were also tested by ANOVAs, and all F values were not significant ($p > .05$, Fs less than 3.1), again with one exception discussed later.

Our main hypothesis was that verbs exhibiting implicit causality initially would make the character in the initiator role more accessible than the character in the reactor role and that this difference in accessibility should facilitate pronoun resolution. But, in addition, some effect of the initial greater accessibility of the character in the initiator role might survive to the end of the sentence. Consistent with this expectation, reaction times were faster to the first test word, which referred to the initiator, than to the second test word in Experiment 2, $F_1(1, 27) = 14.2$ and $F_2(1, 19) = 5.8$, $ps < .05$. Also, in Experiment 3, significantly fewer errors were made on the second character (the initiator) as a test word than on the first, $F_1(1, 19) = 5.9$ and $F_2(1, 19) = 4.1$, $ps < .05$. In addition to these significant effects, the nonsignificant tendencies for reaction times to be faster to test words that referred to initiators than to those that referred to reactors in Experiments 1 and 3 are consistent with our hypothesis that verbs exhibiting implicit causality make the initiator more accessible than the reactor.

Experiments 5 and 6

Experiments 1–4 demonstrated a matching effect in reaction time for responses to a recognition test of a character's name such that responses to a test of a character's name were facilitated if the character matched the referent of the preceding pronoun. We have hypothesized that this happened because the structure of verbs exhibiting implicit causality "privileges" the initiator role over the reactor role as a potential pronominal referent. If the gender of the subsequent pronoun and the information in the continuation following the pronoun are consistent with the implicit causality of the verb, the character in the initiator role is taken to be the pronoun's referent, as demonstrated by the matching effect observed for the initiator in Experiments 1–4. If, however, the gender of the pronoun and the information in the predicate are inconsistent with the potential referent privileged by the verb's implicit causality, this mismatch causes the other character, the reactor, to be selected as the

referent of the pronoun, as demonstrated by the matching effect for the reactor. For both initiator and the reactor, the result is the same: faster recognition responses to a character's name if that character matches the referent of the pronoun in the continuation.

Our account of the n.. ning effects found in Experiments 1–4 emphasizes the importance of consistency between the verb's causal structure and the explanation of the verb's action given in the *because* clause. The relationship between the two is made explicit by the word *because*. This connective may serve to bring to the fore the information about implicit causality inherent in the verb's lexical structure. Experiments 5 and 6 examine whether the presence of this connective is necessary to create the effect observed in Experiments 1–4.

Method

Experiment 5 examines subject-initiating verbs, and Experiment 6 examines object-initiating verbs. The 20 texts for the subject-initiating verbs and the 20 texts for the object-initiating verbs were each modified so that the final, two-clause sentence became two sentences with *because* deleted. This was the only change made to the materials. For example, the final sentences for the first text in Table 1 were changed to: "James infuriated Debbie. He leaked important information to the press," and "James infuriated Debbie. She had to write all the speeches." As these examples suggest, it is still possible, or even likely, that comprehenders will interpret the information in the second sentence as a reason for the action in the first sentence. However, the relation is not made explicit in the text; instead comprehenders must make what Clark (1977) refers to as a bridging inference. We hypothesized that less causally explicit materials might adversely affect pronoun resolution, causing the matching effect to be reduced or to disappear altogether. Of course, splitting the two clauses of the original version of the sentence into two separate sentences would in all likelihood alter subjects' comprehension processes and might also modify discourse relations in ways beyond simply making the causal relationship less explicit, but we lack a sufficiently thorough understanding of discourse representation to predict such changes with any precision. Hence, interpretation of null results from this experiment would of necessity be tentative.

In displaying the two final sentences, the words were presented as in the previous experiments, and there was an additional 200-ms pause after the final word of the first of the two sentences. In all other respects, the experimental procedures and materials were the same as in the previous experiments. (There were no TOO SLOW! messages.) The test words for the experimental texts were always presented at the end of the final sentence of their text. There were the same two variables as in the previous experiments: The final sentence used either the consistent or the inconsistent pronoun, and the test word was either the first character's name or the second character's name. These four conditions were combined in a Latin square design, with 28 subjects in each experiment.

We also collected continuation data on these new materials. We wondered whether the same preference to refer to either the surface subject or the surface object shown in continuations with *because* sentences would also appear without the *because* connective. For the continuation study, we modified the two final sentences of each text so that they used two names of the same gender, and we presented them in this frame: proper name, verb (tense), proper name, pronoun (e.g., "James infuriated Sam. He _____"). Subjects were asked to continue the second sentence, and their continuations were scored according to whether the content indicated that the pronoun had been interpreted as referring to the first character or the second. The texts were divided into two sets, each with half subject-initiating verbs and half object-initiating verbs randomly ordered, and 42 subjects gave continuations for each set. For the subject-initiating verbs, the probability of a continuation indicating that the pronoun had been interpreted according to the causality of the verb was high, .88, as it had been with the connective *because*. However, for the object-initiating verbs, the preference was no longer evident; the probability of a continuation indicating interpretation of the pronoun according to the causality of the verb was only .39. These proportions most likely indicate a preference for a subsequent sentence to refer to the surface subject of a preceding sentence.

Results and Discussion

The data were analyzed as for the previous experiments (with responses slower than 2,000 ms, less than 2%, eliminated), and means are shown in Table 4.

The only difference between these two experiments, 5 and 6, and Experiments 1 and 3 was that the connective *because* was deleted, turning the two-clause final sentences of Experiments 1 and 3 into two separate sentences in Experiments 5 and 6. This difference eliminated the matching effect completely; in Experiments 5 and 6, response time for a test word was *not* affected by whether or not its referent matched the intended referent of the pronoun that preceded it. In fact, the only effect in response times was that, for the object-initiating verbs, responses to the first character name (the name that the pronoun would not be expected to match) were

Table 4
Results of Experiments 5 and 6: Response Times (RTs) and Error Rates

	RT	% errors
Experiment 5: Subject-initiating verbs		
Test first character		
Consistent continuation (referent matches test)	934	2
Inconsistent continuation (referent does not match test)	918	1
Test second character		
Consistent continuation (referent does not match test)	921	2
Inconsistent continuation (referent matches test)	917	1
Experiment 6: Object-initiating verbs		
Test second character		
Consistent continuation (referent matches test)	880	3
Inconsistent continuation (referent does not match test)	887	3
Test first character		
Consistent continuation (referent does not match test)	938	5
Inconsistent continuation (referent matches test)	951	5

slower than responses to the second character name. This effect was significant, $F_1(1, 27) = 4.5$ and $F_2(1, 19) = 5.5$, $ps < .05$. All other Fs, for both experiments, were less than 1.0. There were no significant effects on error rates, $Fs < 1.5$.

Clearly, the presence of the connective *because* contributes to successful pronoun resolution in a dependent clause that follows a verb exhibiting implicit causality. This finding suggests that the lexical structure of the verb and the information contained in the sentence continuations are not sufficient either alone or in combination to bring about successful pronoun resolution. Of course, altering our texts to change the final sentence into two sentences by simply deleting the connecting *because* may have altered discourse relations in other ways as well, so any interpretation of the results of Experiments 5 and 6 must be viewed with caution.

Experiment 7

Experiments 1–4 found evidence of facilitation for a test word whose referent matches the referent of the preceding pronoun in a *because* clause following verbs that exhibit implicit causality. Experiments 5 and 6 suggested that the *because* connective is critical to this matching effect. This suggests a further possibility to be examined: Perhaps the presence of *because* is not only necessary but, in fact, sufficient to create the effect. The results obtained in Experiments 1–4 were obtained using materials with *because* connectives; earlier failures to find similar evidence of pronoun resolution used materials with no *because* clauses (Greene et al., 1992). This final experiment examines whether adding *because* clauses to those earlier materials might allow us to find evidence of pronoun resolution.

Method

Materials. The 32 experimental texts were modified from texts previously used by Greene et al. (1992). An example text is shown in Table 5. Each text was made up of three sentences, with the first sentence introducing two characters of different genders and the second sentence referring to both of them anaphorically. There were two versions of the third sentence, each made up of two clauses connected by *because*. The first clause was the same in both versions and mentioned both characters by name, in the same order as in the first sentence. The first name was the subject of the verb in this clause; the second name was usually a direct or indirect object.

Table 5
Example of Paragraphs from Experiment 7

Sentence	Conclusion
Mary and John were doing the dishes after dinner. One of them was washing while the other dried.	
Mary accidentally scratched John with a knife because	she was so tired and clumsy.
	he suddenly grabbed for a glass.

The verb constructions used in these sentences were, approximately: *scratched, shot at, was being tickled by, tried to catch, saw, read something to, went to visit, threw something at, aimed something at, stole something from, poured something for, saw, broke something playing with, watched, appreciated something from, tried to amuse, tried to cook something for, watched, wanted to call, was playing something for, took over something from, drove, edited something for, made something for, searching for something for, waited to see, tried to repair for, counted something gotten from, was drawing a picture of, heard something about, borrowed something from,* and *started writing to.* None of these verbs fit our analysis of verbs that exhibit implicit causality. One of the second clauses of the final sentence referred to the first character with a pronoun and continued with information consistent with that character in a causal role. The other second clause referred to the second character with a pronoun and continued with information consistent with that character. The mean number of words in the first two sentences was 18.2; the mean number of words in the third sentence was 14.0. The mean number of words between the first character's name in the third sentence and the pronoun was 7.1, and between the second character's name and the pronoun, 2.2. The mean number of words between the pronoun and the end of the sentence was 4.9. There were two test words for each text, the two character names. There was one true–false test statement for each text; half were true and half false. The same filler texts were used as in the previous experiments.

We collected continuation data for the final sentences of these texts in the same way as for the texts used in Experiments 1–4. The first clause of each final sentence plus the word *because* was presented as a sentence fragment for subjects to complete (e.g., "Mary accidentally scratched John with a knife because _____"). Each fragment was completed by at least 32 (or as many as 45) subjects. The mean proportion of continuations that referred to the first character name (out of all continuations that referred to one or the other of the characters) was .46. The variability across items was high, but conditionalizing the response time data (given later) on the relative proportions of continuations did not yield any meaningful differences in the patterns of response times.

Procedure, design, and subjects. The procedure in Experiment 7 was the same as for Experiments 1 and 3. There were two variables in the design: The second clause of the final sentence used a pronoun intended to refer either to the first or to the second character mentioned in the first clause, and the test word was either the first character's name or the second character's name. These four conditions were combined in a Latin square with the 32 texts and 24 subjects (from the same population as the previous experiments).

Results

The data were analyzed in the same way as in the previous experiments, and the means are shown in Table 6. Response times longer than 2,000 ms were eliminated (less than 1% of the data).

The main result is that there was no matching effect. Response time for a test word did not depend on whether the test word's referent matched the intended referent of the pronoun that preceded it. Instead, response times were slower for the first character's name than the second character's name, whichever pronoun was used. This effect was significant, $F_1(1, 23) = 4.8$ and $F_2(1, 31) = 4.8$, $ps < .05$. Other Fs for response times were less than 1.0. There were also more errors on the first character's name, $F_1(1, 23) = 8.1$ and $F_2(1, 31) = 4.3$, $ps < .05$. For errors, the interaction between the

Table 6
Results of Experiment 7: Response Times (RTs) and Error Rates

	RT	% errors
Test first character		
Referent matches test	975	5
Referent does not match test	978	2
Test second character		
Referent does not match test	947	1
Referent matches test	930	2

pronoun and test word variables approached significance in the subjects' analysis, $F_1(1, 23) = 3.6$, $p < .05$, and was significant in the items' analysis, $F_2(1, 31) = 5.4$, $p < .05$. The other *F*s for the errors analysis were less than 2.0.

It is worth repeating here that conditionalizing the response time data on the continuation data did *not* yield a meaningful pattern of results. Neither in this experiment nor in Experiments 5 and 6 could failures to find a matching effect be predicted from continuation probabilities. In Experiments 5 and 6, subjects were likely to continue a sentence containing a subject-initiating verb with a pronoun referring to the subject character, but there was no matching effect. They were not particularly likely to continue a sentence containing an object-initiating verb with a pronoun referring to the object, and there still was no matching effect. The implication of these results is that, while continuation data may sometimes be helpful in eliciting subjects' intuitions, they cannot take the place of other kinds of tests of comprehension.

General Discussion

The lexical representation of interpersonal verbs exhibiting implicit causality guides comprehension of sentences that use those verbs. These verbs entail a psychological relationship between the initiator and the reactor, at least one of whom must have some mental representation of the other. We have argued that the lexical representations of these verbs call for arguments that satisfy the roles of initiator and reactor. The verbs attribute some action or emotion to the reactor that is necessarily a response to a state of affairs for which some action or property of the initiator is the cause. For some verbs, the initiator appears in the subject position in the surface structure of a sentence and the reactor appears in the object position; for others, the surface position of the roles is the reverse. In both cases, the relative accessibility of the initiator in the discourse model constructed during reading is increased. Additionally, because the verbs express an action or state of affairs brought about by the initiator, it is natural for a *because* clause following the verb to explain the initiator's behavior. The increased accessibility of the initiator, the natural fit of the explanation of the verb's lexical structure, and the use of the connective *because* together support pronoun resolution in sentences in which a verb exhibiting implicit causality is followed by an explanatory clause consistent with it. In the sentence "John blamed Mary because she forgot the wine," the action of blaming is initiated by Mary (something she did), and the reason that she brought about blaming is that she forgot the wine. Mary is more accessible than John, it is natural to explain how she caused blaming, and *because* makes the causal relation explicit; these factors together support identification of Mary as the referent of the pronoun. In contrast, for the sentence "John blamed Mary because he was in such a bad mood," the gender of the pronoun is not consistent with the more accessible of the two characters, and the explanation of the blaming action does not immediately fit with the implicit causal structure of the verb. These factors work against identification of Mary as the referent of the pronoun and support the alternative referent, John.

Although we have classified the 40 verbs used in our studies as verbs exhibiting implicit causality, it is important to understand that such a classification is only our best first effort. Some of the 40 verbs may fit into the implicit causality class better than others, and undoubtedly other verbs that we did not consider rightfully belong in the class. Furthermore, implicit causality is only one of many dimensions along which verbs might be classified; when other dimensions are considered, the class of verbs exhibiting implicit causality may break apart into a variety of other classes (see Levin, in press). We have adopted the simplifying assumption that these other dimensions do not interact, for the purposes of our experiments, with implicit causality.

Our data support the proposed analysis of verbs exhibiting implicit causality by showing a matching effect: Both when the *because* clause was consistent with a verb's causality and when it was inconsistent, responses to a character's name as a test word were faster when the character was the referent of the pronoun than when it was not. There are at least two possible ways to describe the decision process that leads to this difference in response times. One possibility is that the test word is matched against the already existing representation of the sentence in memory, and response time and accuracy for the test word reflect its accessibility in that representation. In this case, the test word does not modify the existing representation, and the information provided by the test word interacts with information in the text only in ways that produce no new information about the text. A second possibility is that the test word is used as additional information in that it changes the text representation (Forster, 1981). In terms of our experiments, this could mean that the pronoun's referent had not yet been completely identified before the test word was presented, but that when the referent's name was presented as a test word, subjects at that point matched it against the pronoun and the discourse representation to identify that character as the referent. Of course, presenting the referent's name as a test word does not add any really new information; the name is already in short-term memory because it was just mentioned in the preceding clause (Clark & Sengul, 1979). However, presenting it as a test word could, for example, add to that character's accessibility sufficiently that pronoun resolution could succeed when it had not already. If correct, this second possibility

would make the pronoun resolution that appears in our experiments critically dependent on the presence of the test word. In striking contrast, pronoun resolution in previous experiments (Greene et al., 1992, Experiments 1, 2, 3, 4, and 7) could not have been dependent on the presence of a test word; in those experiments, there was no evidence that the referents of pronouns were identified at all.

The experiments reported by Greene et al. (1992) used sentences like "Mary accidentally scratched John with a knife and then she dropped it on the counter." The main verbs in these sentences do not have implicit causality as a central part of their lexical representations. (See Levin, in press, for a discussion of *scratch*, for example.) Therefore, we suggested, they do not privilege one of their arguments over the other. When discourse models are constructed during reading for sentences like these, the two arguments are not differentially accessible, and the second clause is not naturally attributed to one argument or the other by the structure of the verb. When a pronoun in the second clause is matched against the discourse model, the two arguments do not differ in accessibility, and the pronoun is not identified as referring to one or the other of them. If no referent is identified for a pronoun, then the information predicated of the pronoun is not differentially associated with one character in the discourse representation rather than others.

The results presented here suggest that one way a discourse can support pronoun resolution is by using a verb that increases the accessibility of one possible referent more than that of another and by attributing to the pronoun's referent information that fits naturally with the meaning of the verb. In these circumstances, and possibly in others, pronoun resolution may even be a mandatory component of comprehension (Gernig, 1986). In contrast, as was the case with the materials used by Greene et al. (1992, Experiments 1–7), when a discourse does not support the identification of a unique referent for a pronoun, either because no referent is sufficiently accessible or because several possible referents are all equally accessible, then special goals or strategies may be required. In some of the experiments reported by Greene et al., the procedure was almost identical to that used in the experiments reported in this article: a reading speed normal for college undergraduates (Greene et al. used a constant 250 ms/word pace, compared with the 170 ms/word plus 17 ms per letter we used), and no specific task requiring subjects to identify pronominal referents. The data showed no evidence that unique referents for pronouns were identified. Evidence of pronoun resolution appeared only when test locations were made highly predictable by using just one-sentence texts, when subjects were motivated by a specific task that required pronoun resolution, and when they were given ample time to accomplish the resolution process during reading by presenting the words of the sentences at a rate of about 500 ms each.

As we and others have noted, in natural discourse, pronouns are typically used when only one entity is already highly salient in the comprehender's discourse model (Brennan, 1989; Chafe, 1974; Ehrlich, 1980; Fletcher, 1984; Greene et al., 1992). Use of verbs that exhibit implicit causality is only one of many ways in which natural discourse may make one entity more salient than others, and thereby support pronoun resolution. A variety of other devices may also be used to increase the accessibility of one entity: the cataphoric *this* ("This man walks into a bar. . . ." Gernsbacher & Shroyer, 1989); cleft sentences ("It was Umberto who. . . ." Sidner, 1983b); repetition of a full noun phrase ("Number thirty passes to forty-one. Forty-one shoots, and he misses," Brennan, 1989); and spoken stress (Brennan, 1989). In short, many devices of natural discourse allow it to be designed precisely so that pronoun resolution can be accomplished without requiring any specific strategy on the part of the comprehender. We discuss the process of pronoun resolution here, as in Greene et al., not in terms of what the pronoun does to trigger a search for its referent, but instead in terms of what the discourse does to make such a search unnecessary—how it introduces entities so as to make anaphoric reference felicitous.

More generally, these results and those of Greene et al. speak to the kinds of research needed in discourse comprehension. It has recently been proposed that the representation of discourse constructed by comprehenders without specific goals or strategies is "minimal" (McKoon & Ratcliff, 1992). A minimal representation does not include all the inferences necessary to construct a full, real-life-like mental model of the situation described by a text. Instead, the only inferences constructed are those that are based on easily available knowledge or that are required to achieve coherence with information that is in the same local part of the text. For example, by this view, inferences about "what will happen next" in a story are inferred only if they can be based on well-known information. What will happen next to an actress who falls off a 14th-story roof is not well known and, data have suggested, not explicitly inferred (McKoon & Ratcliff, 1986, 1989a, 1989b, 1989c). The finding that pronoun resolution processes may fail to identify a unique referent for a pronoun pushes the minimalist approach much further. After all, inferring that someone dies after falling from a 14th-story roof might be viewed as quite a complicated inference, unlike a pronoun, which is often thought to be trivially understood by a reader. Clearly, from the pattern of results shown in this article and by Greene et al., pronoun resolution is not a trivial matter. The unanticipated nature of this pattern of results reinforces the minimalist emphasis on the importance of examining the local representation of discourse during comprehension. This pattern of results also underscores the minimalist claim that readers do not necessarily comprehend a discourse in some full, completely correct way; some sorts of "comprehension" may give only an incomplete representation of the meaning of a text.

Prior to this set of experiments, it would have been difficult to guess that stylistically appropriate pronouns were not always understood, that their comprehension depended on the verbs that preceded them in their discourse, and that their comprehension depended on the kind of clause in which they were placed. It would have seemed farfetched to claim that the lexical representation of a verb could determine whether or not a pronoun in a different clause was

understood. Here, we have expressed only the first preliminary ideas about how local representations of discourse might be constructed and what kinds of information they might depend on, and only the first preliminary data to address these problems. But these data should be sufficient to indicate how much we don't know about even the "smallest" parts of discourse comprehension.

References

Au, T. K. (1986). A verb is worth a thousand words: The causes and consequences of interpersonal events implicit in language. *Journal of Memory and Language, 25,* 104–122.

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

Brown, R., & Fish, D. (1983). The psychological causality implicit in language. *Cognition, 14,* 237–273.

Caramazza, A., Grober, E., Garvey, C., & Yates, J. (1977). Comprehension of anaphoric pronouns. *Journal of Verbal Learning and Verbal Behavior, 16,* 601–609.

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

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

Clark, H. H. (1977). Bridging. In P. N. Johnson-Laird & P. C. Wason (Eds.), *Thinking: Readings in cognitive science* (pp. 411–420). New York: Cambridge University Press.

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

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

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

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

Fillmore, C. J. (1971). Verbs of judging: An exercise in semantic description. In C. J. Fillmore & D. T. Langendoen (Eds.), *Studies in linguistic semantics* (pp. 273–296). New York: Holt, Rinehart & Winston.

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

Fodor, J. A., & Bever, T. G. (1965). The psychological reality of linguistic segments. *Journal of Verbal Learning and Verbal Behavior, 4,* 414–420.

Fodor, J. A., Garrett, M., & Bever, T. G. (1968). Some syntactic determinants of sentential complexity, II: Verb structure. *Perception and Psychophysics, 3,* 453–461.

Forster, K. I. (1981). Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing. *Quarterly Journal of Experimental Psychology, 33,* 465–495.

Garvey, C., & Caramazza, A. (1974). Implicit causality in verbs. *Linguistic Inquiry, 5,* 459–464.

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

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

Gernsbacher, M. A., & Shroyer, S. (1989). The cataphoric use of the indefinite this in spoken narratives. *Memory & Cognition, 17,* 536–540.

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

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

Greene, S. B., McKoon, G., & Ratcliff, R. (1992). Pronoun resolution and discourse models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 266–283.

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

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

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

Healy, A. F., & Miller, G. A. (1971). The relative contribution of nouns and verbs to sentence acceptability and comprehensibility. *Psychonomic Science, 21,* 94–96.

Hoffman, C., & Tchir, M. A. (1990). Interpersonal verbs and dispositional adjectives: The psychology of causality embodied in language. *Journal of Personality and Social Psychology, 58,* 765–778.

Hudson, S. B., Tanenhaus, M. K., & Dell, G. S. (1986). The effect of the discourse center on the local coherence of a discourse. In *Proceedings of the Eighth Annual Conference of the Cognitive Science Society* (pp. 96–101). Hillsdale, NJ: Erlbaum.

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

Levin, B. (in press). *English verb classes and alternations: A preliminary investigation.* Cambridge, MA: MIT Press.

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

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

McKoon, G., & Ratcliff, R. (1980). The comprehension processes and memory structures involved in anaphoric reference. *Journal of Verbal Learning and Verbal Behavior, 19,* 668–682.

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

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

McKoon, G., & Ratcliff, R. (1989a). Assessing the occurrence of elaborative inference with recognition: Compatibility checking vs. compound cue theory. *Journal of Memory and Language, 28,* 547–563.

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

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

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review, 99*, 440–466.

McKoon, G., Ratcliff, R., Ward, G., & Sproat, R. (in press). *Syntactic prominence effects on discourse processes*. *Journal of Memory and Language*.

McKoon, G., Ward, G., Ratcliff, R., & Sproat, R. (1993). Morphosyntactic and pragmatic factors affecting the accessibility of discourse entities. *Journal of Memory and Language, 32*, 56–75.

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

Osgood, C. E. (1970). Interpersonal verbs and interpersonal behavior. In J. L. Cowan (Ed.), *Studies in thought and language* (pp. 133–228). Tucson: University of Arizona Press.

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

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

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

Ward, G., Sproat, R., & McKoon, G. (1991). A pragmatic analysis of so-called anaphoric islands. *Language, 67*, 439–474.

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

Received July 16, 1992
Revision received December 4, 1992
Accepted December 8, 1992 ■

JOURNAL OF MEMORY AND LANGUAGE 32, 56–75 (1993)

Morphosyntactic and Pragmatic Factors Affecting the Accessibility of Discourse Entities

GAIL MCKOON, GREGORY WARD, AND ROGER RATCLIFF

Northwestern University

AND

RICHARD SPROAT

AT&T Bell Laboratories

Six experiments provide results showing that the accessibility of discourse entities is affected jointly by pragmatic and morphosyntactic factors. Accessibility was varied pragmatically by making an entity more or less closely related to the topic of its discourse, and it was varied syntactically by introducing an entity either in a verb phrase (*deer* in *hunting deer*) or in a compound (*deer hunting*); the latter should be less accessible according to linguistic data. The accessibility of an entity was examined by measuring the difficulty of understanding a pronoun intended to refer to the entity. Difficulty of understanding the pronoun was measured with reading time for a sentence mentioning the entity, with a test of short term memory, and with a test of long term memory. Results showed that both the pragmatic and syntactic variables affected reading time for the sentence with the pronoun, but that in all cases the relationships among the referent, the pronoun, and information given in the discourse about them appeared to be understood both in their representation in short term memory and in their representation in long term memory.

An important aspect of understanding language, whether listening to a speaker or reading a text, is relating each new piece of information to information that has already been conveyed. This context of prior information is assumed to be represented in "working memory" and used in determining the meanings of individual words, the relations among individual propositions, and the relevance of concepts and propositions to the overall message. The information in working memory is especially critical for the interpretation of pronouns and other anaphoric expressions. In this article, we investigate the structure of information in working memory as it relates to the comprehension of pronouns. We assume a complex structure that is determined by both morphosyntactic and pragmatic factors; following recent work in computational linguistics and discourse analysis, we label this structure a "discourse model." In six experiments, we investigate some of the referential properties of such a model. The experiments investigate the ease with which specific entities in the discourse model may be accessed by means of pronominal reference, and they show that successful reference is a function of both the pragmatic and syntactic context in which the referent was evoked in the prior discourse.

Within cognitive psychology, there have been two distinct traditions of text processing research that have investigated how online language comprehension in general, and anaphor interpretation in particular, re-

This research was supported by NSF Grant 85-16350, NIDCO Grant R01-DC01240, and AFOSR Grant 90-0246 (jointly funded by NSF) to Gail McKoon and by NIMH Grants HD MH44640 and MH00871 to Roger Ratcliff. We thank Steven Greene and Julia Hirschberg for discussions relevant to this work. Correspondence and reprint requests should be addressed to Gail McKoon, Psychology Department, Northwestern University, Evanston, IL 60208.

0749-596X/93 $5.00

late to the representation of information in a discourse model. One tradition has generally focused on syntactic determinants of linguistic structure, and, more narrowly, on structure within a single sentence. Under this view, the relationships among the elements of a sentence are organized according to the syntactic roles that they fill in that sentence. Reference to concepts or entities previously evoked by the text is accomplished by accessing syntactically defined elements; an anaphor accesses the syntactic part of the sentence in which its antecedent occurs. Ease of access is determined by the position of the antecedent in the syntactic structure. Mathews and Chodorow (1988), for example, provide data suggesting that antecedents more deeply embedded in a syntactic structure lead to more difficulty for the interpretation of an anaphor than antecedents not so deeply embedded. In a similar vein, data from experiments by Nicol and Swinney (1989) suggest that the availability of a potential referent is a function of its "syntactic appropriateness" as the antecedent of an anaphor. Syntactic approaches to the on-line representation of discourse information are reviewed by Mathews and Chodorow (1988) and by Fodor (1989).

The other traditional approach to the on-line processes and representations relevant to anaphora has focused on the structure of a discourse as a whole, rather than on single sentences (cf. Haviland & Clark, 1974; Malt, 1985). Kintsch (1974) proposed that a discourse was made up of semantic propositions ("individual idea units") and that these propositions were connected to each other through shared arguments. A connected set of propositions was assumed to consist of a "topic proposition," i.e., the most important proposition of the set, and the importance of all other propositions was defined relative to this proposition. Kintsch and van Dijk (1978) later incorporated this structural proposal into a model of on-line comprehension. In this model, each new set of propositions in a discourse is added to the already existing structure via connections among shared arguments, with preference given to more recently mentioned propositions and arguments. Entities of the discourse that are more topical are more likely to be kept active in short-term memory, and therefore they are more available as referents of anaphoric elements.

The "discourse model" approach that we assume as the background for our research combines elements from the two traditions in psycholinguistics and from computational linguistics, and also introduces several new elements. Following Sidner (1981), Webber (1979), and the propositional tradition (Haviland & Clark, 1974; Kintsch, 1974), we assume that discourse models contain the entities ("arguments," Kintsch, 1974, or "cognitive elements," Sidner, 1981) evoked in a discourse, and these entities are linked together by the relations in which they participate. The entities in question are assumed to be *conceptual* entities—not linguistic ones. As Morgan (1978), Webber (1979), Sidner (1981), and others have pointed out, language and, in particular, referring expressions, are used to refer to objects in the world (or model thereof), and not to other linguistic units.

We also assume that the entities represented in the discourse model are associated with varying degrees of accessibility. Not all noun phrases evoke discourse entities. For example, the anaphor *it* in the sentence *It's snowing outside* does not evoke a discourse entity (cf. Kamp, 1981; Heim, 1982; Webber, 1983), and so the notion of accessibility does not apply. Other anaphors, such as *do so*, have been argued to require explicit linguistic antecedents (McKoon et al., in preparation; Murphy, 1985; Tanenhaus & Carlson, 1990) and therefore may be more sensitive to surface form than to the discourse level of representation. In this article we exclude these

kinds of anaphors and restrict discussion to anaphors that are used to evoke discourse entities in a discourse model and consider their varying degrees of accessibility. We assume that the entire current discourse—and not just individual component sentences—is represented in the discourse model (cf. Kintsch, 1988), although at times, of course, portions of it will be relatively inaccessible and other portions will be particularly salient, or "in focus" (cf. Grosz, 1978; Grosz & Sidner, 1986). Which entities are highly accessible ("in focus") will change as the discourse progresses, partly as a function of recency, and partly as a function of shifts in topic (cf. Malt, 1985).

Our notion of a discourse model differs from previous psycholinguistic proposals in two key ways. First, we claim that the accessibility of discourse entities for subsequent anaphoric reference is determined not by syntax alone and not by topicality alone, but by a variety of syntactic, pragmatic, and semantic factors. The critical consequence of this claim is that there need be no single, most accessible entity (such as the topic) in the discourse, nor is there a single metric (such as syntactic depth of embedding) by which accessibility can be calibrated. Experiments 1 through 6 support this claim by showing that accessibility depends simultaneously on both syntactic and pragmatic factors.

Second, we maintain that the accessibility of an entity in a discourse model is determined not only by the context in which it is introduced but also by the cue with which that entity is later accessed by the comprehension system. Different cues may access the same entity with varying degrees of success; in some contexts, a definite description may work better than a pronoun, and in other contexts, the reverse might be true. Furthermore, the entities that are most accessible given one cue may be different from the entities that are most accessible given another cue. For example, a pronoun may serve to evoke more recent entities, whereas a definite description might serve to evoke more distant entities. Our notion is that reference processing is an interaction between an anaphoric cue and discourse entities in memory. Later in this article, we describe this notion through the metaphor of current global memory models and show how it guides the methodology used in the experiments.

It is important to note the limitations on the theoretical discourse model that we assume. The model is hypothesized to include entities that are explicitly mentioned in the discourse, the relations among those entities (cf. Kintsch, 1974), and their accessibilities relative to potential cues. Whether information of other kinds, such as inferences, "mental models," or causal structures, is also included in the working memory representation of text is an open question (McKoon & Ratcliff, 1992). Thus, for present purposes, our conception of a discourse model represents only the information necessary for processing the kinds of anaphora under investigation, and therefore it differs from the models that have been proposed by some other researchers (Bransford, Barclay, & Franks, 1972; Johnson-Laird, 1983; Morrow, Bower, & Greenspan, 1989; Oakhill, Garnham, & Vonk, 1989; Sanford & Garrod, 1981).

Because the discourse model theory assumed in our research contains elements of previous approaches, it is consistent with a number of previous empirical findings. In Kintsch's model for on-line text comprehension (Kintsch, 1988), the accessibility of an entity depends on the recency with which it was evoked and on how closely connected it is to the discourse topic. Empirically, both of these variables have been demonstrated to affect accessibility as hypothesized: it has been shown that more recently mentioned entities are more accessible (Jarvella, 1971; Caplan, 1972), and that entities more closely connected to the topic are better recalled (Kintsch &

Keenan, 1973) and better recognized (McKoon, 1977). Because the discourse model theory incorporates both recency and topicality as variables affecting accessibility, these findings are consistent with it.

The theory is also consistent with research motivated by more syntactic views of discourse representation. Under these views, the accessibility of an anaphor for an antecedent depends on the syntactic position of the antecedent. Mathews and Chodorow (1988), for example, tested comprehension of the pronoun in sentences like (1a) and (1b):

(1a). After the bartender served the patron, he got a big tip.
(1b). After the bartender served the patron, he left a big tip.

They found that reading time for the clause with the pronoun was faster when the antecedent of the pronoun occurred in subject position than when it occurred in object position. On a strictly syntactic account, this advantage would be due to a search process for the antecedent through the sentence's syntactic structure. An antecedent in subject position, as in (1a), would have an advantage in a left-to-right or top-down search. A discourse model approach would also predict an advantage when the antecedent is in subject position, but not because of a search through a syntactic structure. Instead, the advantage would be due to the greater accessibility in the discourse model of entities evoked in subject position relative to entities evoked in object position.

In our view of discourse models, syntax is assumed to be one of the factors that determines the relative accessibilities of the entities in the model. Several studies have investigated such effects. Rothkopf, Biesenbach, and Billington (1986) and Rothkopf, Koether, and Billington (1988) have shown that a modifier is better recalled when it is presented in predicate adjective position than when it is presented in prenominal position. In Rothkopf et al.'s experiments, texts contained sentences with phrases like *the yellow fruit* or *the fruit that was yellow*. Subjects were better able to answer a later question about the color of the fruit if they had read the second (predicate adjective) version. McKoon, Ward, Ratcliff, and Sproat (in preparation) demonstrated the same point with a different procedure; they showed that a predicate adjective is better recognized than a prenominal one. For example, the adjective *hostile* was presented in either prenominal or predicate position: *The hostile aunt was intolerant* or *The intolerant aunt was hostile*. Later recognition of the word *hostile* was faster and more accurate when it had been read in predicate adjective position. Similarly, concepts presented in direct object position are better recognized than concepts presented in an indirect object position, again demonstrating the effect of syntactic context on later accessibility (McKoon et al., in preparation).

Previous findings such as those just described show either pragmatic influences on accessibility (e.g. Kintsch & Keenan, 1973) or syntactic influences (e.g. Mathews & Chodorow, 1988). What they do not show is that these factors combine in a discourse to jointly affect accessibility for a single discourse entity. This was one of the goals of the experiments presented in this article. Accessibility was examined through its effects on the ease of comprehension of pronouns; the more accessible an entity, the more easily comprehended should be a pronoun being used to refer to that entity.

A second goal of the experiments was to investigate an interesting case of anaphora that has been the topic of much debate in the linguistics literature. This type of anaphora provided us with the means to manipulate accessibility via the syntactic structure by which an entity was introduced into a discourse.

In this type of anaphora, reference is

made to entities evoked by antecedents that appear within morphologically complex words. In the second sentence of (2) below, the pronoun *it* has as its antecedent *Kal Kan*. *Kal Kan* appears within the complex word *Kal Kan cat*, where we use the notion of *word* as defined in recent studies in morphology (cf., Matthews, 1974; Mohanan, 1986): a word may consist of a combination of a stem plus some affixes, normally written as a single orthographic word in English, or else may be a *compound* of several stems, often written as multiple orthographic words, as is the case with *Kal Kan cat*.

2. Patty is a definite Kal Kan cat. Every day she waits for it.

A number of linguistic studies have argued that examples like (3b), in which an antecedent occurs within a compound, are ungrammatical, and so have postulated a grammatical prohibition against complex words containing antecedents for anaphoric elements (e.g., Postal, 1969; Lakoff & Ross, 1972; Simpson, 1983; Mohanan, 1986). In particular, Postal (1969) proposed that no anaphor could have as its antecedent a word that was "part of the sense of" another word. Contrasts such as the one exhibited in (3) (Postal, 1969, p. 230) are claimed to be the result of such a grammatical prohibition:

3a. Hunters of animals tend to like them.
3b. Animal hunters tend to like them.

According to Postal, *them* can be interpreted as "referring to" animals in (3a), but not in (3b). In (3b), *animal* is morphologically contained within the compound *animal hunters*, which by Postal's constraint constitutes what is called an "anaphoric island," and cannot by grammatical rule provide the antecedent for *them*.

However, Ward, Sproat, and McKoon (1991) have argued against this position, presenting dozens of examples of felicitous naturally occurring tokens from a variety of oral and written sources. The example in (2) is one of these tokens; others are given in (4) (the specific sources for the examples are given in Ward et al., 1991):

4a. *Bush* supporters would stay home, figuring *he*'d already won. (*he* = Bush)
4b. Call if you're a *small business* owner, or interested in starting *one*. (*one* = a small business)
4c. For a *syntax* slot, I'd rather see someone with more extensive coursework in *it*. (*it* = syntax)
4d. We went up to *Constable* country; we stayed in the village *he* was born in. (*he* = Constable)
4e. Millions of *Oprah Winfrey* fans were thoroughly confused last week when, during *her* show, she emotionally denied and denounced a vile rumor about herself. (*her* = Ophrah Winfrey)
4f. Our neighbors, who are sort of *New York City*-ites, they have jobs *there* . . . (*there* = New York City)
4g. Do *parental* reactions affect *their* children? (*their* = parents)

Given that examples such as these occur naturally in spoken and written language, it would appear that word-internal elements *can* serve as antecedents for anaphors, contrary to the claims of Postal and others.

In fact, Ward et al. (1991) argue that there is no *grammatical* constraint preventing word-internal elements from serving as antecedents for anaphors. Rather the felicity of such anaphora is a function of the accessibility of the discourse entity evoked by the word internal element to which the anaphor is intended to refer. Consistent with our assumptions about the representation of entities in a discourse model, we claim that both pragmatic and syntactic factors are relevant for the accessibility of the entity. In other words, the factors involved in determining the felicity of anaphora for anaphoric islands are exactly the same as the factors involved in determining the accessibility of discourse entities in general.

According to Ward et al. (1991), the un-

acceptability of anaphora like that in (3b) is due to the inaccessibility of the relevant discourse entity. As mentioned above, modifiers have been shown to be relatively inaccessible (McKoon et al., in preparation; Rothkopf et al., 1986; Rothkopf et al., 1988) and so, assuming that the word-internal element is functioning as a modifier, word-internal elements should not generally be sufficiently accessible to reference by anaphora.

On the other hand, all of the pragmatic, syntactic, and semantic factors that determine accessibility in a discourse model can conspire, singly or jointly, to make word-internal elements sufficiently accessible to permit subsequent anaphora. For example, discourse entities can increase in accessibility through relevance to the listener or reader; *Sheep farmers tend to like them* was judged acceptable by some members of a New Zealand audience. Ward et al. (1991) point out two further ways in which a discourse entity can become more accessible. One way is through contrast with another discourse entity, as in (5), a quote from President Reagan's 1990 farewell speech:

5. Well, action is still needed. If we're to finish the job, Reagan's Regiments will have to become the BUSH Brigades. Soon *he'*ll be the chief, and he'll need you every bit as much as I did.

The other way is through topicality. In a television commercial for Saab, the pronoun *it* in sentence (6) can felicitously refer to the Saab model 9000-CD which was evoked by a word internal to the compound *Saab 9000-CD owners*. Similarly, in the first text in Table 1, the topic of the discourse segment is hunting and the discourse entity corresponding to the referent of the pronoun in the last sentence (i.e., *they/deer*) is closely related to the topic; therefore we would hypothesize that it is relatively accessible.

6. We asked *Saab 9000-CD* owners about *its* road-handling . . .

In sum, we have reason to believe not only that the compound construction illustrated in (3b) serves to render an entity rel-

TABLE 1
EXAMPLES OF TEXTS USED IN EXPERIMENT 1

High topicality, compound

Sam likes the outdoor life. Having grown up in rural Kentucky, he knows a lot about nature and is an expert at fishing and shooting. He goes on hunting trips as often as he can. He used to hunt just small game, like rabbit and quail. However, lately he's taken up deer hunting. He thinks that they are really exciting to track.

Low topicality, compound

Sam has many interests in the outdoors. He's an avid skier, and each winter he takes about a month off from work to ski in Colorado. In the summertime, he visits his parents in Montana where he has a chance to do some mountain climbing. Lately, he's taken up deer hunting. He thinks that they are really exciting to track.

High topicality, verbal complement

Sam likes the outdoor life. Having grown up in rural Kentucky, he knows a lot about nature and is an expert at fishing and shooting. He goes on hunting trips as often as he can. He used to hunt just small game, like rabbit and quail. However, lately he's taken up hunting deer. He thinks that they are really exciting to track.

Low topicality, verbal complement

Sam has many interests in the outdoors. He's an avid skier, and each winter he takes about a month off from work to ski in Colorado. In the summertime, he visits his parents in Montana where he has a chance to do some mountain climbing. Lately, he's taken up hunting deer. He thinks that they are really exciting to track.

Note: Referent noun: deer.

atively inaccessible in some discourse contexts but also that an entity evoked in this construction can be made quite accessible in other discourse contexts. The hypothesis of a joint contribution to accessibility of morphosyntactic and pragmatic factors makes a number of predictions amenable to empirical investigation, which we report on below. To anticipate, in Experiment 1, we varied topicality for entities evoked by antecedents contained in the compound and the corresponding verb phrase constructions, as shown in Table 1. Our prediction was that accessibility for the "referent entity" (*deer* in Table 1) would be increased both by the pragmatic and the syntactic variables; the entity would be more accessible when it was more closely related to the topic and when it was introduced in a verb phrase rather than a compound.

How to Measure Accessibility

Given our notion of a discourse model, accessibility is defined as the ease with which a discourse entity, introduced at one point in a discourse, can be referenced at a later point in the discourse by some cue, such as a pronoun. The empirical goal is to measure accessibility by measuring ease of reference, that is, to measure the ease with which pronouns are understood. This requires at least a minimal model of comprehension processes for pronouns.

In Greene, McKoon, and Ratcliff (1992) and Ward, Sproat, and McKoon (1991), we proposed that a pronoun is completely and correctly understood if its intended referent is sufficiently more highly accessible in the discourse model, relative to the pronoun as a cue, than all other discourse entities. Following current global memory models (Gillund & Shiffrin, 1984; Hintzman, 1988; Murdock, 1982; Ratcliff, 1978; see also Gernsbacher, 1989), a pronoun is assumed to be matched against all entities in the discourse model in parallel. The semantic and grammatical features of the pronoun are matched against the features of the discourse entities. Every entity in the discourse model will match the pronoun to some degree, with the degree of match depending on both the entity's semantic and grammatical features and its accessibility. If the degree of match for some single entity is sufficiently high, and sufficiently higher than the match for all other entities, then (without further processing) that entity is identified as the pronoun's referent; in essence, a sufficiently high degree of match constitutes a decision about the pronoun's referent. If there is no entity that matches sufficiently well, then a referent is not identified. If more than one entity matches sufficiently (but none sufficiently better than the others), then again no single referent is identified. In the cases where a referent is not identified, comprehension may fail in the sense that the pronoun is left without a referent. Alternatively, selection of a referent might be postponed, waiting for more information from the discourse, or for strategic problem solving processes that might be able to identify a referent. In the usual case, where a single entity matches the pronoun sufficiently better than all other entities, the identification of the pronoun with the referent leads to the attachment in the discourse model of information associated with the pronoun to information associated with the referent.

This model for comprehension of pronouns makes the explicit claim that pronouns vary in the ease with which their referents can be identified such that, in some cases, no referent at all is automatically and uniquely identified. Failure to identify a unique referent might occur as the result of a number of factors, including the semantic and pragmatic content of the discourse and the speed required of comprehension processes by the speaker or reader. The possibility that pronouns sometimes fail to evoke unique referents has been discussed previously by Yule (1982), who points out that, in some discourse contexts, the identity of the entity referenced by an anaphor may be irrelevant to the reader or listener. Webber (1983) also suggests that, if there is no im-

mediate need to determine a unique referent, an anaphor may be left unresolved. Empirically, failure to resolve pronouns has been demonstrated by Greene et al. (1992). Their experiments investigated the difficulty of identifying a unique referent for a third person singular pronoun when two possible referents had been evoked in the discourse. Evidence for unique resolution was obtained only when reading rate was slow or readers could anticipate at exactly what point in the discourse the pronoun would occur. When reading rate was more normal (250 ms per word) or readers could not exactly anticipate the pronoun, the data suggested that no unique referent was identified.

The possibility that pronouns may sometimes be left unresolved complicates efforts to measure how difficult they are to comprehend. In particular, the time taken to read a pronoun (or the time to read a sentence containing a pronoun) is not an adequate measure. This is because reading times can reflect either time to successfully resolve a pronoun or time to process the pronoun but fail to resolve. One pronoun read in a given amount of time might be relatively easy to comprehend and so be identified with a unique referent, while another pronoun read in the same amount of time might be relatively difficult and left without a referent. In other words, reading time cannot be interpreted as a measure of comprehension difficulty unless it is combined with some measure of whether the pronoun was successfully resolved. Two methods have been typically adopted in previous research (cf. Chang, 1980; Corbett & Chang, 1983; Gernsbacher, 1989; McKoon & Ratcliff, 1980b). One is to present the intended antecedent of the pronoun as a recognition test word at some point in the discourse after the pronoun. The reasoning that underlies this method is that successful resolution of the pronoun will increase the accessibility of its referent. This increase in accessibility will, in turn, facilitate the recognition decision about the referent when it is presented as a test word. This method was used in Experiments 1 through 3. The second method, used in Experiments 4 through 6, is to use priming in word recognition to show that information given in the discourse with the pronoun is connected in memory to the referent, as it should be if the referent is correctly and completely understood (McKoon & Ratcliff, 1980b).

Experiment 1

Table 1 shows examples of the texts that were used in the experiment. Subjects read texts one line at a time, in a self-paced procedure. After the final line of a text, a single test word was presented for recognition (a decision as to whether or not the word had appeared in the text).

Table 1 also illustrates the design of the experiment: the accessibility of a discourse entity was manipulated pragmatically, by how closely it was related to the topic of its text, and syntactically, by using either the verb phrase or the compound construction. The referent entity (*deer* in Table 1) was introduced in the next to last sentence of its text, and it was the intended referent of the pronoun mentioned in the last sentence. It was also used as the test word that appeared after the final line of the text. The hypothesis was that the accessibility of the referent entity would be increased when it was more closely related to the topic and when it was introduced in a verb phrase. Increased accessibility was expected to result in faster reading time for the final sentence containing the pronoun, faster response time for the test word, or both.

Method

Subjects. Forty subjects participated in the experiment for credit in an introductory psychology class. Each subject participated in one 50-min session.

Materials. Twenty-four sets of four texts were written, each set with one critical referent noun. The four texts of a set implemented the variables of the experiment: the referent noun was used either in a com-

pound or in a verb phrase, and it was either more or less closely related to the topic of its text. The four texts of one set are shown in Table 1. For each of the four texts in a set, the next to last sentence stated the same information about the referent noun and a verb (e.g., *deer hunting* or *hunting deer*). The final sentences of the texts were the same in all four versions and referred to the referent noun with a pronoun (*He thinks they are really exciting to track*). The referent noun was stated only in the next to last sentence. The referent noun was also the test word for the experimental texts.

The mean lengths of both versions of the texts were 58 words, 5 sentences, and 7 lines as they appeared on a CRT screen. The last line of each text was always the entire final sentence of the text with no words from the preceding sentence.

There were 30 additional texts used as fillers in the experiment. These varied from 5 to 7 CRT lines in length, and averaged 50 words. Twenty of these had associated with them a single test word that did not appear in any of the filler or experimental texts. The test word for the other 10 was a word from the text. For each of these 30 texts, there was a true/false test sentence. Half of the test sentences were true and half false.

Procedure. All materials were presented to subjects on a CRT screen, and responses were made on the CRT's keyboard. Presentation and data collection were controlled by a real-time computer system.

The experimental session began with practice on 10 items presented one at a time for lexical decision. Subjects were instructed to respond to these items as quickly and accurately as possible, pressing the ?/ key on the keyboard if the test item was a word and the z key if it was not a word. These items were used to familiarize the subjects with the response keys.

After this practice, the experiment proper began. The texts were presented one at a time, with six of the fillers first, and then the remaining 24 fillers and the 24 experimental texts in random order. For each text, first the instruction *Press space bar for next paragraph* appeared on the screen. When the subject pressed the space bar, there was a pause of 1000 ms, and then the first line of the text appeared. The line remained on the screen until the subject pressed the space bar again, and then the next line of the text appeared just below the first line. The subjects were instructed to press the space bar for the next line when they had read and understood the current line. The text continued in this way, with one additional line every time the space bar was pressed, until the last line of the text. When the space bar was pressed after reading of the last line, the screen was cleared and a test word appeared below where the last line had been. The test word was underlined by a row of asterisks. Subjects were instructed to respond *yes* (with the ?/ key) or *no* (with the z key) according to whether the test word had appeared in the preceding text. The test word remained on the screen until the subject pressed a response key, and then the screen was cleared. For the filler texts, the message *True-False Question* was then displayed, followed by the true/false question for the preceding text. Subjects answered the question by pressing the ?/ key for *true* and the z key for *false*. If the response was incorrect, the message *ERROR* was displayed for 2000 ms. After the true/false question, the next text began with the instruction to press the space bar.

Design. The two variables in the experiment were the topicality of the referent noun, and whether the noun was mentioned in a compound or a verb phrase. These two variables were crossed in a Latin square design, with four sets of materials (six per set) and four groups of subjects. Order of presentation of the texts was random, different for every second subject.

Results

For each text and each subject, means for the reading times of the texts' final sentences and means for response times to the

test words were calculated. Means of these means are presented in Table 2. Analyses of variance were performed on the means from the experimental design with both subjects and items as random variables; $p < 0.05$ was used unless otherwise noted.

First, the data for the test words are considered. For each text, the test word was the referent noun, the antecedent of the pronoun in the final sentence. If, for all four conditions, subjects interpreted the pronoun correctly during the time they were reading the final sentence, then response times to the test word should be equal across the conditions. The processes of interpreting the pronoun might be more or less difficult across conditions, but if the correct referent was always evoked by the pronoun then it should be equally accessible across conditions at the time the test word was presented. This is what the data show: there are no significant differences in response times to the test words (analyses of variance showed F's < 1.2). The standard error of the response times was 23.8 ms. Differences in error rates were also not significant, F's < 1.9.

Reading times show that there were differences in comprehension difficulty for the final sentences. It was hypothesized that interpretation of the pronoun would be difficult when the antecedent of the pronoun was in the modifier position in the compound. The data show this difficulty when the referent noun was low in topicality: reading times were longer when the noun was in a compound compared to when it was not. However, according to the discourse model theory, the difficulty should be reduced when the referent noun is more topical. This hypothesis was confirmed; increased topicality reduced reading times in the compound condition so that they were only slightly longer than in the verb phrase condition.

These effects were supported by analyses of variance. The main effect of compound versus verb phrase was significant, $F_1(1,39) = 10.2$ and $F_2(1,20) = 7.4$, as was the main effect of topicality, $F_1(1,39) = 21.8$ and $F_2(1,20) = 13.3$. The interaction of the two variables was marginally significant, $F_1(1,39) = 3.7$ and $F_2(1,20) = 4.3$. Planned tests showed that the difference between the compound and verb phrase conditions was significant when the referent noun was low in topicality, $F_1(1,39) = 11.2$ and $F_2(1,20) = 14.1$, but not when it was high in topicality, F's < 1.0. The standard error of the reading times was 52.5 ms.

For the true test questions, the mean response time was 2110 ms with 9% errors. For the false questions, the means were 2031 ms and 9% errors.

EXPERIMENTS 2 AND 3

Our interpretation of the results of Experiment 1 depends on the assumption that

TABLE 2
DATA FROM EXPERIMENT 1

Syntactic structure	Response times and error rates for test words			
	Low topicality text version		High topicality text version	
Compound	907 ms	5%	870 ms	2%
Verbal complement	893 ms	4%	886 ms	4%
Filler positive test words	1242 ms	21%		
Filler negative test words	1181 ms	15%		

Syntactic structure	Reading times for final sentences	
	Low topicality text version	High topicality text version
Compound	2117 ms	1785 ms
Verbal complement	1868 ms	1738 ms

subjects understood the correct referents of the pronouns in the final sentences of the texts in all of the experimental conditions. This assumption is consistent with the finding that response times for the test words were equal across experimental conditions. However, the assumption might be wrong. An alternative possibility is that the pronouns were not understood at all, and that this is the reason that response times to the test words did not differ across the experimental conditions. By this alternative, the differences in reading times would represent differing degrees of *unsuccessful* efforts at understanding the final sentences, and there would be no way to determine whether the same pattern of reading times would hold for successful efforts. Experiments 2 and 3 were designed to rule out this alternative.

In both of these experiments, the same basic texts were used as in Experiment 1. However, there were two different possible final sentences. In one final sentence, the same pronoun referring to the critical referent noun was used as in Experiment 1 (*And he says they are really exciting to track* for the text in Table 1). In the second final sentence, a new noun was substituted for the pronoun (*And he says bears are really exciting to track*). This new noun had not been mentioned previously in the text.

In Experiment 2, the final sentence mentioned either the pronoun or the new noun, and following the final sentence, the referent noun was presented as a test word. If the pronoun in the pronoun version of the final sentence is understood as referring to the referent noun, and it is this processing that leads to the facilitation of response times when the referent noun appears as a test word, then response times should be facilitated only when the final sentence contains the pronoun, and not when it mentions the new noun. This was the prediction for the results of Experiment 2.

In Experiment 3, the two final sentences from Experiment 2, one with the pronoun and the other with the new noun, were used and a new test word was introduced. The new test word was a "control" word picked from one of the earlier sentences of the text (e.g., *trips* for the texts in Table 1). There was also a second test word, the same referent noun test word as was used in the previous experiments. Again, we predicted response times to the test words from our assumption that the pronoun in the pronoun version of a final sentence is understood to refer to the referent noun. The pronoun version of the final sentence should facilitate response times for the referent noun test word relative to the new noun version, but response times for the control word should not be affected by which version of the final sentence is read.

Method

Subjects. For Experiment 2, there were 40 subjects and for Experiment 3, 24 subjects, all from the same population as in Experiment 1.

Materials. The basic texts from Experiment 1 were used in Experiments 2 and 3. For each text, a new final sentence was written. This sentence was almost the same as the old final sentence except that the pronoun was replaced by a noun. The new noun had not been mentioned previously in the text, but it plausibly fit the context of the text. There were slight changes in wording from the final sentences used in Experiment 1 to the sentences for Experiments 2 and 3, in order to keep both the pronoun and the new noun versions of the sentences about equally plausible. The mean length of the final sentences with pronouns was 8.4 words, and the mean length of the final sentences with new nouns was 8.9 words. For Experiment 2, the test word for each text was the critical referent noun (e.g., *deer*), the same as was used in Experiment 1. For Experiment 3, there were two possible test words, the referent noun and another control word that had appeared earlier in the text. For both experiments, the same filler paragraphs were used as in Experiment 1.

In these experiments, including all four

versions of the basic texts would have reduced power beyond acceptable limits. We compromised considerations of power with considerations of generality across versions by using two versions in Experiment 2, the high topicality compound version and the high topicality verb phrase version. In Experiment 3, only one version of the basic texts was used, the high topicality, compound version.

Method and design. The procedure was the same as that used in Experiment 1. For Experiment 2, there were two variables: whether the referent noun was stated in a compound or a verb phrase, and whether the final sentence contained the pronoun or the new noun. For Experiment 3, there were also two variables: the final sentence mentioned either the pronoun or the new noun, and the test word was either the referent noun or the control word. For both experiments, the two variables were combined in a Latin square design with four sets of materials and four groups of subjects. The order of presentation of the texts was random, different for every second subject.

Results

The data were analyzed as in Experiment 1, and are presented in Tables 3 and 4.

Experiment 2. When the final sentence contained the pronoun referring to the critical noun, the results of Experiment 2 replicated those of Experiment 1. Whether the critical noun was introduced in a verb phrase or a compound, high topicality should have made it easily accessible, and so, as is shown in Table 3, there should be little effect of syntactic structure on either response times for the referent nouns or reading times for the final sentences.

If processing of the pronoun in the final sentence facilitated responses to the critical noun test word, then replacing the pronoun in the final sentence with a new noun should slow responses to the test word. The data clearly show this effect.

Analyses of variance showed only one significant effect for response times for the referent nouns; when the final sentences contained the new nouns, response times were longer than when the sentences contained the pronouns, $F_1(1,39) = 18.1$ and $F_2(1,20) = 225.8$. The standard error was 27.1 ms. There were more errors on the test words when the final sentences contained the new nouns; these results were marginally significant with $F_1(1,39) = 3.7$ and $F_2(1,20) = 3.5$. There was also only one significant effect for reading times; reading times for the sentences with the new nouns were longer than reading times for the sentences with pronouns, $F_1(1,39) = 8.0$ and $F_2(1,20) = 5.2$. The standard error of the reading times was 102.2 ms. All other F's were less than 2.6.

For the true test questions, the mean response time was 1985 ms with 10% errors

TABLE 3
DATA FROM EXPERIMENT 2

	Response times and error rates for test words				
Syntactic structure	Pronoun final sentence			New noun final sentence	
Compound	948 ms	7%		1070 ms	8%
Verbal complement	926 ms	5%		1045 ms	10%
Filler positive test words		1263 ms	23%		
Filler negative test words		1150 ms	14%		

	Reading times for final sentences	
Syntactic structure	Pronoun final sentence	New noun final sentence
Compound	1961 ms	2199 ms
Verbal complement	2012 ms	2254 ms

TABLE 4
DATA FROM EXPERIMENT 3

	Response times and error rates for test words			
Test word	Pronoun final sentence		New noun final sentence	
Critical noun	884 ms	2%	1028 ms	10%
Control noun	1216 ms	17%	1219 ms	14%
Filler positive test words	1157 ms	23%		
Filler negative test words	1106 ms	8%		

	Reading times for final sentences	
	Pronoun final sentence	New noun final sentence
Compound, high topicality	1884 ms	1951 ms

and for the false questions, the means were 1941 ms and 14% errors.

Experiment 3. In Experiment 3, the final sentence contained either the new noun or the pronoun that was intended to refer to the referent noun. For the referent noun test word, responses should be facilitated only with the pronoun and not the new noun, as in Experiment 2, and the means in Table 4 show this facilitation. For the control test word, there should be no effect of whether the final sentence contained the pronoun or the new noun, and the data showed no effect.

Analyses of variance for response times to the test words showed a main effect for test word (referent noun or control word), $F_1(1,31) = 36.6$ and $F_2(1,23) = 147.8$, and a main effect of final sentence (pronoun or new noun), $F_1(1,31) = 4.8$ and $F_2(1,23) = 11.6$. The interaction of the two variables was significant, $F_1(1,31) = 4.2$ and $F_2(1,23) = 7.2$. Standard error for the response times was 26 ms. For error rates, the main effect of test word was significant, $F_1(1,31) = 17.9$ and $F_2(1,23) = 9.7$, as was the interaction of test word and final sentence, $F_1(1,31) = 6.7$ and $F_2(1,23) = 5.1$. The difference in reading times for the two versions of the final sentences was marginally significant, $F_1(1,32) = 15.2$ and $F_2(1,24) = 3.7$.

For true test statements, the mean response time was 1936 ms (8% errors), and for false test statements, 1941 ms (13% errors).

EXPERIMENTS 4, 5, AND 6

Experiments 1, 2, and 3 appear to show that the time required to comprehend a pronoun is a function of the accessibility of the pronoun's referent in the discourse structure. When accessibility is reduced, either via syntax, by introducing the referent with the compound rather than the verb phrase syntax, or via pragmatics, by making the referent less relevant to the discourse topic, then comprehension takes longer. This was shown in the reading times of the sentences containing the pronouns.

We pointed out that increased reading time does not by itself conclusively show that the pronouns were understood. In addition, some measure of the extent to which the pronouns were actually understood must be provided. Experiments 1, 2, and 3 used an immediate test of the antecedent of the pronoun (the referent noun) to provide evidence of comprehension. Immediate testing provides evidence about the relationships among discourse concepts that are available when both the discourse and the test item are in working memory at the same time (Corbett & Dosher, 1988; van Dijk & Kintsch, 1983; McKoon & Ratcliff, 1980b; 1986; 1989); in the present case, the relevant relationships are those among the pronoun, its intended referent in the discourse model, and the test word. From the results of Experiments 1, 2, and 3, we can conclude that those relationships were available to subjects at the time the test

word was presented. Whether understanding was complete, to the extent that the relationships among the pronoun, its intended referent, and information given in the discourse about the referent were all encoded into long term memory is still an open question (see McKoon & Ratcliff, 1989, for a case in which relationships available at immediate testing were not available at later testing). In Experiments 4, 5, and 6, we used a priming procedure to examine these relationships in long term memory.

The experiments involved a series of study test lists. For each list, subjects read four texts, and then they were given a list of test words for recognition (responding positively if a test word had appeared in one of the studied texts, and negatively if it had not). For the experimental texts, the test words of interest were the referent noun (e.g., *deer*) and a modifier from the final sentence (e.g., *exciting*). These two words were presented in immediately adjacent positions in the test list, with *exciting* following *deer*, and so they formed a "priming" pair. From previous research (McKoon & Ratcliff, 1980a; 1980b; Ratcliff & McKoon, 1978; Ratcliff & McKoon, 1988), it can be predicted that responses for the second word of the pair will be facilitated when they are closely related in memory by virtue of being from the same text (relative to being from different texts). The question is whether facilitation will be even further increased when the modifier *exciting* should be understood (by virtue of processing the pronoun) to describe the referent noun *deer*. Such further facilitation would be evidence that comprehension of the pronoun resulted in long-term memory encoding of the appropriate relationships between the referent and information given in the discourse about the pronoun.

In Experiment 4, the final sentence of a text contained either the pronoun for which the referent noun was the intended antecedent (. . . *they were exciting to track*), or the "new noun" of Experiments 2 and 3 (. . . *bears were exciting to track*). If subjects understand the final sentences completely, then *deer* should be more closely related in memory to *exciting* for the pronoun version of the final sentence than the new noun version, and this increased relatedness should lead to greater facilitation of responses to *exciting* by *deer* for the pronoun final sentence than the new noun final sentence. The results of the experiment followed this prediction.

In Experiment 4, only one version of each text was used, the high topicality, compound version. Experiments 5 and 6 were designed to check that the referent noun and the modifier were closely related in memory for both the high and low topicality versions of the text (Experiment 5) and for both the compound and verb phrase versions (Experiment 6).

Method

Materials. The same basic 24 texts were used as in Experiments 1, 2, and 3. The test words for these texts were the referent noun, the modifier from the final sentence, and two other words from the text. Thirty-two filler texts (30 of them the same as in Experiments 1, 2, and 3) each had four positive test words. Negative test words were chosen from a pool of 142 words that did not appear in any text.

Procedure. The experiments began with ten lexical decision test items, presented for practice on the response keys. This practice was followed by 14 study test lists. The first two study lists each contained four filler texts, and the remaining 12 each contained two experimental texts and two filler texts. The four study texts were presented in random order, one at a time, for 10 s for the filler texts and 11.5 s for the experimental texts. There was a 1.5-s blank interval between each text. After the four texts, a row of asterisks was presented for 1 s to signal that the test list was about to begin. The words in the test list were presented one a time. A word remained on the CRT screen until a response key was pressed (? for positive responses, z for negative re-

sponses). If the response was correct, then there was a blank screen for 200 ms, and then the next test word. If the response was incorrect, the word *ERROR* was displayed for 2 s. There was a total of 26 test words, 16 positive and 10 negative. After the 26th test word, two true/false test statements were presented, one at a time, with the ERROR message displayed for 2 s after incorrect responses. Then the next study test list began.

For study test lists containing experimental texts, the 16 positive test words were: the referent noun and the modifier from each experimental text, two other words from each experimental text, and four words from each filler text. A modifier was always tested later than the third position in the test list, and it was immediately preceded by the referent noun either from its own text or the other experimental text, depending on the experimental condition. The other words from an experimental text were tested later in the test list than the modifier. Otherwise, the order of the test words was random. No word appeared more than once in a test list.

Design. In all three experiments, the first variable was whether the modifier was preceded in the test list by the referent noun from its own text or from the other experimental text. In Experiment 4, the second variable was whether the final sentence of a text was studied in the pronoun version or the new noun version, and only the high topicality, compound versions of the texts were used. In Experiment 5, the second variable was whether the context was high topicality or low topicality. The final sentence was always the pronoun version, and the referent noun always appeared in a compound. In Experiment 6, the second variable was whether the referent noun was presented in a compound or a verb phrase. The final sentence was the pronoun version, and the high topicality texts were used. In each experiment, the two variables were crossed in a Latin square design, with four groups of subjects and four sets of texts. There were 52 subjects in Experiment 4, 32 in Experiment 5, and 24 in Experiment 6.

Results

In Experiment 4, the referent noun should be more closely related in memory to the modifier when the final sentence referred to the referent noun with a pronoun than when it did not. Thus, in the test list, the referent noun should facilitate responses for the modifier more when the final sentence referred to the referent noun. This is the pattern shown in Table 5. With the pronoun in the final sentence, response times to the modifier are facilitated 160 ms when the referent noun comes from the same text relative to a different text. With the new noun in the final sentence, the facilitation is only 53 ms. This interaction was significant, $F_1(1,51) = 9.5$ and $F_2(1,23) = 6.0$. The main effect of same versus different text prime was also significant, $F_1(1,51) = 27.7$ and $F_2(1,23) = 33.9$. Which version of the final sentence was used had no significant effect, F's < 1.0. The standard error of the response time means was 19 ms.

TABLE 5
DATA FROM EXPERIMENT 4

	Response time and error rates for test words				
Prime	Pronoun final sentence			New noun final sentence	
Critical noun from same text	714 ms	9%		764 ms	6%
Critical noun from different text	874 ms	15%		817 ms	19%
Filler positive test words		784 ms	12%		
Filler negative test words		944 ms	25%		

Part of the significant interaction effect (but only part) comes from the pattern of response times in the conditions for which the prime comes from a different text than the modifier; responses are slower when the final sentence contains a pronoun than when it contains the new noun. This difference has no obvious explanation. For errors, the main effect of same versus different text for the prime was significant, $F_1(1,51) = 30.0$ and $F_2(1,23) = 14.9$. Other F's for error rates were less than 1.8.

True test statements averaged 2079 ms in response time, and 17% errors, and false statements, 2077 ms and 12% errors.

In Experiments 5 and 6, the hypothesis was that the relation between the pronoun in the final sentence and the referent noun is encoded in memory equally well, whether the text is presented in the high topicality or low topicality versions, or whether the referent noun is presented in a compound or a verb phrase. As a result, there should be equal amounts of facilitation from the referent noun to the modifier in all cases. The results in Tables 6 and 7 confirm this prediction.

In Experiment 5, there is about the same amount of facilitation with the high topicality texts (55 ms) as with the low topicality texts (48 ms). Overall, the subjects in Experiment 5 were faster than those in Experiment 4 (see response times for filler test items), so the facilitation is somewhat reduced in size. The main effect of whether the prime comes from the same or a different text than the modifier is significant, $F_1(1,31) = 9.9$ and $F_2(1,23) = 7.1$. The interaction with text version did not approach significance, F's < 1.0. The main effect of text version approached significance in the subjects analysis, $F_1(1,31) = 3.5$, but was less than one in the items analysis. The standard error of the response time means was 19 ms. For errors, the main effect of same versus different text for the referent noun was significant with the subjects analysis, $F_1(1,31) = 4.1$, but not with the items analysis, $F_2(1,23) = 2.1$.

True test statements averaged 2071 ms in response time, and 15% errors, and false statements, 1990 ms and 13% errors.

The results of Experiment 6 (Table 7) show that the amount of facilitation is not significantly affected by whether the referent noun appeared in a compound (49 ms of facilitation) or a verb phrase (64 ms). The main effect of same versus different text for the prime was significant, $F_1(1,23) = 10.0$ and $F_2(1,23) = 9.8$. All other F's were less than 1.0. The standard error of the response time means was 20 ms. Same versus different text for the prime also significantly affected error rates, $F_1(1,23) = 31.6$ and $F_2(1,23) = 11.8$. Again, no other F's were greater than one.

True test statements averaged 2126 ms in response time, and 15% errors, and false statements, 2007 ms and 11% errors.

Summary. Experiments 4, 5, and 6 used a priming procedure to examine the long term memory representation of the relations between the referent entity (e.g., *deer*) and information given in the text about that entity. In the final sentence, the information that *they are exciting to track* should be understood such that *exciting* is encoded into long term memory as describing *deer*.

TABLE 6
DATA FROM EXPERIMENT 5

Prime	Response time and error rates for test words				
	High topicality text version			Low topicality text version	
Critical noun from same text	665 ms	8%		691 ms	5%
Critical noun from different text	720 ms	10%		739 ms	10%
Filler positive test words		714 ms	11%		
Filler negative test words		855 ms	26%		

TABLE 7
DATA FROM EXPERIMENT 6

Prime	Response time and error rates for test words, syntactic structure				
	Compound			Verbal complement	
Critical noun from same text	678 ms	1%		671 ms	4%
Critical noun from different text	727 ms	12%		735 ms	13%
Filler positive test words		705 ms	10%		
Filler negative test words		879 ms	31%		

If so, then a response to *deer* in the test list should facilitate a response to *exciting*, more so than if the sentence had said that *bears are exciting to track*. Experiment 4 demonstrated this result, and Experiments 5 and 6 showed that the same result obtained whether *deer* was more or less topical and whether it was introduced in a verb phrase or a compound.

GENERAL DISCUSSION

A discourse model is the representation of information that is built during comprehension of a text or discourse. As comprehension proceeds through a text, the discourse model is continually updated and revised to include new input and to reflect the impact of new input on earlier information. In the discourse model theory assumed as the background for the experiments in this article, the model is made up of the entities evoked by linguistic and contextual information, the relations among the entities, and their accessibilities relative to potential referential cues.

The discourse model that we assume differs from previous psycholinguistic approaches in two ways. First, we propose that the accessibility of a discourse entity is a function of a number of factors, both linguistic and nonlinguistic, arising from explicit information in the text as well as from contextual information, pragmatic knowledge, and speaker/writer and listener/reader goals. In addition, the accessibility of an entity for later reference is determined by the cue with which it is referenced. A given entity may be quite accessible from one cue, but relatively inaccessible from another. Thus, accessibility is an interaction between entities in the discourse model and the cues used by the speaker/writer to evoke those entities.

The experiments presented in this article support the discourse model view by showing that both the morphosyntactic and the pragmatic context in which an entity is introduced into a discourse determine its accessibility for later reference. In Experiment 1, a referent entity (*deer*) was introduced in a morphosyntactic context that made it either more accessible (a verb phrase, *hunting deer*) or less accessible (a compound, *deer hunting*). Reading times for a sentence containing a pronominal anaphor for the referent entity were correspondingly faster when the entity had appeared in a verb phrase versus compound. The referent was also introduced in two pragmatic texts; in one case, it was more closely related to the topic of its discourse than the other. Again, reading times for the sentence with the pronoun reflected accessibility, with faster reading times when the referent was more topical. In fact, when the referent entity was highly related to the discourse topic, reference in the compound condition was not significantly more difficult than reference in the verb phrase condition.

These results validate the claim that short term memory for text comprehension contains a representation of the relative accessibilities of discourse entities, accessibilities that are jointly determined by pragmatic and syntactic factors. The results also support the claim that naturally occurring examples of antecedents in compounds

(e.g., *Kal Kan* in the compound *Kal Kan cat*) are grammatically well formed and that they are neither "performance errors" nor the result of some type of pragmatic salvaging of otherwise ungrammatical constructions (cf. Ward et al., 1991). The fact that such examples are frequently produced in natural discourse does not necessarily entail that they are understood by the hearer/reader. But the psycholinguistic data presented in this article indicates that they are and that they are subject to the same types of pragmatic variables as are other kinds of anaphora. The pragmatic variable in our experiments, topicality, affected reference for both compound and noncompound constructions. Furthermore, placing the word that evokes the referent entity in a compound internal position reduced the accessibility of that entity, just as a modifier position reduces the accessibility of other entities (McKoon et al., in preparation; Rothkopf et al., 1986; Rothkopf et al., 1988).

The results from the six experiments in this article, taken as a whole, also demonstrate the importance of using converging kinds of experimental data. It would not be possible for us to support our conclusions from measurements of sentence reading times alone. For example, we found that reading times were slowed when the referent entity for the pronoun in the final sentence was introduced within a compound. But we could have found that reading times in this condition were quite fast; this could have happened if the pronoun were uninterpretable and subjects quickly realized that it was uninterpretable. In this instance, the reading time data would have seemed to counter our hypotheses. However, an uninterpretable pronoun would have led to slow response times when the referent word was tested immediately after the final sentence, allowing us to correct what would have been erroneous conclusion.

Likewise, it would not be possible, with reading times and immediate testing alone, to conclude that a pronoun was completely understood such that all the relevant relationships among the referent entity and information in the discourse were encoded into long term memory. The delayed testing priming results are required for that conclusion. Thus, only by simultaneous consideration of the sentence reading times, the test word response times, and the priming results can our interpretations of sentence reading times be fully jusitified.

Through these converging sets of data, we argue that the difficulty of comprehension for a pronoun depends on the accessibility of the discourse entity to which the pronoun is being used to refer. Pronoun comprehension is not viewed as a process that depends on the pronoun alone or even primarily. The issue for the comprehension system is not how to use a pronoun to access the intended referent. Instead, the issue is how the discourse model is constructed from the discourse in such a way that pronouns can be automatically and correctly interpreted.

REFERENCES

BRANSFORD, J. D., BARCLAY, J. R., & FRANKS, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology*, 3, 193–209.

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

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

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

CORBETT, A. T., & DOSHER, B. A. (1978). Instrument inferences in sentence encoding. *Journal of Verbal Learning and Verbal Behavior*, 17, 479–491.

FODOR, J. D. (1989). Empty categories in sentence processing. *Language and Cognitive Processes*, 4, 155–209.

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

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

GREENE, S., MCKOON, G., & RATCLIFF, R. (1992). Pronoun resolution and discourse models. *Journal*

of Experimental Psychology: Learning, Memory and Cognition, **18**, 266–283.

GROSZ, B. (1977). The representation and use of focus in a system for understanding dialogs. In *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*, Los Altos: William Kaufmann, Inc.

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

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

HEIM, I. (1982). *The semantics of definite and indefinite noun phrases*. Unpublished Ph.D. dissertation, University of Massachusetts at Amherst.

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

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

JOHNSON-LAIRD, P. (1983). *Mental Models*. Cambridge, MA: Harvard University Press.

KAMP, H. (1981). A theory of truth and semantic representation. In J. Groenendijk, T. Janssen, and M. Stokhof (Eds.), *Formal methods in the study of language. Part 1* (pp. 277–322). Amsterdam: Mathematical Centre Tracts.

KARTTUNEN, L. (1976). Discourse referents. In J. McCawley (Ed.), *Syntax and semantics VII: Notes from the linguistic underground* (pp. 363–386). New York: Academic Press.

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

KINTSCH, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, **95**, 163–182.

KINTSCH, W., & VAN DIJK, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, **85**, 363–394.

KINTSCH, W., & KEENAN, J. M. (1973). Reading rate and retention as a function of the number of propositions in the base structure of sentences. *Cognitive Psychology*, **5**, 257–274.

LAKOFF, G., & ROSS, J. (1972). A note on anaphoric islands and causatives. *Linguistic Inquiry*, **3**, 121–125.

LEVI, J. N. (1978). *The syntax and semantics of complex nominals*. New York: Academic Press.

LIEBER, R. (1984). Grammatical rules and sublexical elements. In *Papers from the 20th parasession on lexical semantics, Chicago Linguistic Society* (pp. 187–199).

MALT, B. (1985). The role of discourse structure in understanding anaphora. *Journal of Memory and Language*, **24**, 271–289.

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

MATTHEWS, P. (1974). *Morphology*. Cambridge, UK: Cambridge University Press.

MCKOON, G. (1977). Organization of information in text memory. *Journal of Verbal Learning and Verbal Behavior*, **16**, 247–260.

MCKOON, G., & RATCLIFF, R. (1980a). Priming in item recognition: The organization of propositions in memory for text. *Journal of Verbal Learning and Verbal Behavior*, **19**, 369–386.

MCKOON, G., & RATCLIFF, R. (1980b). The comprehension processes and memory structures involved in anaphoric reference. *Journal of Verbal Learning and Verbal Behavior*, **19**, 668–682.

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

MCKOON, G., & RATCLIFF, R. (1989). Assessing the occurrence of elaborative inference with recognition: Compatibility checking vs compound cue theory. *Journal of Memory and Language*, **28**, 547–563.

MCKOON, G., & RATCLIFF, R. (1992). Inference during reading. *Psychological Review*, **99**, 440–466.

MOHANAN, K. P. (1986). *The theory of lexical phonology*. Dordrecht: Reidel.

MORGAN, J. (1978). Toward a rational model of discourse comprehension. In D. Waltz (Ed.), *Proceedings of TINLAP-2* (pp. 109–114). New York: Association for Computational Machinery.

MORROW, D., BOWER, G., & GREENSPAN, S. (1989). Updating situation models during narrative comprehension. *Journal of Memory and Language*, **28**, 292–312.

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

MURPHY, G. L. (1985). Psychological explanations of deep and surface anaphora. *Journal of Pragmatics*, **9**, 171–198.

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

OAKHILL, J., GARNHAM, A., & VONK, W. (1989). The on-line construction of discourse models. *Language and Cognitive Processes*, **4**, 263–286.

POSTAL, P. (1969). Anaphoric islands. In R. Binnick et al. (Eds.), *Papers from the fifth regional meeting of the Chicago Linguistic Society* (pp. 205–255).

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

RATCLIFF, R., & MCKOON, G. (1978). Priming in item

recognition: Evidence for the propositional structure of sentences. *Journal of Verbal Learning and Verbal Behavior*, 17, 403–417.

RATCLIFF, R., & MCKOON, G. (1988). A retrieval theory of priming in memory. *Psychological Review*, 95, 385–408.

ROTHKOPF, E., BIESENBACH, B., & BILLINGTON, M. (1986). *Syntax violations as error feedback during rapid reading: Suggestions for a new readability measure supplement* (Technical memorandum). AT&T Bell Laboratories.

ROTHKOPF, E., KOETHER, M., & BILLINGTON, M. (1988). *Why are certain sentence constructions mnemonically robust for modifiers?* (Technical memorandum). AT&T Bell Laboratories.

SANFORD, A. J., & GARROD, S. C. (1981). *Understanding written language*. New York: Wiley.

SIDNER, C. (1979). *Towards a computational theory of definite anaphora comprehension in English discourse*. Cambridge, MA: MIT dissertation.

SIDNER, C. (1981). Focusing for interpretation of pronouns. *American Journal of Computational Linguistics*, 7.4, 217–231.

SIMPSON, J. (1983). *Aspects of Warlpiri morphology and syntax*. Cambridge, MA: MIT dissertation.

TANENHAUS, M. K., & CARLSON, G. N. (1990). Comprehension of deep and surface verb phrase anaphors. *Language and Cognitive Processes*, 5, 257–280.

VAN DIJK, T. A., & KINTSCH, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

WARD, G. L. (1985). *The semantics and pragmatics of preposing*. Philadelphia: University of Pennsylvania dissertation. Reprinted in 1988, New York: Garland.

WARD, G., SPROAT, R., & MCKOON, G. (1991). A pragmatic analysis of so-called anaphoric islands. *Language*, 67, 439–474.

WEBBER, B. (1979). *Elements of discourse understanding*. New York: Garland Press.

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

WILSON, D., & SPERBER, D. (1979). Ordered entailments: An alternative to presuppositional theories. *Syntax and Semantics*, 11, 299–323.

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

(Received July 17, 1991)

(Revision received February 18, 1992)

Psychological Review
1992, Vol. 99, No. 3, 440-466

0033-295X/92/$3.00

Inference During Reading

Gail McKoon and Roger Ratcliff
Northwestern University

Most current theories of text processing assume a constructionist view of inference processing. In this article, an alternative view is proposed, labeled the *minimalist hypothesis*. According to this hypothesis, the only inferences that are encoded automatically during reading are those that are based on easily available information, either from explicit statements in the text or from general knowledge, and those that are required to make statements in the text locally coherent. The minimalist hypothesis is shown to be supported by previous research and by the results of several new experiments. It is also argued that automatically encoded minimalist inferences provide the basic representation of textual information from which more goal-directed, purposeful inferences are constructed.

In reading, comprehension processes are generally assumed to combine information from two sources: explicit statements from the text being read and general knowledge already known to the reader. Interactions of information from these two sources produce the representation of a text that is encoded into memory. The issue addressed in this article is the extent to which these interactions lead to the encoding of inferences. We claim that there is only minimal automatic processing of inferences during reading. Our hypothesis is that readers do not automatically construct inferences to fully represent the situation described by a text. In the absence of specific, goal-directed strategic processes, inferences of only two kinds are constructed: those that establish locally coherent representations of the parts of a text that are processed concurrently and those that rely on information that is quickly and easily available. This minimalist claim is supported in this article with several new experiments and with conclusions drawn from a review of previous research.

For different readers, minimalist processing with little strategic processing will occur in different situations. For some readers, it might be a rare occurrence; for others, it might happen in such situations as reading a magazine on an airplane, reading the newspaper through the morning fog over breakfast, or reading texts in a psychology experiment. However, more often than not, readers do have specific goals, especially when learning new information from texts, and so they often engage in strategic processes designed to achieve those goals. The minimalist claim for these situations is that minimal inferences provide the database for more strategic processes. They provide the database for strategic inferences that are constructed during reading, and they provide a minimalist representation of a text in memory from which strategic inferences can be constructed by retrieval operations.

The minimalist position is presented as an hypothesis from which to work toward explicit processing models. The hypothesis distinguishes between those inferences that are labeled *automatic* and those that are labeled *strategic*; however, this distinction is not always clear cut. In situations where a reader adopts special strategies, some strategic inferences may be easy to construct, perhaps nearly as easy as minimal inferences. Some strategic inferences may also be obligatory, in the sense that the text cannot be completely understood without them (Gerrig, 1986). It is our hope that an understanding of what information is provided quickly and automatically will provide the basis for an understanding of which effortful strategic and goal-based processes are relatively easy to construct and which more difficult. In fact, if a strict automatic–strategic demarcation is not eventually tenable, then the product of the minimalist program will be a set of results that label inferences in terms of speed of availability, ease of processing, probability of occurrence, and dependence on contextual environment. These results are critical in the development of processing models.

For present purposes, an *inference* is defined as any piece of information that is not explicitly stated in a text. This definition includes relatively simple inferences as well as complex, elaborative inferences and inferences that add new concepts to a text as well as those that connect pieces of the text. For example, by this definition it would be an inference to encode the relation between a pronoun and its referent or to encode two instances of the same word as referring to the same concept. It would also be an inference to compute *2* as the referent of *the number that is four less than the product of three times two* or to combine the clues of a mystery novel to give the murderer. Defining *inference* this broadly emphasizes the different degrees of processing that are required to produce different inferences. Some inferences

Experiment 1 was prompted in part by discussions with Al Collins and Ed Smith. This research was supported by National Science Foundation (NSF) Grant 85-16350 and Air Force Office of Scientific Research Grant 90-0246 (jointly funded by NSF) to Gail McKoon and by NSF Grant 85-10361 and National Institute of Mental Health Grants HD MH44640 and MH00871 to Roger Ratcliff. We thank Tom Trabasso for stimulating discussions about Experiment 3 and for comments that led to Experiment 4. We also thank Richard Gerrig, Art Glenberg, and Jan Keenan for their many helpful comments and Art Glenberg for the materials used in Experiment 5.

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

seem to be made automatically, without awareness. Others seem to involve conscious, problem-solving types of processing.

The automatic inferences that are the focus of this article are assumed to be supported by information that is quickly and easily available, and this kind of information is assumed to come from one of two sources: well-known information from general knowledge and explicit information from the text being read. Inferences based on general knowledge have been demonstrated in the encoding of such inferences as elaborations about "what will happen next" in a story *if* what will happen next is very predictable, the encoding of inferences about aspects of the meanings of words *if* they are highly typical aspects, the encoding of inferences about instances of categories *if* the instances are highly typical, and so on. For inferences based on explicit textual information, the information may be in short-term memory or it may be easily retrievable from the long-term memory representation of the text that is under construction.

Inferences based on explicit textual information are used to establish local coherence for a text. These inferences include connections among instances of the same concept, pronominal reference, and perhaps causal relations. *Local coherence* is defined for those propositions of a text that are in working memory at the same time; in other words, propositions that are no farther apart in the text than one or two sentences. Many of the inferences that establish local coherence are based on information that is easily available because it is in short-term memory. Other local inferences, such as the relation between *the dog* and *the collie*, are based on combinations of explicitly stated information and well-known general knowledge. In either case, inference processes are assumed to proceed automatically. Only when neither explicit short-term memory information nor general knowledge leads to a coherent local representation of a text are other processes, perhaps strategic, problem-solving types of processes, engaged to provide local coherence.

According to the minimalist position, only the two classes of inferences, those based on easily available information and those required for local coherence, are encoded during reading, unless a reader adopts special goals or strategies. Automatically processed inferences are the main focus of this article for two reasons. First, they represent the most controversial point of debate between advocates of a minimalist position and advocates of a more constructionist view of text processing. There are many potential inferences that would be automatically generated during reading according to constructionist theories but not according to a minimalist view.

Second, although much of reading may have as its goal the generation of strategic inferences (e.g., in education, problem solving, planning, or decision making), these inferences must depend on the information automatically provided by a text. Automatic inferences are those that are encoded in the absence of special goals or strategies on the part of the reader, and they are constructed in the first few hundred milliseconds of processing. They therefore merit attention because they form the basic representation of a text from which other, more purposeful, inferences are constructed. In terms of theory development, our aim is to understand what kinds of information are quickly and easily available. Such an understanding is required to build processing accounts of the construction of automatic inferences. In turn, representational and processing models for automatic encoding would optimally serve as the starting point for explanations of more strategic encoding processes.

It is interesting to note the history of our approach to this minimalist position. About 12 years ago, we began experiments (prompted by discussions with Ed Smith and Al Collins) designed to demonstrate the use of goal hierarchies during reading (e.g., Experiment 1 discussed later). After a series of eight experiments, we could find evidence for the use of local goals but no evidence at all for the use of higher order goals. It was only much later, after several years and a number of other results (e.g., McKoon & Ratcliff, 1986), that we finally came to adopt the minimalist position.

The minimalist position contrasts with the framework that underlies most previous and current psychological investigations of inference processing during reading. Modern investigators began with the studies of Bransford and Franks and their colleagues, who adopted a strong constructionist approach to text processing (Bransford, Barclay, & Franks, 1972; Bransford & Franks, 1971; Johnson, Bransford, & Solomon, 1973). They interpreted their experimental results as demonstrating that encoding processes constructed inferences that were necessary to represent the situation described by a text. For example, a complete description of the sentence "Three turtles rested on a floating log, and a fish swam beneath them" would include the inference that the fish swam under the log. From the constructionist framework, this inference should be automatically encoded. From the minimalist position as proposed in this article, the inference would not be automatically encoded because it is not necessary to achieve local coherence, nor is the information that the fish swam under the log general knowledge.

Following Bransford et al.'s (1972) early work, constructionist hypotheses were advocated and tested by Richard Anderson and his colleagues (R. C. Anderson & Ortony, 1975; R. C. Anderson et al., 1976) and currently are embodied in some mental models approaches to text processing (Black & Bower, 1980; Glenberg, Meyer, & Lindem, 1987; Johnson-Laird, 1980; Mandler & Johnson, 1977; Morrow, Greenspan, & Bower, 1987; Rumelhart, 1977; Stein & Glenn, 1979; Trabasso & van den Broek, 1985; van Dijk & Kintsch, 1983). These models propose that the automatically encoded, mental representation of a text is a model of the situation described by the text. The representation is supposed to contain many nonminimal inferences, including elaborations on explicitly stated pieces of information and global connections among propositions. These constructionist models stand in direct opposition to the minimalist approach.

In this article, support for the minimalist position is provided in three ways. The first section of the article demonstrates a contrast between inferences that are constructed for local coherence and inferences that might be constructed to combine more global elements of a text. Several constructionist theories of text processing propose that global inferences are automatically constructed to connect pieces of information that are widely separated in a text; global inferences provide the overall structure of the text, such as the framework of a typical fairy tale or the causes of characters' actions. For local inferences, a review of recent research shows that several kinds are encoded during reading, as would be expected from a minimal-

ist theory. In contrast, the results of Experiments 1 through 4 show that causal global inferences are not automatically encoded, in contradiction to some global theories.

A second body of research that supports the minimalist position is research that has examined elaborative inferences. These inferences represent information that is not required for local coherence. For example, semantic inferences might add contextually appropriate features of meaning to the representation of a concept, instrumental inferences might add the typical instrument for a verb (e.g., *spoon* for *stirring coffee*), and predictive inferences might add information about "what should happen next" in a story. A review of previous studies shows that, for instrumental and predictive inferences, the data contradict the constructionist hypothesis and support the minimalist hypothesis. For inferences about the contextually appropriate meanings of words, the data are consistent with both hypotheses.

Finally, several studies that examined the use of lifelike situation models during reading are considered. It has been proposed that a situation model represents textual information in a way that corresponds to a "real-life" situation (cf. Glenberg et al., 1987). For example, for a character described in a text as moving from one room to another, the situation model would automatically keep track of the character, associating the character first with the objects in one room, then the next room, and so on as the character moved (Morrow, Bower, & Greenspan, 1989). In the third section of this article, studies designed to demonstrate the automatic encoding of lifelike situation models are shown to have alternative interpretations, and a new experiment demonstrates the plausibility of one such interpretation. The alternative interpretations are consistent with the minimalist view, and no elaborated situation model is required.

The remarkable conclusion to be drawn from both the new experiments and the review of previous experiments is that the widely accepted constructionist view of text processing has almost no unassailable empirical support (see also Alba & Hasher, 1983). The constructionist view has been discussed and tested for the past 20 years. Yet, it is difficult to point to a single, unequivocal piece of evidence in favor of the automatic generation of constructionist inferences. In the General Discussion section, we suggest that future research should investigate a variety of kinds of inferences, aiming toward a deep understanding of the processing and informational bases of each kind. We suggest that such investigation will lead to a gradual expansion of the kinds of inferences identified as minimal: The immediately available information in short-term memory may be more complexly structured than originally supposed, and the immediately available information from general knowledge may be more varied than we now believe. It is the goal of the minimalist hypothesis to motivate this expansion.

It should be stressed that the minimalist and constructionist positions disagree on the question of what inferences are encoded automatically, as the basis for more strategic inferences or when readers do not have special goals and strategies, and that it is these automatic inferences that are the topic of this article. All of the inferences that might be (and often are) strategically generated as the result of special goals adopted by motivated readers are critically important to language understanding, problem solving, and learning. The minimalist position separates these inferences from minimal inferences, and so they are outside the scope of this article. However, at some point the connection must be made between the mental representations provided by minimal inferences and the processes that operate on them to form strategic inferences, and the issue must be addressed of how minimal inferences support other kinds of inferences. These problems are no less important than those described in this article.

Local Versus Global Coherence

The minimalist hypothesis makes an important distinction between the inferences that are required to establish local coherence and those that might connect more globally separated pieces of information. This distinction is not one that would be made from a constructionist viewpoint; a constructed representation of the situation described by a text would not necessarily include aspects of the situation that were mentioned in close proximity, and it would not necessarily exclude aspects that were more widely separated. However, in support of the minimalist position, the distinction is clearly apparent in the results of empirical studies. On the one hand, there is a large body of evidence favoring the hypothesis that local inferences are automatically generated. On the other hand, there is little evidence for the automatic generation of global inferences during reading, and Experiments 1 through 4 provide explicit evidence that one kind of global inference, causal inferences, is not generated.

Local Coherence

A major claim of the minimalist view is that inferences are constructed during reading to the extent that the information on which they depend is readily available. If the required information is not readily available, then an inference will not be constructed (unless the text is not locally coherent). An obvious potential source of readily available information is the information in short-term memory, and so it is hypothesized that inferences based on this information are automatically constructed. To support the minimalist position, it must be shown both that the supporting information is readily available and that the supported inference is encoded.

For the processing of text through short-term memory, we follow the model proposed by Kintsch and van Dijk (1978), although for other purposes we would update this model to the more complex representations of discourse models (cf. Greene, McKoon, & Ratcliff, 1992; Grosz, Joshi, & Weinstein, 1983; Sidner, 1983a, 1983b; Ward, Sproat, & McKoon, 1991; Webber, 1983). In Kintsch and van Dijk's model, the information in short-term memory during reading is assumed to be made up of explicitly stated words of the text plus the propositions that are being formed from them. The amount of information in short-term memory at any point in reading a text is loosely defined to be several clauses or sentences, depending on their length (cf. Daneman & Carpenter, 1980). The relevant issue for the current discussion is not an exact specification of the amount of information in short-term memory at any point in processing, but rather the contrast between information that can be described as being locally available and global informa-

tion. When local inferences are examined empirically, they involve pieces of explicitly stated information that are close together in a text. When global inferences are examined empirically, they involve pieces of information that are so widely separated in the text that it is clear they could not be in short-term memory at the same time (without retrieval from long-term memory).

In the Kintsch and van Dijk (1978) model, the clauses in short-term memory are converted into semantic propositions. These propositions are connected together through overlap of their arguments, and they are ordered with respect to the most salient or topical proposition. Sentences and clauses do not usually provide an explicit representation of their underlying propositions and the connections among them; this information must often be inferred. For example, in the sentence "The mausoleum that enshrined the czar overlooked the square" the propositions are (roughly and informally) *mausoleum enshrined czar* and *mausoleum overlooked square*, where the two propositions refer to the same mausoleum. To form the appropriate locally coherent structure for the sentence, it must be encoded that the mausoleum both overlooked the square and enshrined the czar. In Kintsch and van Dijk's model, the processes that construct propositions are assumed to recognize that different occurrences of the same argument are in fact the same however the argument might be referenced (e.g., by a noun or an anaphor). Thus, the model assumes the encoding of the basic inferences necessary to form propositions through argument overlap. The minimalist view incorporates these inferences because they are based on the easily available information of short-term memory.

Empirical evidence confirms the assumption that inferences necessary to establish argument overlap are encoded. The encoding of inferences that establish propositional semantic units is well documented. Recall of a text depends on the number of propositions in the text (Kintsch & Keenan, 1973), and propositional units tend to be recalled as a whole (Kintsch & Glass, 1974). However, recall studies do not provide completely convincing evidence about encoded structures; with unlimited time in free recall, subjects may edit their responses to make them seem grammatical (i.e., by deleting incomplete propositions). Other evidence about propositional structures comes from priming studies with recognition memory. Ratcliff and McKoon (1978; see also McKoon & Ratcliff, 1980b; Ratcliff & McKoon, 1981b) gave subjects short lists of sentences to study. After each study list, subjects were given a recognition test list made up of single words from the sentences and unrelated distracter words. A subject's task was to decide as quickly as possible if each word in the test list had or had not appeared in a studied sentence. If a target test word from one of the sentences was immediately preceded in the test list by another word from the same sentence, then response time for the target was speeded. This priming effect was significantly greater if the two words from the sentence were from the same proposition than if they were from different propositions. For example, for the mausoleum sentence, response time for the target SQUARE was faster when *square* was primed by *mausoleum* from the same proposition than when it was primed by *czar* from the other proposition. These propositional priming effects have been shown to be due to automatic retrieval processes (Ratcliff & McKoon, 1981a), indicating that the structures reflected by priming were encoded during reading.

Evidence . . inferences to establish propositional units are encoded during reading supports the minimalist position only if it can also be shown that the information on which the inferences are based is easily available. Studies that indicate immediate availability are provided in recent work by Swinney and his colleagues (cf. Nicol & Swinney, 1989; Swinney & Osterhout, 1990), who used a cross-modal on-line lexical decision task. They used sentences like "The policeman saw the boy that the crowd at the party accused of the crime." In this sentence, *boy* should be encoded as the person who was accused (in the proposition *crowd accused boy*), and so *boy* should be quickly available after the word *accused*. To test this, sentences were presented auditorily, and at various points during the sentences, lexical decision test items were displayed visually. The lexical decision test items were strong associates of critical words in the sentences. The reasoning was that there should be facilitation in response time for an associate at any point where its related critical word was being used in comprehension. For example, the lexical decision for an associate of *boy* should be facilitated after the word *accused* because *boy* is the object of *accused*. The data showed this result and also that the associate was not facilitated after the word *party*, a point in the sentence where *boy* would not be used in building the underlying structure of the sentence. Similar evidence of immediate availability has been reported by Tanenhaus, Carlson, and Seidenberg (1985) and by Ga-nsey, Tanenhaus, and Chapman (1989). This evidence is all consistent with the idea that the information necessary to make connections among propositions is quickly available. The total combination of evidence—that inferences about propositional connections are encoded (Ratcliff & McKoon, 1978) and that the information on which they depend is quickly available (Nicol & Swinney, 1989)—exactly fits the minimalist hypothesis.

A second kind of inference that is often needed to establish argument overlap is the connection between an anaphor and its referent. If a text mentions some pronoun and predicates information about the pronoun, then the information about the pronoun should be connected to a referent of the pronoun and to other information given by the text about that referent. The processing of coreference has been extensively studied. For example, Corbett and Chang (1983; also Chang, 1980; Clark & Sengul, 1979; Ehrlich & Rayner, 1983) used sentences like "Rachel tried to catch Sally, but she was not able to do it," with the possible referents of *she* presented for recognition test at the end of the sentence. They found that responses to the intended referent were faster than responses to the unintended referent (but see Gernsbacher, 1989; Greene, McKoon, & Ratcliff, 1992). Nicol (1988, cited in Nicol & Swinney, 1989) has demonstrated the availability of potential referents of pronouns more immediately than at the end of sentences. She used a cross-modal on-line lexical decision task (as presented earlier), and sentences like "The boxer told the skier that the doctor for the team would blame him for the recent injury." When test words were presented immediately after the pronoun *him*, there was facilitation of response times for associates of the potential referents of the pronoun (*boxer* and *skier*). However, there was no facilitation for an associate of the noun that could not be a referent (*doctor*). This pattern of data is consistent with information

about potential referents being quickly available, and so the result is consistent with the minimalist hypothesis.

A more stringent test of the minimalist position would be a combination of studies that showed both the encoding of appropriate connections between referent and anaphor and the immediate availability of the information that supports the connections. Such studies have not been done for pronouns, but they have been done for nominal anaphors (Dell, McKoon, & Ratcliff, 1983; McKoon & Ratcliff, 1980a). These experiments used short texts that, in the first sentence, mentioned a character such as a burglar. "A burglar surveyed the garage set back from the street. Several milk bottles were piled at the curb. The banker and her husband were on vacation. The criminal/A cat slipped away from the streetlamp." In the last sentence, either the character introduced in the first sentence was referenced again with a category label (*the criminal*), or a new character (*a cat*) was introduced, with no mention of the character from the first sentence. When the last sentence referred to the burglar as *the criminal*, information about the burglar should have been directly connected from the first sentence to the last sentence. McKoon and Ratcliff (1980a) showed that these connections were encoded using recognition priming. Subjects were given study lists of texts to read. After each study list, they were given single words for recognition. Among the test words was a noun from the last sentence, and it was immediately preceded in the test list by the character from the first sentence (e.g., *streetlamp* immediately preceded by *burglar*). When the noun and the character were directly connected together in the text by the anaphor (*The criminal*), response times on the noun were speeded relative to when the noun and the character were not directly connected (when it was the cat that slipped away from the streetlamp). This result shows the encoding of connections based on anaphoric inferences.

Results indicating the immediate availability of information supporting the connections were obtained by Dell et al. (1983). They used a word-by-word reading procedure in which each word of a text was displayed for 250 ms, and recognition test words could be presented after any word of the text. One test point was immediately after the first noun of the last sentence (the anaphor *criminal* or the word *cat*). At this test point, response times to the antecedent (*burglar*) and to another word from the same proposition as the antecedent (*garage*) were both facilitated in the criminal version of the last sentence relative to the cat version, consistent with immediate availability of the referent for the anaphor. Corbett (1984) also found results that indicate the immediate availability of potential referents for anaphors using a different paradigm. He found that reading times for anaphors like *wooden toy* were faster when there was only one possible referent in the text (*wooden block*) than when there was also a nonreferent from the same general category (*rubber ball*). Thus, taken in combination, these studies support the minimalist hypothesis by showing that the information necessary to establish anaphoric connections is available immediately during reading.

In the van Dijk and Kintsch (1983) processing model, the propositional connections established by repetitions of concepts and anaphoric relations are the only means of establishing local coherence. However, as Kintsch and van Dijk point out, propositional connections are not sufficient to guarantee coherence. Keenan, Baillet, and Brown (1984) made this point with the sentence pair "Tom Jones plans to go to the dentist. A plane flew over Tom Jones." According to the minimalist position, inferences will be encoded if they are required for local coherence. The problem is to define exactly what constitutes local coherence. No formal definition is available, although researchers have made several suggestions. Lack of a formal definition does not mean that local coherence cannot be investigated empirically. Other concepts in psycholinguistics that lack formal definitions (such as *proposition*) have been used to excellent advantage (cf. Kintsch & Keenan, 1973; Kintsch, Kozminsky, Streby, McKoon, & Keenan, 1975; Kintsch & van Dijk, 1978), and empirical investigation should lead to more formal descriptions and definitions of local coherence. For present purposes, we assume that a set of two or three sentences is locally coherent if it makes sense on its own or in combination with easily available general knowledge. It is not locally coherent if information from elsewhere in the discourse is required.

Suggestions for the kinds of inferences that might be involved in local coherence include bridging inferences and causal inferences. Haviland and Clark (1974) outlined several kinds of bridging inferences, and Keenan and Kintsch (1974; also McKoon & Keenan, 1974) provided data to indicate that bridging connections are encoded into the memory representation of a text. An example of a text used by Keenan and Kintsch is "Police are hunting a man in hiding. The wife of Bob Birch disclosed illegal business practices in an interview on Sunday." For this text, a bridging inference is required to provide the relation between Bob Birch and the man in hiding. Keenan and Kintsch found evidence that this inference is encoded during comprehension. They used a verification test (given 15 min after the text was read). Response times for the statement "Bob Birch is the man who is hiding" were just as fast for the text that required the bridging inference as for another version of the text that made the inference explicit. From this result, Keenan and Kintsch argued that this kind of bridging information was encoded during reading. Whether the result is fully consistent with the minimalist position is not clear. The information that Bob Birch is the man who is in hiding is not known before reading the text, and so it would not be quickly and easily available. Therefore, the minimalist prediction would be that it was constructed by a relatively slow inference process; this prediction has not been tested.

Another potential contributor to local coherence is causality; propositions that are in short-term memory at the same time have been said to be connected by their causal relations. One way to demonstrate the importance of causal relations would be to show that causally relevant propositions are preferentially maintained in short-term memory during reading. Fletcher, Hummel, and Marsolek (1990) found evidence for such maintenance, although it could be argued that, with their materials, causally relevant propositions were maintained in short-term memory by virtue of (anaphoric) repetitions of their content rather than by virtue of their causality.

Other demonstrations of the effects of causal relations have used pairs of sentences that were designed to vary in their causal relatedness. Keenan et al. (1984; see also Bloom, Fletcher, van den Broek, Reitz, & Shapiro, 1990; Myers, Shinjo, & Duffy, 1987) found that the reading time for the second sen-

tence of a pair was slowed as the causal relatedness of the pair was decreased. There are two possible interpretations of this result: One is that reading time was slowed by the process of constructing (or attempting to construct) a causal chain to relate the two sentences—less related sentences require the construction of a longer chain. The other interpretation is that reading time slowed because of difficulty in finding an already existing causal chain in long-term memory. By this interpretation, closely related sentences are causally connected through a relation provided by long-term memory. The causal chain that connects two closely related sentences may be long or short, but it will be quickly processed because it is already available and does not have to be constructed. Less closely related sentences would represent a mixture of processes, some connected by difficult-to-access relations in long-term memory, some connected by newly constructed relations, and some perhaps left without any causal connection.

Given these different interpretations, it is not clear whether the causal connections investigated in these studies were encoded automatically. From the minimalist point of view, the causal relations encoded automatically during reading should be those that are quickly available from long-term memory; those that are not available from long-term memory but are required to establish local coherence should also be encoded. This claim has not been tested empirically. One problem is to define what causal inferences are necessary for coherence; we return to discussion of this problem after considering research on global inferences.

Leaving aside the uncertain situation with causal relations, the minimalist hypothesis is well supported with respect to local coherence: Current data are consistent with the claims that inferences based on quickly available information are encoded during reading. The minimalist position would be contradicted if it could be shown that some inference was encoded even though it was neither quickly available nor necessary for local coherence. The minimalist position would also be contradicted if it could be shown that there were kinds of quickly available information that did not support inferences. However, there is no such evidence to contradict the minimalist claims. In the next section, we show that the situation for global inferences is much different than that for local inferences. Although the local inferences for propositional structures posited by the minimalist view are relatively easy to demonstrate empirically, there is no evidence that global inferences for global structures are automatically generated during comprehension.

Global Inferences

Many researchers have proposed that global inferences connect widely separated pieces of textual information and that they do so automatically as a necessary part of comprehension. Sometimes these inferences are analyzed as the linking elements of a story "grammar" so that initiating settings, characters, goals, and events are linked to their consequent events and outcomes (Mandler, 1978; Mandler & Johnson, 1977; Rumelhart, 1975; Stein & Glenn, 1979; Thorndyke, 1977). More often, global inferences are the links that connect explicit pieces of information into an overall causal chain or network (Black & Bower, 1980; Graesser, 1981; Graesser, Robertson, & Anderson, 1981; Omanson, 1982a, 1982b; Trabasso & van den Broek, 1985; Trabasso & Sperry, 1985). From the minimalist point of view, these inferences should not be automatically constructed during reading. They are usually not required to establish local coherence, and they are usually not supported by well-known information. Only if a text is locally incoherent at some point should global information be recruited to establish local coherence. Of course, readers will often construct global inferences when such inferences are required by the readers' goals. Minimalist inferences will be constructed in the absence of special goals or strategies and to provide the bases for goal-driven inferences.

Experiments 1 through 4 examined whether global causal inferences are generated automatically during comprehension. Because the experiments directly challenge the hypothesis that global inferences are encoded automatically, it is necessary to explain clearly what kinds of inferences are both causal and global. As an illustration, we use the method of analysis of causal relations developed by Trabasso and his colleagues (cf. Trabasso & van den Broek, 1985).

Table 1 shows a short story and its analysis, adapted from an article by Suh and Trabasso (1988). The meaning of each sentence in the story is identified as setting, initiating event, goal, action, outcome, or reaction. These elements make up the definition of an *episode*. For an episode to occur, there must be a setting in which it occurs, one or more initiating events in the setting, and reactions to the events. If the reactions lead to a goal, then one or more actions will result, and they in turn will have outcomes. This episode structure is recursive in that outcomes may provide the initiating events for further reactions, goals, and outcomes. The definition of the episode structure requires that each goal be linked directly to its initiating event or events and each outcome be linked directly to the goal it fulfills (or fails to fulfill). It is assumed that these direct links must be encoded during reading. If the links are not explicitly stated, then they will be inferred. If the necessary pieces of information to create the links are not locally available, then they will be retrieved from memory. The links between initiating events and goals, and between goals and outcomes, that are assumed for the story in Table 1 are shown at the bottom of the table. The mother's birthday is the initiating event for the goal of wanting to buy a present, and the outcomes of this goal are that everything was too expensive and no present was bought. These outcomes plus the original initiating event, the birthday, provide the initiating events for the second goal, knitting a sweater. For this second goal, an inference is required, namely, that the sweater was to be the mother's birthday present. This is labeled a *global inference* if it is the case that the initiating event, the mother's birthday, is no longer available in working memory when the second goal is read. The specific analysis for the story in Table 1 is from Trabasso and van den Broek (1985), but other causal analyses (e.g., Black & Bower, 1980; Graesser, 1981; Mandler & Johnson, 1977; Omanson, 1982a, 1982b; Rumelhart, 1975; Stein & Glenn, 1979; Thorndyke, 1977) would also assume that the inferred link between the birthday and knitting the sweater was encoded during reading into the mental representation of the story.

A number of empirical results have been obtained that are consistent with causal analyses of stories. The largest body of

Table 1
A Short Story From Suh and Trabasso (1988)

Setting: *Once there was a girl named Betty.*
Initiating Event 1: *One day, Betty found that her mother's birthday was coming soon.*
Goal 1: *Betty really wanted to give her mother a present.*
Action: *Betty went to the department store.*
Outcome 1: *Betty found that everything was too expensive.*
Outcome 2: *Betty could not buy anything for her mother.*
Reaction: *Betty felt sorry.*
Initiating Event 2: *Several days later, Betty saw her friend knitting.*
Setting: *Betty was good at knitting too.*
Goal 2: *Betty decided to knit a sweater.*
(*Story continues*)
Goal 1 is linked directly to its Initiating Event 1.
Outcomes 1 and 2 are linked directly to their Goal 1.
Goal 2 is linked directly to its initiating events, which are Initiating Event 1 and Outcomes 1 and 2.

data comes from recall studies. The probability of recalling any particular fact can be predicted from its position in the causal network representation of its story (cf. Black & Bower, 1980; Omanson, 1982a; van den Broek, 1988). Causal information that is on a direct causal chain from the beginning of a story to the end is more likely to be recalled than information that is not on the chain (Trabasso & van den Broek, 1985; van den Broek & Trabasso, 1986). Also, the probability of recalling a piece of information increases with the number of causal connections it has to other pieces of information (Trabasso & Sperry, 1985).

These recall findings have often been cited as support for the hypothesis that global causal inferences are encoded during reading. However, recall does not necessarily measure encoding. It may be that recall sometimes gives an accurate measure of encoded information, but it may also measure the results of the retrieval and editing processes that operate on encoded information, and these processes may give nonrandom distortions of the encoded information. For recall of stories, it is easy to see that subjects might edit the facts of their encoded representations into causally connected structures, eliminating facts that they remembered but decided not to write down and working extra hard to remember facts that would turn an otherwise unrelated list of sentences into a coherent story. Thus, the causal structures found in recall protocols may be a reflection of editing processes and not an accurate reflection of the representation in memory that was formed by encoding processes.

This point is reinforced by empirical demonstrations of the roles of retrieval and editing processes. For example, Alba, Alexander, Hasher, and Caniglia (1981) showed that subjects could recognize statements from stories for which they knew the topic as well as they could recognize statements from stories for which they did not know the topic, even though recall was much worse when they did not know the topic. Another clear example of the operation of retrieval processes is provided in a study by Singer (1978). He showed that the effectiveness of a cue for recall was determined by backward associations at the time of recall from the cue back to the text to be recalled, not forward associations inferred when the text was read. Other results by Corbett and Dosher (1978) and Baillet and Keenan (1986) also demonstrate that recall experiments do not provide convincing evidence that inferences are generated during reading.

The processes that can be involved in recall, including editing and inference generation, are important processes to study, but they are not the focus of this article. Our aim is to separate out and focus on the inferences that are automatically included in a text representation at encoding. In this way, a clearer demarcation can be drawn between processes that occur at encoding and those that can occur at retrieval.

In Experiments 1 through 4, we use experimental procedures other than recall to compare causal global inferences to inferences based on locally available information. From the minimalist hypothesis, we expected that global inferences would not be automatically encoded during reading. This finding is also predicted by results from experiments by Glanzer, Fischer, and Dorfman (1984). They interrupted subjects' reading in the middle of a text and gave them an unrelated task to perform. When the subjects resumed reading the text, the best aid to comprehension was not global information about the topic of the text, but local information from the context immediately preceding the interruption.

Empirical Tests for Global Causal Inferences

The basic hypothesis that runs through all of Experiments 1 through 4 is that, barring special strategies by readers, causal global inferences are not constructed if a text is locally coherent. Only when a text is not locally coherent will global information be brought in to aid comprehension. Of course, readers can and often do adopt special strategies, either during reading or recall, to involve global information in local processing. However, in the typical laboratory experiment without special instructions, such strategies do not appear to be used during reading.

The hypothesis that global inferences are not automatically constructed for locally coherent texts is suggested by consideration of simple examples. Suppose a story relates that, when a killer's rifle won't work properly, he reaches for his hand grenades. This sequence of events makes sense without global knowledge of the killer's goal, to assassinate a president. On the other hand, if a text is not locally coherent, then global information should be used. When a character in a story decides to buy fruit and yogurt as a result of finding her bicycle broken, a reader needs the global information that she is trying to lose weight to make sense of the scenario.

Experiment 1 contrasted the availability of local and global information during reading of short texts. Causal global inferences were identified using the definitions given by Trabasso (Suh & Trabasso, 1988; Trabasso & Sperry, 1985; Trabasso & van den Broek, 1985) and described earlier. All of the texts were locally coherent, and results indicated that local information is available during comprehension. The texts did not require global causal information for coherence at the local level, and results indicated that it was not used. Experiment 2 extended these results with texts of two types. One type was coherent at the local level, but local information contradicted global information. The data showed no effects of this contradiction. The second type of text was not coherent locally, although it could be made coherent through global information. In this case, the data showed that global information does become available for use at the local level.

Experiments 3 and 4 used long, naturalistic stories to investigate the representations of inferences in memory. The cause of some specific event in a story was separated from the event by several paragraphs. The empirical question was whether global inferences would connect the event to its cause in the memory representation of the story. The data indicated that this does not happen. Thus, over all four experiments, there is no evidence that global causal inferences are constructed during reading.

Experiment 1

This experiment was designed to assess the availability of local and global information at the end of reading short texts. Each text had two paragraphs, an introduction paragraph and a continuation paragraph, as shown by example in Table 2. In the introduction, a general goal (e.g., killing the president) and a goal subordinate to the general goal (using a rifle) are described. For the continuation paragraph, there were three different versions: Control, Try Again, and Substitution. In the Control continuation, both goals are achieved (the president is shot), and a new goal is introduced. In the Try Again continuation, a problem arises in achieving the subordinate goal and the character tries this goal again (using the rifle in a different way). In the Substitution continuation, a problem also arises with the subordinate goal, but instead of trying again, the character replaces it with a new subordinate goal (hand grenades). The new subordinate goal, like the old one, is designed to achieve the original general goal (killing the president).

Subjects read each text one sentence at a time, at a pace they controlled themselves. Availability of a goal was tested by presenting a recognition test word for the goal immediately after the final sentence of the text.

For the general goal in the texts, the minimalist and constructionist positions make different predictions. All of the continuations were written to be coherent in themselves; the general goal is not needed to comprehend any of them. Thus, according to the minimalist prediction, the general goal should not be used during comprehension of any of the continuations, and so the availability of the general goal should be equal across the different continuations. Responses to the general goal test word should not differ across the continuations in speed or accuracy. In contrast, according to a constructionist theory, responses to the general goal test words should be faster in the Try Again and Substitution continuations than in the Control continuation. This is because the character in the text is still trying to achieve the general goal at the end of both the Try Again and Substitution conditions but not at the end of the Control condition (where a new general goal has taken over).

For the subordinate goal, the minimalist and constructionist positions can make the same predictions. Locally, the original subordinate goal is necessary for comprehension only in the Try Again continuation; in neither the Control condition nor the Substitution continuations is the original subordinate goal still necessary to understand the character's actions. Thus, responses to the subordinate test word should be faster, more accurate, or both in the Try Again condition relative to the other two. For a constructionist theory, the character is still trying to achieve the original subordinate goal in the Try Again continuation, and so responses should be facilitated in this condition relative to the Control. In the Substitution continuation, there might or might not be facilitation, depending on whether the switch to a new subordinate goal eliminated all facilitation for the original subordinate.

Table 2
An Example of a Story From Experiment 1

Part of story	Story
Introduction	The crowd's cheers alerted the onlookers to the president's arrival. The assassin wanted to kill the president. He reached for his high-powered rifle. He lifted the gun to his shoulder to peer through its scope.
Control continuation	The assassin hit the president with the first shot from his rifle. Then he started to run toward the west. The searing sun blinded his eyes.
Try again continuation	The scope fell off as he lifted the rifle. He lay prone to draw a sight without the scope. The searing sun blinded his eyes.
Substitution continuation	The scope fell off as he lifted the rifle. So he reached for his hand grenades. The searing sun blinded his eyes.

General goal test word: *Kill*
Subordinate goal test word: *Rifle*

Note. The labels of the parts of the stories were not presented to the subjects.

Method

Materials. Each of the 30 experimental texts was made up of an introduction and three different continuations. The introduction introduced a general goal for the main character in the story (e.g., killing the president for the text in Table 2) and a subordinate goal that was a way of obtaining the general goal (e.g., using a rifle). The general goal was mentioned only once in the introduction and was not mentioned explicitly in the continuations. The subordinate goal was mentioned once in the introduction and again in the first sentence of each continuation. The introductions were always four sentences in length. The general goal and the subordinate goal were used as test words (e.g., *kill* and *rifle*).

In the Control continuation, the first sentence described successful fulfillment of the subordinate goal and so, by implication, the general goal. Then, the second sentence described a new general goal for the character. Examples of the original general goals in the introductions and new general goals in the Control continuations include going out for an evening's entertainment and then finding out where to buy furniture, cleaning house and then painting a barn, eating and then back scratching, getting a front-page story and then moving a printing press, investing money and then stopping at a dry cleaners, and holding a sale and then going to Europe.

The second continuation, the Try Again condition, described a problem with fulfilling the subordinate goal and presented a new method for fulfilling the same subordinate goal. Examples of the new and old methods include having a lecture from a doctor instead of from a social worker, going somewhere by train instead of by car, adopting a baby through a lawyer instead of an agency, borrowing money from a bank instead of a relative, asking a sister to do something instead of a

friend, and going to cheerleader practice instead of going home. Note that in each case, the two methods of achieving the subordinate goal are coherent alternatives even though the general, superordinate goal for these examples is not given here. For example, having a lecture from a doctor instead of a social worker makes sense without knowing that the general goal is to obtain information about the world's population problems.

The third continuation, the Substitution condition, also described a problem with fulfilling the original subordinate goal and presented a new subordinate goal that would fulfill the original general goal. Examples include raking the lawn instead of trimming the hedge, a lecture about the world's food supply instead of about birth control, going to see fireworks instead of to the beach, giving money to a charity instead of to specific people, selling stocks instead of borrowing money, and going to a night club instead of to a movie. In each of these examples, the alternative makes sense without the general goal. For example, trimming the hedge can be understood as a substitute for raking the lawn without knowing the general goal of getting ready for a lawn party.

Each continuation was three sentences in length, and the final sentences of the three continuations were identical. The continuations were all locally coherent, as shown by the examples, in that they could make sense without knowledge of the general goal stated in the introduction.

In some of the Try Again and Control continuations, the second sentence contained words that might be semantically associated (preexperimentally) to one or the other of the test words. For example, the words *sight* and *scope* are associated with rifle. However, the number of items with such associations was about the same for the Try Again and Substitution continuations for both the general goal and subordinate goal test words. Thus, overall, associations between words in the texts and test words were equated across all conditions but the Control.

There were also 42 filler texts, each with one test word. Nine of the fillers were five sentences in length, 9 were six sentences in length, and 9 were eight sentences in length. For each length, six of the tests had positive test words and three had negative test words. The other 15 texts were seven sentences in length and had negative test words.

Procedure. The presentation of stimuli and collection of responses was controlled by a real-time computer system. Stimuli were displayed on a cathode-ray tube (CRT) screen, and responses were made by pressing keys on the CRT's keyboard.

The experiment began with a practice list of 8 texts, each one to three sentences in length. Then the 72 texts of the experiment proper were presented, eight fillers first and then the remaining texts in random order.

Presentation of each text began with an instruction displayed on the CRT screen asking the subject to press the space bar. When the space bar was pressed, there was a 200-ms pause, and then the first sentence of the text was displayed. The sentence remained on the screen until the subject pressed the space bar again; then the screen was cleared, there was a 500-ms pause, and then the next sentence of the text was displayed. Presentation of the sentences continued in this way until the final sentence of the text. After the final sentence was displayed and the space bar pressed, a row of asterisks appeared with a test word immediately below it. The subjects' instructions were to indicate whether the test word had appeared in the immediately preceding text, by pressing the "/" key for a positive response and the *z* key for a negative response. The test word was erased from the screen immediately after the response. If the response was incorrect, the word *ERROR* was presented for 2,000 ms, and then the screen was cleared and there was a pause of 200 ms. If the response to the test word was correct, then there was a 200-ms pause. After the pause, the response time for the test word was displayed for 800 ms, then there was a 500-ms pause, and the instruction to press the space bar to begin the next text was displayed. Subjects were instructed to read the texts carefully and to respond as quickly and accurately as they could to the test words.

Design and subjects. With one group of 18 subjects, the test word for the experimental texts was always the general goal; for the other group of 18 subjects, it was always the subordinate goal. The experimental texts were presented with the Control, the Try Again, or the Substitution continuations. This variable was combined with three sets of subjects (6 per set in each group) and three sets of texts (10 per set) in a Latin square design. The subjects participated in the experiment for credit in an introductory psychology course.

Results

The mean reading time for the final sentence of each text and the mean response times and error rates for each subject and each test word were calculated; means of these means are displayed in Table 3. There were no specific predictions about final sentence reading times; they are included for completeness. Subjects who were tested with the subordinate goal test words read faster than subjects who were tested with the general goal test words.

According to the minimalist local coherence position, only local information and not the original general goal is necessary for comprehension of the continuations. Thus, response time and accuracy for the general goal test words should not vary across the different continuations, as is shown in the data in Table 3. For the subordinate goal test words, in the Control and Substitution continuations, a new goal was substituted so that the original subordinate need no longer be involved in comprehension at the end of the continuations; as a result, response times for the subordinate test word should be relatively slow and/or inaccurate. In the Try Again continuation, the original subordinate goal was still necessary for comprehension of the character's actions, so response times should be relatively fast and/or accurate. This is the pattern of data shown in Table 3.

For the test words expressing the original general goal, analysis of variance (ANOVA) showed no significant differences in response times, error rates, or reading times for the final sentences. For the test words expressing the original subordinate goal, the response times were significantly different across the continuations, $F(2, 34) = 5.5$, with subjects as the random variable, and $F(2, 58) = 3.4$, with test words as the random variable. The standard error of the response times was 11.6 ms. There were no significant differences in error rates or in the reading times of the final sentences (Fs < 1).

Discussion

The critical comparison between the minimalist local coherence hypothesis and the global constructionist hypothesis rests in their predictions for the general goal test word. According to the constructionist hypothesis, the character in the text is still trying to achieve the general goal at the end of the Try Again and Substitution continuations, and so responses to the goal test words should be facilitated in these conditions relative to the Control condition. According to the minimalist hypothesis, responses for the general goal test words should not differ across the three conditions because all the continuations are locally coherent and none require the general goal for local

Table 3
Results From Experiment 1: Mean Response Times (in Milliseconds), and Error Rates for Test Words and Mean Reading Times (in Milliseconds) for Final Sentences

	General goal tested			Subordinate goal tested		
	Test words			Test words		
Type of continuation	RT	% error	Reading times	RT	% error	Reading times
Control	717	11	1,551	638	5	1,399
Try again	717	12	1,588	594	8	1,399
Substitution	718	12	1,585	644	7	1,337

Note. RT = response time.

comprehension. The data support the minimalist view because there are no differences across the conditions. The data for the subordinate goal test words do show differences across conditions, indicating that the experiment did not lack power.

Experiment 2

Experiment 2 was devised to provide additional tests for global inferences. The procedure was the same as in Experiment 1: Subjects read short texts sentence by sentence, and recognition test words were presented after the final sentence. Examples of the texts are shown in Table 4. There were two kinds of texts, Globally Inconsistent and Locally Inconsistent, each with an Introduction plus a Control continuation and a Problem continuation.

The first text is labeled *Globally Inconsistent.* This reflects the fact that, in the Problem continuation, watching videotapes is not consistent with the stated goal of working out an injured arm. The text provides a test for global inferences because the inconsistency should amplify the use of global information at the local level, and so responses to the test word *workout* should be facilitated relative to the Control condition. However, the Problem continuation, like the Control continuation, is locally coherent; neither requires use of the general goal for comprehension. If only local information is used in comprehension, then there should be no facilitation of *workout* in the Problem condition relative to the Control condition.

The second type of text is labeled *Locally Inconsistent* because replacing a broken bicycle with grapefruit and yogurt does not make sense on the local level. However, it does make sense in the global context of trying to lose weight. For this text, both the global inference and the minimalist positions agree: The global goal information about losing weight should be recruited during local processing, and responses to the goal test word (*weight*) should be facilitated in the Problem condition relative to the Control condition.

Method

Materials. Each of the experimental texts used in the experiment was made up of an introduction and two different continuations. The introductions, always four sentences in length, described some goal for the main character of the story (a workout in the first example in Table 4). This goal was mentioned explicitly only once in the introduction and not mentioned explicitly in either continuation. One word expressing the goal (e.g., workout) was used as the test word for the text. In the first continuation, the Control condition, the goal was fulfilled and a new goal described (the Control versions were similar to the Control versions used in Experiment 1). In the Problem continuation, some problem that prevented attainment of the original goal was described, and then a new goal was substituted. The final sentences of the two continuations were always the same, and all continuations were three sentences long.

There were two sets of experimental texts (20 in each set) that differed in the relation between the substitute goal in the Problem continuation and the original goal. In the Globally Inconsistent set of texts, the new goal was inconsistent with the original goal; some examples of new and original goals include fixing a lock in the attic instead of preparing the grounds for a lawn party, going to a restaurant instead of on a picnic, buying a conservative gown instead of buying something to look unusual, donating money instead of finding a cure for loneliness, watering the chickens instead of cleaning the house, buying a heated swimming pool instead of saving on electric bills, flying to Las Vegas instead of investing wisely, and serving take-out hamburgers instead of a sumptuous feast. In each case, the substituted goal cannot lead to achievement of the original goal—there is no way that take-out hamburgers can provide a sumptuous feast, and presenting the two goals in conjunction, as is done here, makes the inconsistency clearly apparent. What makes the inconsistency not obvious to readers of the texts is that the two goals are not simultaneously available. The continuation becomes locally coherent because there is a plausible relation between the problem and the substitute goal (e.g., take-out hamburgers are a plausible alternative when someone forgets to buy steak).

In the Locally Inconsistent set, the substitute goal was consistent with the original goal (as dieting is another way to lose weight in Table 4), but the relation between the problem (a broken bike) and the substitute (buying grapefruit and yogurt) could not easily be determined at a local level. Some examples of problems and the actions that resulted from them include going to McDonald's after finding a stopped clock, looking for a scarf when the power goes out, substituting a quilt for a clock, calling customers when the vegetables are overcooked, and looking in the cupboards when the car won't start. In these examples, it is not clear why the action results from the problem because the general goal is not given. For example, the general goal connecting the quilt to the clock was the search for something decorative to place above a fireplace mantel.

In addition to the test word that described the goal introduced in the introduction, each text also had two other test words, one for each continuation. About half of these were words that appeared in the continuation (positive test items), and about half were words that did not appear in any text at all (negative test items).

There were also 40 filler texts. Each was seven sentences in length, and each had two test words. Of all the filler test words, 30 were posi-

Table 4
Examples of Stories Used in Experiment 2

Part of story	Story
	Globally inconsistent
Introduction	Curtis spied a tennis court in the park.
	His arm was healing from an injury and needed a workout before the big match.
	So he needed an opponent.
	Curtis waved to a friend to join him.
Control continuation	The friend came over and was an exhausting opponent.
	Curtis decided to go borrow some change for a drink.
	Curtis ran happily along the path.
Problem continuation	Curtis' friend did not want to be Curtis's opponent.
	So Curtis decided to go home and study videotapes of his serve instead.
	Curtis ran happily along the path.
Goal test word: *Workout*	
	Locally inconsistent
Introduction	Diane wanted to lose some weight.
	She thought she should lose at least 20 pounds.
	Diane thought cycling might help her lose some weight.
	She went to the garage to find her bike.
Control continuation	Diane peddled 5 miles each day for 3 months and became very slim.
	She decided to go back to school to complete her degree.
	It took several years, but Diane finally reached her goal.
Problem continuation	Diane's bike was broken and she couldn't afford a new one.
	So she went to the grocery store to buy grapefruit and yogurt.
	It took several years, but Diane finally reached her goal.
Goal test word: *Weight*	

tive words from their texts and 50 did not appear in any text. The positive words were always chosen from the latter halves of their texts (because the goal test words from the experimental texts were always from the beginning of their texts).

Procedure. The presentation of stimuli and collection of responses were controlled by a real-time computer system. Stimuli were displayed on a CRT screen, and responses were indicated using keys on the CRT's keyboard.

The experiment began with a practice list of 30 texts, each one or two sentences in length. Then the 80 texts of the experiment proper were presented in the same manner as the practice texts.

Presentation of each text began with an instruction displayed on the CRT screen asking the subject to press the space bar. When the space bar was pressed, there was a 500-ms pause and then the first sentence of the text was displayed. The sentence remained on the screen until the subject pressed the space bar again; then there was a 50-ms pause, the screen was cleared, there was another 50-ms pause, and then the next sentence of the text was displayed. Presentation of the sentences continued in this way until the final sentence of the text. After the final sentence was displayed and the space bar pressed, a row of addition signs appeared with a test word immediately below it. The subjects' instructions were to indicate whether the test word had appeared in the immediately preceding text by pressing the "/" key for a positive response and the *z* key for a negative response. If the response was incorrect, the letters of the word *ERROR!!* were presented one at a time for 600 ms each, and then the screen was cleared and a row of addition signs and a second test word were presented. If the response to the first test word was correct, then the test word was erased from the screen, there was a 100-ms pause, and then the row of addition signs with the second test word appeared. If the response to the second test word was correct, the instruction to press the space bar to begin the next text was displayed. If the response was incorrect, the error message was presented before the instruction to begin the next text. The order of presentation of the texts was randomly chosen, a different randomization for each second subject. For the experimental texts, the first test word was always the word expressing the goal mentioned in the introduction. For the filler texts, the correct response for the first test word was always negative. Subjects were instructed to read the texts carefully and to respond as quickly and accurately as they could to the test words.

Design and subjects. For one group of 50 subjects, the experimental texts were the Globally Inconsistent set, and for a second group of 50 subjects, they were the Locally Inconsistent set. Each of the sets was divided into two subsets. The subsets were combined in a Latin square design with two sets of subjects (25 subjects per set) and the two continuations, Problem and Control.

Results

Mean response times and error rates for the test words were calculated for each subject and each test word, and mean reading times were calculated for each sentence of each text. Means of these means are shown in Table 5.

From both the minimalist and global inference points of view, the Problem continuations of the Locally Inconsistent texts should require the use of global information. The original goal is needed for the continuations to be understood. Thus, responses to the general goal test word should be faster in the Problem Condition than the Control condition, which is what

Table 5
Experiment 2: Reading Times for Sentences and Correct Response Times (in Milliseconds) and Error Rates for Test Words

Data	Control continuation	Problem continuation
Locally inconsistent texts		
Reading time		
Sentences 1–4	2,166	2,184
Sentence 5	2,000	1,973
Sentence 6	2,371	2,120
Sentence 7	1,524	1,567
Goal test word	1,086[a]	1,030[b]
Globally inconsistent texts		
Reading time		
Sentences 1–4	2,078	2,144
Sentence 5	1,951	2,019
Sentence 6	2,454	2,345
Sentence 7	1,609	1,681
Goal test word	1,137[c]	1,164[d]
Filler test word		
Positive test word	889[a]	
Negative test word	1,004[b]	

[a] Percentage error was 6 for this entry. [b] Percentage error was 4 for this entry. [c] Percentage error was 5 for this entry. [d] Percentage error was 8 for this entry.

the data show. For the Globally Inconsistent texts, the two points of view make different predictions; according to the local coherence position, there is no local problem with comprehension, and so there should be no significant difference between mean response times for the goal word in the Problem and Control continuations. According to the global coherence position, the general goal should still be involved in comprehension in the Problem Continuation, and so response times for the test word should be facilitated. As Table 5 shows, no significant facilitation was observed (the nonsignificant difference is in the wrong direction).

The results just described represent an interaction shown significant by ANOVA, $F(1, 49) = 4.68$, with subjects as the random variable, and $F(1, 38) = 7.32$, with test words as the random variable. There was also a significant main effect, that responses for the goal test words were slower for the Globally Inconsistent than the Locally Inconsistent texts, $F(1, 49) = 41.2$, with subjects as the random variable, and $F(1, 38) = 6.29$, with test words as the random variable. The standard error for the response times was 17 ms.

Post hoc tests showed the advantage for response times for the goal word with the Locally Inconsistent Problem texts to be significant, $F(1, 49) = 5.65$, with subjects as the random variable, and $F(1, 38) = 11.1$, with test words as the random variable. For the Globally Inconsistent texts, response times for the goal words were actually slower with the Problem text than the Control text, but this difference was not significant, $F(1, 49) = 1.31$ and $F(1, 38) = 2.58$.

The error rates for the goal test words were generally in accord with the response times. The interaction between Locally versus Globally Inconsistent text and continuation type was significant, with subjects as the random variable, $F(1, 49) = 4.21$, but not with test words as the random variable, $F(1, 38) = 2.37$. No other effects were significant ($Fs < 1.07$).

The reading time data is presented in Table 5 for completeness. There are two points worth noting. First, reading times for Sentence 6 are slow in all conditions, reflecting the point at which a new goal is introduced. However, reading times show less slowing for the Locally Inconsistent Problem continuations than for the other three conditions. This suggests that connecting a new goal to a previously mentioned higher order goal may be easier when the new goal is perceived to be directly related to previously mentioned goals. Second, the patterns of reading times are about the same for the two kinds of texts, Globally and Locally Inconsistent. Thus, the differences in response times for the goal test words cannot easily be ascribed to differences in reading times.

Discussion

Experiments 1 and 2 offer three tests of the notion that causal global inferences are encoded during reading. In both the Try Again and the Substitution conditions of Experiment 1, the general (global) goal should have been tied into comprehension at the ends of the stories. The same is true for the Globally Inconsistent Problem continuations of Experiment 2. However, in none of the three cases was there evidence that the general goal was more available after these continuations than after the Control continuations. Instead, the results support the hypothesis that global information is not automatically used during local comprehension.

Experiments 1 and 2 also offer two tests of the idea that the availability of concepts depends on whether they are required to establish local coherence. In both the Try Again continuations of Experiment 1 and the Locally Inconsistent Problem continuations of Experiment 2, concepts that were required for local coherence showed facilitation relative to the Control condition.

Experiments 1 and 2 used an on-line testing procedure. It is often argued that there are several possible interpretations of results obtained with this procedure (cf. McKoon & Ratcliff, 1980a; McKoon & Ratcliff, 1986; McKoon & Ratcliff, 1989a; Potts, Keenan, & Golding, 1988; Ratcliff & McKoon, 1988). Up to this point, we have assumed that a response to a test word reflects the state of availability of the concept tested, that is, the state of availability at the end of the text that precedes the test word. However, another possibility is that the response reflects a backwards context-checking process by which the test word is matched against the preceding text to determine if it fits the context (Forster, 1981). A poor match could inhibit the response, and a good match could facilitate it. Still another possibility is that the preceding text and the test word are jointly matched against memory as a compound cue (Ratcliff & McKoon, 1988); again, a good match would facilitate the response and a poor match would inhibit it. Fortunately, the data for Experiment 1 provide the means to decide among the interpretations. Both the backwards context checking and the joint matching interpretations lead to the same prediction: Response

times for the general goal test words should be facilitated in the Try Again and Substitution conditions relative to the Control condition. This is because the texts in the former two conditions are still discussing information relevant to the general goal, whereas the Control is not (in the Control continuation, a new general goal has been introduced). For example, the test word *clean* should have provided a good context-checking match when the Try Again and Substitution continuations discussed water or brooms, but not when the Control continuation discussed painting a barn. However, this prediction does not fit the data; there were no significant differences in response times across conditions for the general goal test word. By this reasoning, we interpret the results of Experiments 1 and 2 as reflecting inference processes that occur during reading. The processes of backwards context checking and jointly matching text and test word against memory may also have been part of the processing of the test word, but they were not responsible for differential response times and accuracy rates across experimental conditions. However, it should be stressed that this is not a general conclusion about the on-line processing of test words. In other experiments, backwards context checking or a joint matching process might be responsible for on-line testing results.

The results of Experiments 1 and 2 support the local coherence, minimalist hypothesis over global inference theories. A recent experiment by Suh and Trabasso (1988) also can be interpreted to support the minimalist hypothesis (although Suh and Trabasso interpreted their results differently). They tested for the use of global information during reading of texts like that in Table 1 and found increased availability of global information at places that might have corresponded to coherence breaks, that is, points at which local coherence may not have been possible without the use of global information.

Despite the support for the minimalist hypothesis in Experiments 1 and 2, it could be argued that the texts in all these experiments were short and unnaturalistic. Also, only one experimental methodology was used, testing single word recognition immediately after reading. In Experiments 3 and 4, longer and more natural texts were used. The procedure in Experiments 3 and 4 was one that would allow examination of possible global inferences in the memory representations of the stories.

Experiment 3

The stories for Experiment 3 were 600-word narratives of the sort that might describe a television adventure story (see Tables 6 and 7). They were written to express a series of goals for a main character, with each goal eventually being fulfilled through some outcome. The goals were embedded such that fulfillment of any goal required that all of its subordinate goals had to be fulfilled first. For example, in the *Kidnapped* story, Jon had to help Ali with the microfilm to get into the fortress, and he had to get into the fortress to find his daughter, and so on. Once the most subordinate goal was fulfilled (e.g., Jon gets the microfilm), then the other higher goals could each be fulfilled in turn. If global causal inferences are constructed during reading, then each goal should be connected to its eventual outcome by inferred relations. This should be true even though the goal and the outcome events are far from each other in the text. However, if only local relations are constructed, then the goals will not be connected directly to their outcomes.

Whether the goals of the stories were connected to their outcomes in the encoded representations of the stories was tested with a priming procedure. Subjects read two stories and then were presented with a list of test statements for verification. For each story, there were statements that tested goals and statements that tested their outcomes. Theories that assume the encoding of global causal relations during reading would predict that a goal was connected to its outcome during reading and therefore that the connection would be encoded into the memory representation of the story. It follows, then, that a test statement about the goal should facilitate responses to an immediately following test statement about the outcome. This should be true even when several paragraphs intervene between the statements in the text. The facilitation given to the outcome statement by the goal statement should be greater than any facilitation that might be given by some other statement that was equally far away in the text.

Method

Materials. Twelve stories were written, each with a series of embedded goals. An example story is shown in Table 6, and the structure of the goals used in the experiment is shown in Table 7. (Table 7 does not represent the complete goal structure for all the goals for all the characters, only those goals relevant to the test conditions used in the experiment.) The stories were written so that each subgoal had to be fulfilled before the next highest subgoal could be attempted. So, for example, Jon had to find his daughter before he could attempt to rescue her. For each story, there was a series of true–false test sentences. One of these, an outcome target, expressed the outcome of one of the goals; for example, "Ali drove with Jon hidden in the trunk" expressed the means by which Jon achieved the goal of entering the fortress. A second, goal prime, test sentence expressed the goal ("Jon had to find help to get into the fortress"). A third test sentence (action near goal prime), a control condition, expressed some action that was near to the goal in terms of number of words in the story but not directly related to the goal; and a fourth sentence (near prime), another control condition, expressed an action that was near to the outcome in terms of number of words. Four more test sentences represented the same four conditions (goal, outcome, and two controls) with a different goal of the story. Finally, there were eight other sentences used as fillers in the test lists, three true sentences and five false sentences. The stories ranged from 579 to 613 words in length and from 53 to 59 lines when presented on a CRT screen. Each story was divided into seven paragraphs. The test sentences that represented the experimental conditions ranged from 7 to 11 words in length. Test sentences were taken as exactly verbatim from the stories as possible, allowing for shortening and using names or descriptions instead of anaphors.

There were also 12 other stories that were part of another experiment (Ratcliff & McKoon, 1988). These were about the same length, and each of them had seven true and five false test sentences that were used in the test lists.

Procedure. The experiment was conducted with a CRT screen and keyboard as in Experiments 1 and 2. The experiment began with a practice list of 40 strings of letters presented for lexical decision to give subjects practice at responding quickly and accurately with the keys on the CRT keyboard. After the lexical decision, there was 1 study-test list for practice and then 12 study-test lists for the experiment proper.

Each study-test list began with an instruction to press the space bar

Table 6
An Example Story From Experiment 3: Kidnapped

Jon was a CIA agent who often worked behind the Iron Curtain. He had made many enemies, and one of them, a KGB agent, kidnapped his daughter, Karyn, while she was on a trip to the Bahamas. It was all part of a plan to get revenge because Jon had foiled one of the enemy agent's plots many years before.

Jon wanted to get Karyn back from the KGB agent who had kidnapped her as quickly as possible. He had worked against the KGB agent, Vladimir, many years ago and was very worried about his daughter's safety. Although the authorities told Jon that he should stay at home and let the professionals do their job, Jon decided that he had to get to the Bahamas. Anxious, as any father would be, he made a reservation on the first plane he could find. In a few hours Jon arrived in the Bahamas.

Jon believed the only way he would get Karyn back safely was to find her himself. He had to find out where the kidnapper had taken her. Soon after Jon checked into a hotel, a young man delivered a ransom note from his enemy, Vladimir. As soon as the messenger left, Jon quietly followed him. He hoped that the young man would lead him to Vladimir. After some time, Jon arrived at a large, old fortress that was once used as a prison. As Jon watched the messenger go into the fortress, he was sure this was where his daughter was being held by Vladimir. He hoped that she was alright.

The fortress appeared to be completely impenetrable. Jon knew that if he was to rescue Karyn he would have to find help getting into it. Jon returned back to town, hoping to find a mercenary to help him. After visiting several bars, Jon met an old friend, Ali Al-Dib, a double agent he had known for many years. They had worked both against and with each other but they always remained friends. Jon and Ali had some beers and talked over old times. Jon discovered that Ali had done business with Vladimir on several occasions. Jon explained his situation to Ali and asked him to help rescue Karyn.

Ali was busy with his own mission, stealing some microfilm that contained the locations of missile silos of certain west European countries. He hoped to sell it to the highest bidder. Ali agreed to help Jon if Jon would help him first. Jon thought it was a fair exchange and agreed to the bargain. They sat up late that night trying to come up with a plan to get the microfilm, which was hidden in the British embassy. They came up with a deceptively simple plan. Since Jon knew some people at the embassy, he would go in first and keep them occupied while Ali stole the microfilm. It worked.

Ali contacted Vladimir and asked him if he would be interested in buying the microfilm. Vladimir wanted to see it first, so Ali drove to the fortress with Jon hidden in the trunk. The guards recognized Ali and let him into the fortress without searching his car, so they did not find Jon in the trunk. While Ali kept Vladimir busy examining the microfilm, Jon ran from room to room and finally found the room where his daughter was being held hostage.

They escaped, undetected, and hid in Ali's car. Soon, Ali finished his business with Vladimir and got into the car. He drove Jon and Karyn to the airport before Vladimir realized Karyn had been rescued. In just a few hours, Jon and Karyn were safely back home.

Test sentences
Outcome target: *Ali drove with Jon hidden in the trunk.*
Goal prime: *Jon had to find help to get into the fortress.*
Action near goal prime: *Jon met an old friend who was a double agent.*
Near prime: *Jon kept the people at the embassy occupied.*

on the CRT keyboard. When the space bar was pressed, there was a 500-ms pause, and then the first paragraph of the first story was displayed. The paragraph remained on the screen until the subject pressed the space bar again; then the screen was erased, and after a 100-ms pause, the next paragraph was presented. Presentation continued in this way through all the paragraphs of the story. After the last paragraph, there was a 3-s pause, and then the second story was presented in the same way. After a 3-s pause after the second story, a row of asterisks was displayed for 500 ms, and then the test sentences were presented one at a time. Each sentence remained on the screen until the subject pressed a response key ("?/" for true and *z* for false), and then the screen was erased and there was a 100-ms pause. If the response was correct, the next sentence was presented immediately. If the response was incorrect, the letters of the word *ERROR!!* were displayed one at a time for 600 ms each. Then the screen was erased and the next sentence presented. After all 24 sentences of the test list, the instruction to press the space bar for the next study list was presented.

Table 7
Goal and Outcome Structure Kidnapped

Goal and structure
Goal 1: Rescue his daughter
Goal 2: Find his daughter himself
Goal 3: Get into the fortress
Goal 4: Help Ali with the microfilm
Outcome 4: Got the microfilm
Outcome 3: Got into the fortress
Outcome 2: Found his daughter
Outcome 1: Escaped with his daughter

The stories presented in each of the 12 lists were chosen randomly, except that there was one story from the experiment (and one from the other experiment) in each list. These two stories were presented in random order. The test sentences of a list were presented in random order (sentences from the two stories interspersed), except for two restrictions: The test sentences used in the experimental design were not presented in the first test position, and the test sentence immediately preceding a prime–target pair was not from the same story as the target. A different randomization was used for every second subject.

Subjects and design. There were two groups of subjects. For the first group (21 subjects), the outcome test sentence was primed by its goal test sentence, the test sentence near to it in the text, or a test sentence from the other story of the study list (Control). These three conditions were combined in a Latin square design with sets of subjects (7 per set)

and sets of outcome test sentences (8 per set). For the second group of subjects (32 subjects), an outcome test sentence was primed by its goal, another action near to the goal, or a test sentence from the other story (Control). Again, the three conditions were combined in a Latin square with sets of subjects and sets of outcome test sentences. The subjects participated in the experiment for credit in an introductory psychology course.

Results

Means were calculated for each subject and test sentence in each condition, and means of these means are shown in Table 8. For the target test sentences, only responses preceded by a correct response to the priming sentence are included in the means.

The minimalist prediction is that responses to the outcome targets should receive the largest amount of facilitation when the prime is the sentence near to the outcome in the text. There should also be some facilitation when the prime is a sentence farther away in the text (because the sentences are from the same text), but the amount of this facilitation should not depend on whether the prime is related to the target as goal and outcome. This is the pattern of data shown in Table 8. Relative to the Control condition (the prime from another story), responses to the outcome target are fastest with the near prime and about equally fast with the Goal and Action Near the Goal primes.

For the first group of subjects, an ANOVA showed that the overall difference in response times for the target test sentences was significant, $F(2, 40) = 19.0$, with subjects as the random variable, and $F(2, 22) = 17.6$, with test sentences as the random variable. Post hoc tests showed that response times in the near priming condition were faster than response times in the goal priming condition, $F(1, 40) = 4.6$, and $F(1, 22) = 5.3$. Standard error of the response time means was 38 ms. There were no significant differences in error rates.

For the second group of subjects, an ANOVA also showed that the overall difference in response times for the target test sentences was significant, $F(2, 62) = 18.3$, with subjects as the random variable, and $F(2, 22) = 15.2$, with test sentences as the random variable. Post hoc tests showed that the difference between the goal and action near goal priming conditions was not significant (Fs < 1.0). The standard error of the means was 30 ms. Differences in error rates were not significant (Fs < 1.3).

For the first group of subjects, the mean reading time per paragraph was 12.150 s, and for the second group of subjects, it was 12.939 s.

Table 8
Results From Experiment 3: Response Times (in Milliseconds) and Error Rates for Outcome Target Sentences and Filler Test Sentences

	Subject Group 1		Subject Group 2	
Priming condition	RT	% error	RT	% error
Outcome target sentences				
Goal prime	1,567	6	1,541	6
Action near goal prime			1,576	7
Near prime	1,451	7		
Control prime	1,781	11	1,772	10
Filler test sentences				
True items	1,605	8	1,628	10
False items	1,801	15	1,823	26

Note. RT = response time.

Discussion

If global inferences connected goals to outcomes in the stories of Experiment 3, then the outcome test statements should have been primed more by the goal test statements than by the action near goal test statements. However, the two priming effects were not significantly different. Once again, as with Experiments 1 and 2, the data failed to provide evidence of global inferences.

One problem that might be raised with Experiment 3 is that the data show no evidence of any kind of structure for the stories at all. Responses to the target statements were facilitated more by other statements from the same story than by statements from a different story, but within a story the only effect was one of surface distance, with the near primes giving more facilitation than the other within-story primes. However, as discussed earlier, previous investigations of mental representations of texts have demonstrated some internal structure, specifically, that propositions sharing arguments are connected together (McKoon, 1977; McKoon & Ratcliff, 1980b; Ratcliff & McKoon, 1978). In Experiment 4, we looked for evidence of this kind of structure.

Experiment 4

If propositions from the stories of Experiment 3 are connected by argument repetition during reading, then evidence of those connections should be observable in priming effects. For example, all the propositions about Ali should be connected together, whether he was explicitly called Ali or referred to as "double agent" (cf. McKoon & Ratcliff, 1980b). These propositions should be more closely connected to each other than they are to other propositions that do not refer directly to Ali. We tested for these differences in connections with the same procedure as in Experiment 3, priming in verification of statements from the stories.

Method

Materials. The 12 stories from Experiment 3 were used, with a new set of test sentences. For each story, there were two target test sentences. Each of these targets had two primes. One of the primes was near the target in terms of the argument repetition structure of the story, and the other was relatively far from the target. The average distance of the two primes from the target in terms of number of words was about the same (191 words and 192 words, respectively). For example, in the *Kidnapped* story, one target was "Jon met Ali, who was an old friend." The near prime for this target was "Karyn's father took the first plane to the Bahamas," which shares an argument with the target because Jon and Karyn's father are the same person. The far prime for this target was "Vladimir wanted to see the microfilm before he

bought it," not so closely connected to the target by argument repetition. The number of words in the prime and target test sentences ranged from 7 to 11. There were also 8 filler test sentences for each story, 3 true sentences and 5 false sentences.

Procedure. The procedure was the same as for Experiment 3, except there were no stories from another experiment so that the total number of study-test lists was six.

Design and subjects. Each target was primed by another test sentence near it in argument repetition structure, another test sentence far from it in argument repetition structure, or a sentence from the other story in the study list (Control). These three conditions were combined in a Latin square with the 12 stories (4 per set) and 24 subjects. The subjects participated for credit in an introductory psychology course.

Results

The data were analyzed as in Experiment 3, and the results are shown in Table 9.

As expected, response times for the targets were speeded with the near prime, relative to both the prime from the other story and the far prime from the same story. An ANOVA showed that overall differences were significant, $F(2, 46) = 8.5$, with subjects as the random variable, and $F(2, 22) = 5.9$, with test sentences as the random variable. The difference between the near and far conditions was significant by post hoc test, $F(1, 46) = 4.6$ and $F(1, 22) = 6.7$. The standard error of the means was 26 ms. There were no significant differences in error rates ($Fs < 1.9$). The mean reading time for all paragraphs was 17.260 s.

Discussion

The motivation for Experiment 4 lay in a potential problem with interpretation of the results of Experiment 3. We want to claim that, for the stories of Experiment 3, readers encoded the same local relations as have been demonstrated in past experiments. The inferences that they failed to encode were the global ones for which we tested. However, Experiment 3 gave no evidence that readers had, in fact, encoded any relations at all other than proximity in surface distance. Experiment 4 provided this evidence, showing that relations based on argument repetition were represented in memory. Thus, the mental representation does show structure, but the structure is based on argument repetition and not on global inferences about causality.

Table 9
Results From Experiment 4: Response times (in Milliseconds) and Error Rates for Target Test Sentences and Filler Test Sentences

Priming condition	RT	% error
Target test sentences		
Near prime	1,502	5
Far prime	1,579	5
Control prime	1,651	10
Filler test sentences		
True items	1,586	10
False items	1,661	21

According to the minimalist hypothesis, the inferences that build the argument repetition structure are based on information that is easily available, in this case, the names and descriptions of the characters in the stories. For example, Jon, the CIA agent, is the main character in the story in Table 6. Whenever Jon is mentioned in the story, and new propositions are to be attributed to him, his name serves to make available other information encoded about him earlier in the story and to make it likely that these different pieces of information will be connected through repetition of their argument *Jon*. So long as a definite description of an entity is a strong enough cue to evoke previous information about the entity, then the different pieces of information can be connected together.

An argument that might be advanced against the minimalist interpretation of the results of Experiments 3 and 4 is that a recognition test procedure does not tap the level of representation at which inferences are encoded, but instead some more superficial level of representation. However, this argument is countered by previous research. First, discussed later, recognition does give evidence for some kinds of elaborative inferences (those supported by well-known, easily available information). Second, recognition also gives evidence for structural inferences when the minimalist hypothesis predicts that such inferences should be encoded. McKoon and Ratcliff (1980b) used recognition to show that the organization of a list of sentences was inferred from well-known (schema) knowledge. Similarly, McKoon, Ratcliff, and Seifert (1989) used recognition to show that the relations between stories were inferred from schema knowledge. Recognition can also be used to show that both structural and elaborative inferences are constructed when subjects are given instructions to use special strategies during reading (Seifert, McKoon, Abelson, & Ratcliff, 1986; M. McDaniel, November, 1991, personal communication).

In sum, Experiments 1 through 4 strongly support the minimalist hypothesis over the constructionist hypothesis. With both simplistic and natural texts and with both on-line and delayed memory procedures, there was no evidence that causal global inferences were constructed. Evidence for global inferences appeared only for texts that were not locally coherent. These results emphasize a striking contrast between local and global inferences. Local inferences have been easy to demonstrate empirically in a large number of studies. However, in the same kinds of experiments in the same laboratory situations, there is no evidence for the kinds of causal global inferences posited by a number of theorists.

It is important to recognize that the results of Experiments 1 through 4 demonstrate failures to encode global—not local—causal inferences. The minimalist claim is that local causal inferences will be encoded either if they are easily available from long-term memory or if they are required to establish local coherence.

Van den Broek (1990) and Fletcher and Bloom (1988) have proposed a model by which the causal inferences necessary for local coherence are encoded. The architecture and processes of the model are the same as in van Dijk and Kintsch's (1983) model, except that the propositions of a text are connected by causal relations in addition to argument-repetition relations. The model has the same short-term memory limit on processing as the minimalist position: Only propositions that are in

short-term memory at the same time are connected by inferences; information from other parts of the text is used only if the local information is not coherent. Van den Broek provided a definition for *coherence* in terms of four criteria of causality. Coherence is maintained for an event if there are antecedents for the event that are temporally prior, operating at the time of the event, necessary for the event to occur, and sufficient for the event to occur. The event is connected to antecedents that fulfill these criteria just as propositions containing the same argument are connected in the Kintsch and van Dijk model. Only if there is no antecedent fulfilling all the criteria does a coherence break occur (van den Broek, 1990, p. 434). Then, either propositions of the text that are no longer in short-term memory are retrieved, or new propositions are generated to provide the connections necessary for coherence. Evidence consistent with this model has been provided by Bloom et al. (1990) and by Fletcher and Bloom (1988).

The results of Experiments 1 through 4 show that the global causal inferences defined by recent theories are not part of automatic encoding processes. However, the results say nothing about their roles in other more goal-driven encoding processes or in retrieval processes. Although the focus of this article is on inferences that are constructed automatically during reading, it must be stressed that understanding the processes that construct inferences important to a reader's goals and the processes underlying recall are also extremely important. Practically speaking, we use goal-driven reading processes and recall processes ubiquitously, and setting up optimal reading and recall processes is the aim of many educational efforts. The problem raised by the results presented in this article is to accommodate a minimalist representation of textual information with the more constructionist information that appears in recall and question answering and that readers use in those frequently occurring situations where they have specific goals. One possibility is that information beyond the minimal is constructed by retrieval processes that follow local connections through memory. A model like this, based on Raaijmakers and Shiffrin's (1981) recall process, has been developed by Fletcher and van den Broek (1989), with some empirical support. In general, however, there is little current theorizing about the more strategic aspects of text processing.

Elaborative Inferences

The most important claim of many mental models theories of text comprehension is that the mental representation of a text automatically depicts the events described by the text in a lifelike way. Various parts of the description must be constructed by elaborative inferences, because a text seldom provides an explicit description of an event that is sufficiently complete to describe the situation in a lifelike way. Thus, it is essential to mental models theories to show that elaborative inferences are automatically encoded during reading.

In contrast, the minimalist hypothesis does not make any claim about the extent to which a mental representation depicts the event described by a text. Instead, the minimalist hypothesis applies other criteria to decide whether inferences will be constructed: whether the text is locally coherent and whether the information necessary for an inference is easily available. Usually, these criteria are not consistent with a full description of a textual event. This is because the information necessary for a complete description is usually not all easily available. Also, for local information, a coherent description is not necessarily a complete description.

The minimalist criteria for elaborative inference processes are advantageous in that they provide guides to empirical research. Specifically, demonstrations that a criterion for elaborative inference is met (e.g., a demonstration that inference-supporting information is quickly available; McKoon & Ratcliff, 1989b, Experiment 2) can be separated from demonstrations that elaborative inferences are encoded, thus avoiding circularity. The criterion provided for elaborative inferences by a constructionist hypothesis does not so obviously lead to independence: There is no a priori way to know, for any particular inference, whether it is required in the representation of an event. As a result, there is no way to independently verify whether a particular inference should be encoded.

In the sections that follow, specific kinds of elaborative inferences are considered. Each case allows evaluation of the minimalist hypothesis, the constructionist hypothesis, or both. According to the minimalist hypothesis, for each kind of inference, encoding should depend on the availability of the information necessary to support inference processes. If supporting information is not quickly available, then an inference should not be constructed (unless necessary for local coherence). According to the constructionist hypothesis, the encoding of inferences should not depend completely on the availability of supporting information; instead encoding should depend on whether an inference is required for a lifelike description of the event described by the text.

Consideration is limited to those kinds of elaborative inferences for which there is sufficient research to provide a reasonably coherent body of data. These are instrumental inferences, inferences about the meanings of words, and predictive inferences about what will happen next in a story. For other elaborative inferences, such as expectations (Duffy, 1986), and inferences deriving from the argument structures of verbs (Boland, Tanenhaus, & Garnsey, 1990; McKoon & Ratcliff, 1989c; Tanenhaus, Carlson, & Trueswell, 1989) the accumulated data are not sufficiently constraining to test the minimalist and constructionist hypotheses.

Instrumental Inferences

When elaborative inferences were first studied extensively, in the 1970s, it was argued that a description of the event described by "Mary stirred her coffee" (Dosher & Corbett, 1982) should include the instrument *spoon* (cf. Johnson et al., 1973; Paris & Lindauer, 1976). Early evidence to support the encoding of instrumental inferences came from cued-recall studies, in which recall of a text was facilitated by a cue that was an instrument highly associated with a verb in the text but not stated explicitly in the text (Paris & Lindauer, 1976). Subsequently, Singer (1978, 1979) and Corbett and Dosher (1978) showed that cued-recall results could not decide issues of encoding.

More recent research argues against the constructionist hypothesis. Dosher and Corbett (1982) looked at the relation be-

tween an inference sentence and its implicit instrument, for example, "Mary stirred her coffee" and *spoon*. They examined whether the relation would affect responses to the instrument when it was presented as a test item in a Stroop task. Results showed that Stroop responses were not affected. There was no effect regardless of whether the instruments were the most likely for their sentences, and there was no effect for instruments that were tools or for instruments that were body parts. Only when subjects were instructed to explicitly guess the instrument in advance of the Stroop test were responses affected. In other words, unless an instrument was explicitly requested, there was no evidence that it was involved in comprehension of the inference sentence. These results argue strongly against the constructionist hypothesis because a complete description of an event like stirring coffee seems to require an instrument.

On the other hand, the results are compatible with the minimalist hypothesis, given the assumption that the instruments were not automatically available during reading of the inference sentence. The assumption can be tested as follows: If the availability of the instruments is increased to a sufficiently high level, then they should be encoded. This test was part of a study by McKoon and Ratcliff (1981). Availability was increased by explicitly mentioning an instrument several sentences before the inference sentence for which it would be the implicit (but highly typical) instrument; for example, spoon would be mentioned several sentences before the sentence "Mary stirred her coffee." The instrument was presented as a test word immediately after the inference sentence. Responses to the test word were facilitated (relative to a control condition), suggesting that the relation between sentence and instrument was available in an immediate test situation. This availability should lead to encoding according to the minimalist hypothesis. That the instrument was encoded was confirmed by a priming effect in a delayed memory test. Presenting the instrument as a test word immediately before a noun from the inference sentence (e.g., *spoon* immediately before *coffee*) facilitated responses to the noun (relative to a control condition). The facilitation indicates a close association between the instrument and the noun in the memory representation of the sentence, which in turn indicates that the instrument was encoded with the sentence.

Overall, empirical results from studies of instrumental inferences favor the minimalist hypothesis. Highly typical instruments of verbs are strong candidates for inclusion in a mental model of a stereotypical event such as stirring coffee, yet there is no evidence that they are used in comprehension or that they are encoded (unless subjects engage in special strategies; Dosher & Corbett, 1982). In contrast, the minimalist hypothesis predicts the finding that increasing the availability of the instruments during comprehension leads to their encoding.

Inferences About the Meanings of Words

Instrumental inferences were one of the kinds of elaborative inferences studied in the 1970s in the effort to document constructed mental representations. Another kind were inferences about the meanings of words. For example, R. C. Anderson and Ortony (1975) used cued recall to examine the meaning of *container* in the sentence "The container held the apples." The cue for the sentence was either *basket* or *bottle*, and *basket* was more effective.

Results like this suggest that contextually appropriate aspects of the meanings of words might be encoded into the mental representations of texts, and current research confirms this idea. McKoon and Ratcliff (1988; see also Barsalou, 1982; Tabossi, 1982; Tabossi & Johnson-Laird, 1980) used texts in which a specific feature of a noun was made salient (e.g., a text about painting a picture of a tomato should make salient the color of the tomato, red). After a series of texts, test sentences were presented for verification. Sentences that tested a feature that had been made salient in the text (e.g., "tomatoes are red") were verified faster than control sentences.

If features of the meanings of words are automatically encoded into memory, then according to the minimalist hypothesis they must have been easily available during comprehension. Easy availability should show up when the features are tested immediately after reading. Immediate facilitation of this sort was obtained by McKoon and Ratcliff (1988; see also Tabossi, 1982; Tabossi & Johnson-Laird, 1980), using a sentence verification task, and by Greenspan (1986) using lexical decision.

Research that has examined the contextually defined meanings of category terms is also consistent with the minimalist hypothesis. If a text mentions the category *animals* in the context of milking some animals on a farm, subjects have difficulty in later rejecting the word *cow* as having appeared in the text (McKoon & Ratcliff, 1989b). This result can be taken to indicate that something like the concept *cow* was encoded into the mental representation of the text. It should follow that the concept *cow* is easily available during comprehension. This availability appears as facilitation when *cow* is tested immediately after the text, with both recognition and lexical decision (McKoon & Ratcliff, 1989b), and also as faster reading time for a follow-up sentence that explicitly mentions *cow* (Roth & Shoben, 1983).

These patterns of results are consistent with both the minimalist and constructionist views, but the minimalist view is the more constrained. The hypothesis that encoded inferences must be based on immediately available information would be contradicted if some inference was encoded, but its supporting information was not quickly available during reading (and subjects did not engage in special strategies). However, there are no inferences that pattern this way. In contrast, the constructionist view makes no claims about the relation between availability during reading and subsequent encoding, and so no constraints are placed on the constructionist hypothesis.

Predictable Events

If someone falls off a 4-story roof, then the real-life result will be death. Because the outcome is so predictable, a mental model for a text such as "the actress fell from the fourteenth story" should automatically include the inference that she died. It would not be reasonable, from the mental model point of view, to leave her suspended in midair. On the other hand, the inference about death is not necessary for local coherence if the text ends with the sentence about the fall. The event of falling from a 14-story building is not familiar enough to make the inference easily available. So the minimalist hypothesis pre-

dicts that the inference about death will not be included automatically in the mental representation.

To test for inferences of this kind, McKoon and Ratcliff (1986, 1989d, 1989e) used a speeded recognition memory test. Subjects read several short texts before reading a list of test words. Each test word was followed by a signal, and the subjects were instructed to give a response immediately when the signal was presented. The delay between test word and signal was short enough that slow, strategic processes (that might construct inferences at the time of the test) were eliminated. The critical test words were those that represented inferences about predictable events where the events were known to be highly predictable from previous norming studies. For the actress text, the critical test word was *dead*. The correct response for these test words was *no*, because they had not been explicitly stated in any of the studied texts. However, if the inference was generated during reading, then a negative response should be difficult and subjects should tend to make errors (relative to a control condition, in which the subjects read a text that did not predict the critical event).

When a critical word was presented for test, it was preceded by a priming word (displayed for 200 ms). In one condition, the priming word was the neutral word *ready*. In this condition, subjects did not make significantly more errors when they had read the text that predicted the critical word than when they had read the control text. This result indicates that the predictable event was not clearly and explicitly encoded during reading of the predicting text, counter to the constructionist hypothesis. However, in a second condition, the prime for the critical test word was a word from the text (e.g., the word *actress*). In this condition, subjects did make more errors when they had studied the predicting text relative to the control text.

McKoon and Ratcliff (1986, 1989a, 1989d, 1989e, 1990; Potts et al., 1988) interpreted this increase in errors with the prime from the text as evidence for partial encoding of the inferences. Although the failure to find an elevated error rate with the neutral prime indicates that the inference could not have been explicitly and completely encoded, the increase in error rate with the prime from the text indicates that the inference was encoded to some degree. On the basis of this result, McKoon and Ratcliff suggested that inferences were encoded to varying degrees, with some inferences encoded minimally by a set of features or propositions that do not completely instantiate the inference. This proposal is supported by findings that inferences are encoded to a higher degree if they are based on well-known information, such as semantic associations or category membership (McKoon & Ratcliff, 1989b, 1989d).

If inferences about predictable events are not explicitly encoded, then according to the minimalist position, the reason should be that they are not quickly and easily available during reading. Several experiments have shown that this is the case (see McKoon & Ratcliff, 1989e). When the textual information that would generate the inference is immediately followed by a test for the inference, then responses on the test are not affected (McKoon & Ratcliff, 1989d; Till, Mross, & Kintsch, 1988). For example, in the sentence, "The diver jumped, spun, and hit the cement," the information necessary to know that he was hurt is given only by the final word of the sentence. When the test word *hurt* was presented immediately after the final word of the sentence, subjects had no difficulty in deciding that it had not appeared in the text (relative to a control condition). However, when more time (and/or other textual material) intervenes between text and test, then the inference does affect responses (McKoon & Ratcliff, 1986; Potts et al., 1988; Till et al., 1988). In contrast, when well-known information from general knowledge is available to support an inference about a predictable event (e.g., the predictable event of sitting after approaching a chair), then the inference does affect responses to a test word, and it does so even when the test word is presented immediately after the textual information that would generate the inference (McKoon & Ratcliff, 1989d). This contrast, between those predictable event inferences that are supported by well-known information and those that are not, is exactly in accord with the minimalist hypothesis.

The results on predictable inferences, like the results on instrumental inferences, disconfirm the constructionist hypothesis: Inferences that should be explicitly represented in a mental model, like the death of the actress, are not. However, the results on predictable inferences are also not consistent with an all-or-none minimalist position. Instead, the results suggest that inferences vary in the degree to which they are encoded. This suggestion is taken up in the General Discussion section.

Inferences From Situation Models

A number of different theories embrace the constructionist hypothesis. Theories proposed to explain story understanding hypothesize that readers construct connections between different parts of a story, such as goals and outcomes (Mandler, 1978; Mandler & Johnson, 1977; Rumelhart, 1975, 1977; Stein & Glenn, 1979; Trabasso & van den Broek, 1985). Theories proposed to explain the understanding of descriptions of events assume that the mental representation is "filled out" with inferred information (Bower, Black, & Turner, 1979; Glenberg et al., 1987; Johnson-Laird, 1980; Morrow et al., 1989; van Dijk & Kintsch, 1983). The hypothesis that unifies these theories is that the mental representation of a text automatically specifies, in some complete way, the real-life situation described by the text. The mental representations are labeled *mental models* or *situation models*.

These terms, *situation model* or *mental model*, do not in principle have to incorporate elaborative inferences beyond those postulated by the minimalist position. For example, a situation model might be proposed that contains only those elaborative inferences that are easily available from general knowledge. These inferences might connect propositions of the text in ways that simple argument repetition would not, relating the propositions to reflect well-known knowledge. Such a model has been proposed by Kintsch (1988), and this model is discussed further in the General Discussion section. In this section, evidence pertinent to constructionist situation models is reviewed.

The constructionist hypothesis is that readers automatically construct a full representation of the real-life situation described by a text. This hypothesis has been tested directly in a number of experiments. These experiments differ from those that investigate elaborative inferences in that they use a situation rather than a text as their starting point. For the experiments on elaborative inferences, the issue of concern is the

relation between a given text and the encoding of some specific inference. The issue of concern for experiments on situation-based inferences is the relation between the mental representation of a text and a real-life situation (or a lifelike situation learned in an experiment).

Many of the early experimental results thought to demonstrate the use of lifelike situation models during reading have since been reinterpreted. An excellent review of this work has been provided by Alba and Hasher (1983), and only several main points are repeated here. One kind of experiment used passages that were extremely difficult to understand and recall unless prior knowledge of the situation was invoked (e.g., the "washing clothes" passages used by Bransford & Johnson, 1972; see also Dooling & Lachman, 1971). It was originally claimed that making available prior knowledge of the situation led to the use of a full mental model during reading; however, prior knowledge may have simply provided a specific context for the interpretations of individual words and the construction of locally coherent structures, in accord with a minimalist approach (see also Alba et al., 1981). A second kind of experiment used short sentences that could be combined to describe an event (e.g., ants eating jelly on a kitchen table; Bransford & Franks, 1971). In a recognition test, subjects' confidence that they had studied a sentence describing the whole event was greater than their confidence that they had studied shorter sentences describing parts of the event, even though they had never studied the sentence describing the whole. However, it was later shown that this result could be obtained with meaningless material, such as nonsense syllables, suggesting that subjects had actively engaged in special encoding strategies (cf. J. R. Anderson & Bower, 1973; Flagg, 1976; Flagg & Reynolds, 1977; Katz & Gruenewald, 1974; Moeser, 1976; Reitman & Bower, 1973). Third, several experiments were thought to demonstrate the use of knowledge about prototypical situations (schemas) during reading (Bower et al., 1979; Graesser, 1981); for example, subjects were more likely to recognize a highly typical schema action as previously studied than a less typical action. However, Alba and Hasher pointed out that the effect can be explained as a response bias. Also, consistent with a minimalist position, it was later shown that the use of schema knowledge depends on how available it is during comprehension. Only when schema relations are extremely well-known are they automatically used to relate events during reading (McKoon et al., 1989; Seifer et al., 1986).

More recently, the lifelike situation models that have been tested empirically derive from theories proposed by van Dijk and Kintsch (1983) and by Morrow et al. (1987). The characteristics of these theories have been listed by Glenberg et al. (1987):

> A situation model is the result of interactions between information given in a text and knowledge about linguistics, pragmatics, and the real world; a situation model can be modified as new information comes in to produce a completely new interpretation of the text; the information in a situation model can be manipulated to produce emergent relations; a situation model is perceptual-like; a situation model guides interpretation of referential terms; and a situation model guides the generation of inferences (p. 69).

Since the 1970s and the realization that cued-recall experiments could not distinguish inferences generated at recall from inferences generated at encoding, there have been surprisingly few studies designed to investigate interactions between textual information and knowledge of lifelike situations. Some experiments (e.g., Johnson-Laird, 1980; Mani & Johnson-Laird, 1982; Perrig & Kintsch, 1985) used descriptions of situations (such as a textual description of the layout of a town; Perrig & Kintsch, 1985) but used procedures that invite subjects to engage in strategic processing (by extended study or a problem-solving type of task). Others have confounded learning instructions with situational versus other kinds of information (Schmalhofer & Glavanov, 1986). Only a few experiments have used procedures where exposure to a text is limited to one reading at an approximately normal reading rate.

Several of these experiments have been conducted by Morrow and colleagues (Morrow et al., 1987, 1989). They investigated whether knowledge of lifelike situations affects comprehension of narratives. With a map, subjects were taught about the rooms in a laboratory and the objects in those rooms. After the subjects memorized this information, they were presented with a series of narratives, each describing a character moving through the rooms. The subjects were interrupted at various points during the narratives with questions about whether two objects were located in the same room. Results showed that subjects were faster to answer the questions when the objects were located in a room that was relevant to the character's current location or the character's goal location.

There are two problems with taking these results as strong evidence against the minimalist hypothesis. First, subjects knew that they would be tested on the objects in the rooms as they read the narratives (all test questions were about pairs of objects). Subjects could plausibly adopt a strategy to perform well on the test questions (i.e., up to the level that would be expected of Stanford undergraduates); the strategy would be to rehearse objects while reading the narratives. At any point during reading, the probability of rehearsing the objects from a particular room could well depend on the room's relevance to the information being read at that point. If so, then the objects would be made available, not as the result of automatic (priming) processes but as the result of strategic retrieval processes by which a relevant room would be used as a retrieval cue for rehearsal of its objects. By this account, Greenspan et al.'s (1987) results are not due to the reader moving (metaphorically) through a situation model, complete with objects in their correct rooms, but instead to the reader's appreciation of the relative saliencies of concepts in local parts of the discourse and the use of the most salient concepts as retrieval cues. This account is consistent with the minimalist hypothesis, and support for it has recently been provided by Wilson, Rinck, McNamara, Bower, and Morrow (1992).

The second problem with these studies is that situation models do not predict which parts of a situation will be relevant to different narratives. Morrow et al. (1989) stated that, for the sentence "We flew from Paris to New York last week" (p. 300), comprehension is unlikely to involve information about the Atlantic ocean. This may be true, but then it becomes unclear why comprehension of a sentence about a character going from a conference room to a laboratory should make available (unstated) information about the shelves in a library that the character passes through on his way. In fact, one might argue that

flying over the Atlantic ocean makes that ocean (and its perils) more salient than the shelves on the wall of a room that is quickly left behind. The problem is that there is no way of predicting what aspects of a situation are salient in any given situation and therefore no way of predicting which inferences should be included in the mental model.

This problem is critical for a constructionist situation model approach to discourse comprehension. If empirical data is to support the inclusion of lifelike information into a mental representation of discourse, then there must be clear, theoretically motivated distinctions between inferences that should be included in the situation model and inferences that should not be.

Experiment 5

One way to begin to define constructionist inferences is to consider the real-life situation that a text describes and assume that whatever information is in the real situation is also in the mental model. This is the approach taken by Glenberg et al. (1987). Subjects read short narratives like "A girl was enjoying the warm spring weather. She walked up to the entrance of a park, and bent down to pick up a flower for her sister. Then she walked into the park and down to a small stream where some ducks were feeding. She smiled to see seven tiny ducklings trailing behind their mother." If readers construct a situation model while reading this text, then at the end of the text, their model should include the girl, and the girl should have the flower with her, exactly as would be the case in real life. This model should be different from the model constructed for a second, control, version of the text. The control version was the same as the first version except that the girl bent down to smell the flower, she did not pick it or take it with her. In the model at the end of the control version, the girl would not have the flower with her. To test for the use of a situation model during reading, Glenberg et al. presented a recognition test word at the end of a text; for this example, the test word was *flower*. Glenberg et al. predicted correctly that responses to the test word would be facilitated when the girl had picked the flower to take with her compared with when she had only smelled it.

This result appears, at first, to provide elegant support for the notion that a situation model is used during comprehension. However, there is an alternative interpretation of the data. It might be that the differential response times to the test words result from their differential salience (or topicality) in a propositional representation. A flower picked to take with the girl for a present might be treated during comprehension as more relevant to the topic of its discourse than a flower smelled for a moment and then left behind (see discussions of discourse models by Grosz et al., 1983; Sidner, 1983a, 1983b; Webber, 1983). In Experiment 5, this alternative interpretation was tested by changing Glenberg et al.'s texts to add words that were topical but not model relevant. For example, the two versions of the flower text were changed so that the girl *bent down to an ornamental display* to pick a flower or smell a flower. The added propositions about the ornamental display contain the concept *flower*, and so they should vary in topicality as *flower* varies in topicality (cf. Kintsch, 1974). However, obviously, the display cannot accompany the girl, so the display cannot move with the girl in a situation model. Whether the girl picks the flower or merely smells it, the display is not part of the current situation at the end of the text. Thus, there are clearly contradictory predictions: In a situation model, when the girl picks the flower, the flower should be currently available at the end of the text but the ornamental display should not be. In a propositional representation, when flower is more salient, the display should also be more salient.

The procedure for Experiment 5 was the same as that used by Glenberg et al. (1987). Subjects read each text at a rate they controlled themselves, and, at some point during the text, a test word was presented for recognition.

Method

Materials. The 24 experimental texts were based on paragraphs used by Glenberg et al. (1987). For each text, there was a critical noun used by Glenberg et al. as the test word. In one version of the text, this noun stayed with the main character of the story as the action moved forward through the story. In the other version of the text, the noun was left behind the character as the action moved forward. We hypothesized that the critical noun was more salient in the texts for which it stayed with the main character than in the texts in which it was left behind. The texts were modified from those used by Glenberg et al. by the addition of a location for the critical noun. The location was always mentioned with the critical noun. For example, in the story just given as an example, the phrase "to an ornamental display" was added to give a location for the critical noun *flower*. The location was something that could not move with the main character.

The first sentence of each text was the same in both versions and served to introduce the main character. It averaged 10 words in length. The second sentence mentioned the critical noun and the location and was presented in one of two versions to manipulate whether or not the critical noun stayed with the character (a mean of 17 words in both cases). The third and fourth sentences completed the story (averaging 14 and 12 words, respectively). There was also a yes–no question associated with each text to test general comprehension of the story; the correct answer to 13 of the questions was yes, and the correct answer to 11 was no.

Filler texts (the same filler texts as were used by Glenberg et al., 1987) were chosen from a pool of 58 texts, ranging from 20 to 60 words in length. For 22 of these texts, there was a test word that had appeared in the text, and for the remainder, the test word had not appeared in any text. For 28 of the filler texts, the correct answer to the comprehension question was yes.

Procedure. The texts and test items were presented on a CRT screen, and responses were recorded on the CRT's keyboard. The CRT was controlled by a real-time microcomputer system.

The experiment began with a list of 30 lexical decision test items used to familiarize subjects with the response keys. After this practice, 10 filler texts were presented, and then the experiment proper began, with the 24 experimental texts and 24 filler texts randomly ordered. Presentation of each text began with the message "Press space bar" to initiate the text. When the space bar was pressed, the first sentence of the text was displayed. It remained on the CRT screen until the space bar was pressed again; then the screen was erased, and the next sentence was displayed. Sentences were presented in this way until the final sentence before the test word. When the space bar was pressed after reading of this sentence, the test word was displayed, with a row of asterisks underneath it. The test word remained on the screen until a response key was pressed, "/" for a positive response if the test word had appeared in the text just read or *z* key for a negative response if the test word had not appeared in the text. If the response was not correct, then the word *ERROR* was displayed for 1,500 ms before the next

sentence or yes–no question. If the text was one of the experimental texts, the test word was presented after the third sentence, and the fourth sentence was presented after the test word. For the filler texts, the test word was always presented after the last sentence of the text. After the text and its test word, the yes–no question appeared, and it remained on the screen until a response key was pressed. If the response was correct, the message to initiate the next text was presented. If the response was an error, then *ERROR* was displayed for 1,500 ms.

Subjects and design. There were four experimental conditions: The critical object either stayed with the main character until the end of the text remained behind, and the test word was either the critical noun or its location. These four conditions were combined in a Latin square design, with four groups of subjects (9 per group) and four groups of texts (6 per group). The subjects were 36 undergraduates from the same population as in Experiments 1 through 4.

Results

Means were calculated over responses for each subject and item in each condition, and means of these means are shown in Table 10.

Glenberg et al. 1987 showed that response times for the test word that was the critical noun were faster when the noun stayed with the character than when it was left behind. Table 10 shows a clear replication of this result; responses times for the critical noun test word were 66 ms faster when the noun stayed with the character. The important question is whether this same result obtains for the location test word. Our hypothesis was that the speed up in response times for the critical noun was due to its increased salience, not to the fact that it stays with the character. If this hypothesis is correct, then the speed up should also be obtained for the noun's location. The data confirm this hypothesis: Location response times were 55 ms faster when the critical noun was more salient (i.e., when the noun stayed with the character). This speed up in response times for the location test words is not predictable from a situation model. Thus, the results support the hypothesis that it was salience, not availability in a situation model, that was responsible for Glenberg et al.'s finding.

An ANOVA showed significantly faster response times for both noun and location test words when the noun stayed with the character, $F(1, 35) = 6.9$, and $F(1, 20) = 9.2$, and marginally significantly faster response times for the noun test words than the location test words, $F(1, 35) = 5.5$, and $F(1, 20) = 2.9$. The interaction between the two factors was not significant ($Fs < 1$). The standard error of the response time means was 18 ms. There were significantly more errors on the location test words than the noun test words, $F(1, 35) = 20.4$, and $F(1, 20) = 11.7$. There were no other significant effects on error rates ($Fs < 1.7$). Correct positive responses to the yes–no questions averaged 1,747 ms (12% errors) and correct negative responses, 1,810 ms (23% errors).

Table 10
Results of Experiment 5: Response Times (in Milliseconds) and Error Rates on Test Words

	Test word			
	Critical noun		Location	
Text version	RT	% error	RT	% error
Critical noun				
With main character	1,078	13	1,148	27
Behind main character	1,144	11	1,203	24
	Filler test word			
Positive filler	1,344[a]		20[b]	
Negative filler	1,199[a]		14[b]	

Note RT = response time.
[a] RT. [b] Percentage error.

Discussion

The Glenberg et al. (1987) and the Morrow et al. (1987; 1989) experiments fail to provide convincing evidence that real-life situation models are used automatically during comprehension because the results of both sets of experiments have interpretations that are consistent with the minimalist hypothesis. Of course, when readers have special goals or strategies, they can construct representations of quite complicated situations. Perrig and Kintsch's (1985) subjects were able to construct representations of the layout of a town from a text they read four times. Johnson-Laird's (1980) subjects, knowing they would have to draw a picture, were able to construct the relative positions of three objects from a textual description. As Johnson-Laird pointed out, constructing the information needed for a complete situation model can require considerable effort (see also Glenberg & Langston, 1992). If a passage in a story describes a complex scene with many interrelated objects, then a reader "would probably form only a rather vague idea of the actual spatial layout" (Johnson-Laird, 1980, p. 103). However, if the reader's goal were to answer a question about the relative location of a specific object, then the reader could use appropriate strategies and sufficient time during reading to construct the answer to the question. These strategies are clearly available to readers; the inferences required to represent lifelike information can be constructed. However, there is no empirical evidence to conclusively show that the inferences are constructed during reading by automatic processes.

An important conclusion to be drawn about constructionist inferences is that they contrast sharply with inferences that establish local coherence and inferences that make use of well-known information. Although these latter inferences (about propositional connections, reference, and well-known semantic relations) can be demonstrated easily in the prototypical laboratory experiment, there are no equivalently convincing demonstrations of the automatic encoding of real-life situational inferences.

General Discussion

It is widely believed that readers automatically construct inferences to build a relatively complete mental model of the situation described by a text (Glenberg et al., 1987; Johnson-Laird, 1980; Morrow et al., 1987; Rumelhart, 1975; Trabasso & van den Broek, 1985; van Dijk & Kintsch, 1983). However, our conclusion is that readers do not automatically encode the infer-

ences that would make up such a model. We base this conclusion on several points:

1. The empirical evidence that has been put forward to demonstrate the automatic encoding of a life-like situation model can be explained by the minimalist hypothesis.
2. Elaborative inferences that should be part of a lifelike situation model, for example, instrumental inferences, are not explicitly and automatically encoded.
3. Global inferences to connect widely separated parts of a story are not automatically encoded.

A wide range of data has been shown to be consistent with the minimalist hypothesis. For local inferences based on information in working memory, the minimalist claim is that they will be encoded because they are quickly and easily available. This claim has been verified for the inferences that connect propositions through argument repetition and anaphora. Minimalist tests of other kinds of local inferences, for example, the minimalist causal inferences proposed by van den Broek (1990) and Fletcher (Fletcher & van den Broek, 1989), await further research. For inferences based on general knowledge, again the minimalist claim is supported. Inferences about the instruments taken by verbs, about the contextually relevant meanings of words, and about the prototypical members of categories are encoded if the information on which they are based is easily available during reading.

In contrast, there is little data to support the constructionist position. Experiments 1 through 4 showed that causal global inferences are not automatically encoded during reading. Inferences that might be assumed to be encoded under all circumstances, such as instrumental inferences, are not. Experiments that have been cited as verifying constructionist situation models (e.g., Morrow et al., 1987, 1989) have alternative interpretations (Alba & Hasher, 1983), and Experiment 5 demonstrated the validity of one such alternative interpretation.

Besides being consistent with data, the minimalist hypothesis has several advantages over the constructionist hypothesis. The minimalist hypothesis is falsifiable in that it has clearly testable predictions: An inference is not constructed unless it is necessary to establish local coherence or it is supported by well-known, easily available information. These predictions provide a direct focus for empirical tests of the hypothesis. In contrast, constructionists have rarely provided an account of exactly which inferences should be encoded and which should not be.

The minimalist hypothesis is also a vital alternative hypothesis that more elaborative theories will have to take into account before they can become serious theories of text processing. So long as there is no convincing body of empirical evidence to support the constructionist view of automatic encoding processes, then the minimalist hypothesis remains viable. By providing an alternative to an otherwise widely accepted view, the minimalist hypothesis can lead to potentially fruitful investigations of inference processes.

It should be stressed, as it has been throughout this article, that the controversy between the constructionist and minimalist views is about the inferences that a reader encodes automatically, in the absence of specific goals and strategies. Neither minimalist nor constructionist theories propose models of how strategic, goal-specific inference processing is carried out, and this issue remains on the agenda for future research.

Two strategies for research are suggested by the minimalist hypothesis. One strategy is broadly exploratory and based on the hypothesis that the kinds of information that support inferences are those that are easily available. Experiments are designed to test for the inferences that might be supported by a variety of different kinds of knowledge. Another general issue is the organization of the local information in a text, and so experiments are designed to investigate the relative availabilities of the different entities evoked by a text. Overall, the strategy is to hold to the minimalist hypothesis while searching as widely as possible for evidence to force its rejection.

The current empirical situation is that there is no conclusive evidence to support the constructionist hypotheses, and therefore no reason to reject the minimalist hypothesis. An exploratory strategy is one way to face the challenges imposed by this situation—many different kinds of inferences can be examined, and the wider the range, the more stringent will be the test of the minimalist claims.

The second strategy for research is to construct explicit models of minimalist processes and then test the models empirically. This is the strategy adopted by Kintsch (1988; Kintsch, Welsch, Schmalhofer, & Zimny, 1990; Kintsch & Welsch, 1991) and by van den Broek (1990) and Fletcher (Fletcher & van den Broek, 1989). Van den Broek and Fletcher's model for local coherence was discussed earlier in this article. It provides a definition of local coherence in terms of causality and assumes that if locally available propositions (those in short-term memory at the same time) provide adequate causal relations for each other, then no other, more global causal relations are constructed. Thus, it is an example of a minimalist processing model.

Kintsch (1988) has proposed a construction–integration model whereby the concepts stated in a text and information from general knowledge associated with the concepts all interact to produce an encoded representation of the text (see also Ratcliff & McKoon, 1988). The construction–integration process can both change the relations among propositions that are explicitly stated in a text and add propositions to the representation. The construction–integration process is explicitly defined: Integration of long-term memory information and text information takes place through a repeated recycling of activation so that information associated only weakly to a relatively small portion of a text is further weakened, whereas information associated more strongly to multiple concepts in the text is strengthened. This process can change the organization of explicitly stated propositions because long-term memory associations can strengthen connections between propositions that would otherwise be only weakly connected. In this sense, the integration process represents the text as a situation model in a way that is consistent with minimalist rather than constructionist claims.

The construction–integration process can also add inferences to the text representation. For example, if a text contains the word *mint*, strong associations to both meanings are immediately activated, providing support for potential inferences. If other concepts in the text are associated with one of the meanings and not the other, then activation of these associations leads to an increase in activation of associations to the contextually appropriate meaning of *mint* and a decrease in activation of associations to the contextually inappropriate meaning. This

process results in the encoding of propositions (inferences) about the contextually appropriate meaning and implements the same claim as is made by the minimalist position, that information is added to a text representation to the extent that it is supported by easily available information.

Kintsch's (1988) model goes beyond the minimalist position in that it allows for the encoding of inferences that are based on information not immediately available. In the review of the literature presented in this article, the one finding inconsistent with the minimalist position was the finding that inferences about predictable events are encoded—not explicitly encoded as would be predicted by constructionist theories, but encoded to some degree. Kintsch showed how a minimalist processing model can account for this result. Inferences from long-term memory that are only weakly associated to individual concepts in a text can become more strongly activated over time if they are associated to several different concepts in the text. For the sentence "The townspeople were amazed to find that all the buildings had collapsed except the mint," the concept *earthquake* is not strongly associated with any individual concept in the sentence, and so it is not immediately available to support inferences. However, as the integration process proceeds, the recycling of activation from multiple sources may make this concept available as an inference to be added to the text representation.

Our goal in establishing the minimalist hypothesis is to stimulate research designed to find the principles by which inferences are generated. These principles might be defined across a number of kinds of inferences or within models of text processing. Either way, we believe that there will be two outcomes. First, both the minimalist and the constructionist hypotheses will be modified away from their all-or-none positions toward a more graded view of inference processing. This has already happened with inferences about predictable events. It cannot be said either that these inferences are completely and explicitly encoded or that they are not encoded at all. Instead, they are encoded to some degree, and finding evidence for them depends on finding the appropriate retrieval environment.

Second, the class of minimal inferences will be expanded. Currently, minimal inferences have been shown to include those that are supported by well-known semantic associations and well-known category membership relations. Expansion to include inferences that are supported by knowledge of the argument-taking properties of verbs has been tentatively suggested (Boland et al., 1990; Hudson, Tanenhaus, & Dell, 1987; McKoon & Ratcliff, 1989c; Tanenhaus et al., 1989). For example, in the sentence "he cleared the papers off" an argument of the verb (the place the papers were removed from) is missing. McKoon and Ratcliff (1989c) provided data indicating that readers inferred the missing argument from mention in a preceding sentence. Minimal inferences are also currently said to include those that are based on local information in a text. This notion too can be tentatively expanded toward a less simplistic view. Instead of regarding local information as a relatively undifferentiated list of propositions, we can borrow the discourse models used in artificial intelligence (Grosz et al., 1983; Sidner, 1983a, 1983b; Webber, 1983). According to these models, the information in a text is represented as the set of entities evoked by the text and the relations among them. Each entity in the model is assumed to have some degree of accessibility, which is determined by the syntactic, semantic, and pragmatic environment in which it is linguistically expressed. The varying degrees of accessibility should be reflected in the processes that construct local inferences. Initial evidence that inference processes are, in fact, affected in this way has been found in several studies (Greene et al., 1992; Hudson et al., 1987; McKoon, Ward, Ratcliff, & Sproat, in press; Ward et al., 1991). For example, comprehension of a pronoun is facilitated if the pronoun refers to an entity that is topical in its text, and if the referent entity was first mentioned in a salient syntactic position (Ward et al., 1991).

As the class of minimal inferences expands, the sharp contrast between the minimalist and constructionist positions may be redefined. For example, inferences previously thought to be constructed because they were necessary to a situation model may instead be understood to be based on easily available information and so become incorporated into a minimalist representation. At the same time, it may become clear which inferences cannot be constructed automatically, and for these inferences, models of strategic, goal-based generation processes will be required.

It is very important not to misunderstand the goal of the minimal inference position. It is easy to see it as a rejection of all goal-based, purposeful inference processing because this article is focused on minimal inferences. This is not the case. The aim is to try to separate the inferences and relations that are automatically and rapidly produced from those that are the result of slower, goal-based strategic processes. From such a separation, we can begin to understand the characteristics of the database provided in the first few hundred milliseconds of processing. Information about this database can then be used to tell us what information strategic processes have to work with (and therefore which strategic inferences will be difficult and which easy) and perhaps even identify strategic inferences that the processing system cannot avoid constructing.

References

Alba, J. W., Alexander, S., Hasher, L., & Caniglia, K. (1981). The role of context in the encoding of information. *Journal of Experimental Psychology: Human Learning and Memory, 7*, 283–292.

Alba, J. W., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin, 93*, 203–231.

Anderson, J. R., & Bower, G. H. (1973). *Human associative memory*. New York: Holt, Rinehart & Winston.

Anderson, R. C., & Ortony, A. (1975). On putting apples into bottles—A problem of polysemy. *Cognitive Psychology, 7*, 167–180.

Anderson, R. C., Pichert, J. W., Goetz, E., Schallert, D., Stevens, K., & Trollip, S. (1976). Instantiation of general terms. *Journal of Verbal Learning and Verbal Behavior, 15*, 667–679.

Baillet, S. D., & Keenan, J. M. (1986). The role of encoding and retrieval processes in the recall of text. *Discourse Processes, 9*, 247–268.

Barsalou, L. W. (1982). Context-independent and context-dependent information in concepts. *Memory & Cognition, 10*, 82–93.

Black, J. B., & Bower, G. H. (1980). Story understanding as problem solving. *Poetics, 9*, 223–250.

Bloom, C. P., Fletcher, C. R., van den Broek, P., Reitz, L., & Shapiro, B. P. (1990). An online assessment of causal reasoning during comprehension. *Memory & Cognition, 18*, 65–71.

Boland, J. E., Tanenhaus, M. K., & Garnsey, S. M. (1990). Evidence for the immediate use of verb control information in sentence processing. *Journal of Memory and Language, 29*, 413–432.

Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology, 11*, 177–220.

Bransford, J. D., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology, 3,* 193–209.

Bransford, J. D., & Franks, J. J. (1971). The abstraction of linguistic ideas. *Cognitive Psychology, 2,* 331–350.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 11,* 717–726.

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

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

Corbett, A. T. (1984). Pronominal adjectives and the disambiguation of anaphoric nouns. *Journal of Verbal Learning and Verbal Behavior, 17,* 683–695.

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 and Verbal Behavior, 17,* 479–491.

Daneman, M., & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19,* 450–466.

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

Dooling, D. J., & Lachman, R. (1971). Effects of comprehension on retention of prose. *Journal of Experimental Psychology, 88,* 216–222.

Dosher, B. A., & Corbett, A. T. (1982). Instrument inferences and verb schemata. *Memory & Cognition, 10,* 531–539.

Duffy, S. A. (1986). Role of expectations in sentence integration. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 208–219.

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

Flagg, P. W. (1976). Semantic integration in sentence memory? *Journal of Verbal Learning and Verbal Behavior, 15,* 491–504.

Flagg, P. W., & Reynolds, A. G. (1977). Modality of presentation and blocking in sentence recognition memory. *Memory & Cognition, 5,* 111–115.

Fletcher, C. R., & Bloom, C. (1988). Causal reasoning in the comprehension of simple narrative texts. *Journal of Memory and Language, 27,* 235–244.

Fletcher, C. R., Hummel, J. E., & Marsolek, C. J. (1990). Causality and the allocation of attention during comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 233–240.

Fletcher, C. R., & van den Broek, P. (1989). *A model of narrative comprehension and recall.* Manuscript submitted for publication.

Forster, K. I. (1981). Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing. *Quarterly Journal of Experimental Psychology, 33,* 465–495.

Garnsey, S. M., Tanenhaus, M. K., & Chapman, R. M. (1989). Evoked potentials and the study of sentence comprehension. *Journal of Psycholinguistic Research, 18,* 51–60.

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

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

Glanzer, M., Fischer, B., & Dorfman, D. (1984). Short-term storage in reading. *Journal of Verbal Learning and Verbal Behavior, 23,* 467–486.

Glenberg, A. M., & Langston, W. E. (1992). Comprehension of illustrated text: Pictures help to build mental models. *Journal of Memory and Language, 31,* 129–151.

Glenberg, A. M., Meyer, M., & Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory and Language, 26,* 69–83.

Graesser, A. C. (1981). *Prose comprehension beyond the word.* New York: Springer.

Graesser, A. C., Robertson, S. P., & Anderson, P. A. (1981). Incorporating inferences in narrative representations: A study of how and why. *Cognitive Psychology, 13,* 1–26.

Greene, S. B., McKoon, G., & Ratcliff, R. (1992). Pronoun resolution and discourse models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 266–283.

Greenspan, S. L. (1986). Semantic flexibility and referential specificity of concrete nouns. *Journal of Memory and Language, 25,* 539–557.

Grosz, B. J., Joshi, A. K., & Weinstein, S. (1983). Providing a unified account of definite noun phrases in discourse. *Proceedings of the 21st Annual Meeting of the Association of Computational Linguistics.* Association of Computational Linguistics.

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

Hudson, S. B., Tanenhaus, M. K., & Dell, G. S. (1987). The effect of the discourse center on the local coherence of a discourse. *Proceedings of the 8th Annual Cognitive Science Meetings.*

Johnson, M. K., Bransford, J. D., & Solomon, S. K. (1973). Memory of tacit implications of sentences. *Journal of Experimental Psychology, 98,* 203–205.

Johnson-Laird, P. N. (1980). Mental models in cognitive science. *Cognitive Science, 4,* 71–115.

Katz, S., & Gruenewald, P. (1974). The abstraction of linguistic ideas in "meaningless" sentences. *Memory & Cognition, 2,* 737–741.

Keenan, J. M., Baillet, S. D., & Brown, P. (1984). The effects of causal cohesion on comprehension and memory. *Journal of Verbal Learning and Verbal Behavior, 23,* 115–126.

Keenan, J. M., & Kintsch, W. (1974). The identification of explicitly and implicitly presented information. In W. Kintsch (Ed.), *The representation of meaning in memory* (pp. 153–176). Hillsdale, NJ: Erlbaum.

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

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95,* 163–182.

Kintsch, W., & Glass, G. (1974). Effects of propositional structure on text recall. In W. Kintsch (Ed.), *The representation of meaning in memory* (pp. 146–152). Hillsdale, NJ: Erlbaum.

Kintsch, W., & Keenan, J. M. (1973). Reading rate and retention as a function of the number of propositions in the base structure of sentences. *Cognitive Psychology, 5,* 257–274.

Kintsch, W., Kozminsky, E., Streby, W. J., McKoon, G., & Keenan, J. M. (1975). Comprehension and recall of text as a function of content variables. *Journal of Verbal Learning and Verbal Behavior, 14,* 196–214.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85,* 363–394.

Kintsch, W., & Welsch, D. M. (1991). The construction-integration model: A framework for studying memory for text. In W. E. Hockley & S. Lewandowsky (Eds.), Relating theory and data: Essays on human memory (pp. 363–367). Hillsdale, NJ: Erlbaum.

Kintsch, W., Welsch, D. M., Schmalhofer, F., & Zimny, S. (1990). Sen-

tence memory: A theoretical analysis. *Journal of Memory and Language, 29,* 133–159.

Mandler, J. M. (1978). A code in the node: The use of a story schema in retrieval. *Discourse Processes, 1,* 14–35.

Mandler, J. M., & Johnson, N. J. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology, 9,* 111–151.

Mani, K., & Johnson-Laird, P. N. (1982). The mental representation of spatial descriptions. *Memory & Cognition, 10,* 181–187.

McKoon, G. (1977). Organization of information in text memory. *Journal of Verbal Learning and Verbal Behavior, 16,* 247–260.

McKoon, G., & Keenan, J. M. (1974). Response latencies to explicit and implicit statements as a function of the delay between reading and test. In W. Kintsch (Ed.), *The representation of meaning in memory* (pp. 166–176). Hillsdale, NJ: Erlbaum.

McKoon, G., & Ratcliff, R. (1980a). The comprehension processes and memory structures involved in anaphoric reference. *Journal of Verbal Learning and Verbal Behavior, 19,* 668–682.

McKoon, G., & Ratcliff, R. (1980b). Priming in item recognition: The organization of propositions in memory for text. *Journal of Verbal Learning and Verbal Behavior, 19,* 369–386.

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

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

McKoon, G., & Ratcliff, R. (1988). Contextually relevant aspects of meaning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 331–343.

McKoon, G., & Ratcliff, R. (1989a). Assessing the occurrence of elaborative inference with recognition: Compatibility checking vs. compound cue theory. *Journal of Memory and Language, 28,* 547–563.

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

McKoon, G., & Ratcliff, R. (1989c, November). *Inferences based on lexical information about verbs.* Paper presented at the 30th Annual Meeting of the Psychonomic Society, Atlanta, GA.

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

McKoon, G., & Ratcliff, R. (1989e). Textual inferences: Models and measures. In D. A. Balota, G. B. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 403–421). Hillsdale, NJ: Erlbaum.

McKoon, G., & Ratcliff, R. (1990). Dimensions of inference. In A. Graesser & G. Bower (Eds.), *The psychology of learning and motivation* (Vol. 25, pp. 313–328). San Diego, CA: Academic Press.

McKoon, G., & Ratcliff, R., & Seifert, C. (1989). Making the connection: Generalized knowledge structures in story understanding. *Journal of Memory and Language, 28,* 711–734.

McKoon, G., Ward, G., Ratcliff, R., & Sproat, R. (in press). Morphosyntactic and pragmatic factors affecting accessibility of discourse entities. *Journal of Memory and Language.*

Moeser, S. D. (1976). Inferential reasoning in episodic memory. *Canadian Journal of Psychology, 31,* 41–70.

Morrow, D., Bower, G., & Greenspan, S. (1989). Updating situation models during narrative comprehension. *Journal of Memory and Language, 28,* 292–312.

Morrow, D., Greenspan, S., & Bower, G. (1987). Accessibility and situation models in narrative comprehension. *Journal of Memory and Language, 26,* 165–187.

Myers, J. L., Shinjo, M., & Duffy, S. A. (1987). Degree of causal relatedness and memory. *Journal of Memory and Language, 26,* 453–465.

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

Omanson, R. C. (1982a). An analysis of narratives: Identifying central, supportive and distractive content. *Discourse Processes, 5,* 195–224.

Omanson, R. C. (1982b). The relation between centrality and story category variation. *Journal of Verbal Learning and Verbal Behavior, 21,* 326–337.

Paris, S., & Lindauer, B. K. (1976). The role of inference in children's comprehension and memory for sentences. *Cognitive Psychology, 8,* 217–227.

Perrig, W., & Kintsch, W. (1985). Propositional and situational representations of text. *Journal of Memory and Language, 24,* 503–518.

Potts, G. R., Keenan, J. M., & Golding, J. M. (1988). Assessing the occurrence of elaborative inference: Lexical decision versus naming. *Journal of Memory and Language, 27,* 399–415.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88,* 93–134.

Ratcliff, R., & McKoon, G. (1978). Priming in item recognition: Evidence for the propositional structure of sentences. *Journal of Verbal Learning and Verbal Behavior, 17,* 403–417.

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

Ratcliff, R., & McKoon, G. (1981b). Does activation really spread? *Psychological Review, 88,* 454–462.

Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review, 95,* 385–408.

Reitman, J. S., & Bower, G. H. (1973). Storage and later recognition of exemplars of concepts. *Cognitive Psychology, 4,* 194–206.

Roth, E. M., & Shoben, E. J. (1983). The effect of context on the structure of categories. *Cognitive Psychology, 15,* 346–378.

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.* San Diego, CA: Academic Press.

Rumelhart, D. E. (1977). Understanding and summarizing brief stories. In D. Laberge & J. Samuels (Eds.), *Basic processes in reading: Perception and comprehension* (pp. 263–303). Hillsdale, NJ: Erlbaum.

Schmalhofer, F., & Glavanov, D. (1986). Three components of understanding a programmer's manual: Verbatim, propositional, and situational representations. *Journal of Memory and Language, 25,* 279–294.

Seifert, C. M., McKoon, G., Abelson, R. P., & Ratcliff, R. (1986). Memory connections between thematically similar episodes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 220–231.

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

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

Singer, M. (1978, August–September). *The role of explicit and implicit recall cues.* Paper presented at the 86th Annual Convention of the American Psychological Association, Toronto, Ontario, Canada.

Singer, M. (1979). Processes of inference during sentence encoding. *Memory & Cognition, 7,* 192–200.

Stein, N. L., & Glenn, C. (1979). An analysis of story comprehension in

elementary school children. In R. O. Freedle (Ed.), *New directions in discourse processing.* Hillsdale, NJ: Erlbaum.

Suh, S., & Trabasso, T. (1988, November). *Convergent evidence on inferences during comprehension of text.* Paper presented at the Annual Meeting of the Psychonomic Society, Chicago.

Swinney, D., & Osterhout, L. (1990). Inference generation during auditory language comprehension. In A. Graesser & G. Bower (Eds.), *The psychology of learning and motivation* (Vol. 25, pp. 17–33). San Diego, CA: Academic Press.

Tabossi, P. (1982). Sentential context and the interpretation of unambiguous words. *Quarterly Journal of Experimental Psychology, 34,* 79–90.

Tabossi, P., & Johnson-Laird, P. N. (1980). Linguistic context and the priming of semantic information. *Quarterly Journal of Experimental Psychology, 32,* 595–603.

Tanenhaus, M. K., Carlson, G. N., & Seidenberg, M. S. (1985). Do listeners compute linguistic representations? In D. Dowty, L. Kartunnen, & A. Zwicky (Eds.), *Natural language parsing.* Cambridge, England: Cambridge University Press.

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

Thorndyke, P. W. (1977). Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology, 9,* 77–110.

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.

Trabasso, T., & Sperry, L. L. (1985). Causal relatedness and importance of story events. *Journal of Memory and Language, 24,* 595–611.

Trabasso, T., & van den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language, 24,* 612–630.

van den Broek, P. (1988). The effects of causal relations and hierarchical position on the importance of story statements. *Journal of Memory and Language, 27,* 1–22.

van den Broek, P. (1990). The causal inference maker: Towards a process model of inference generation in text comprehension. In D. A. Balota, G. B. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 423–446). Hillsdale, NJ: Erlbaum.

van den Broek, P., & Trabasso, T. (1986). Causal networks versus goal hierarchies in summarizing text. *Discourse Processes, 9,* 1–15.

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

Ward, G., Sproat, R., & McKoon, G. A. (1991). Pragmatic analysis of so-called anaphoric islands. *Language, 67,* 439–474.

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

Wilson, S. G., Rinck, M., McNamara, T. P., Bower, G. H., & Morrow, D. G. (1992). *Mental models and narrative comprehension. Some qualifications.* Unpublished manuscript.

Received September 10, 1990
Revision received March 20, 1991
Accepted June 12, 1991 ■

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

0278-7393/92/$3.00

Pronoun Resolution and Discourse Models

Steven B. Greene
Stanford University

Gail McKoon and Roger Ratcliff
Northwestern University

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

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

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

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

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

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

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

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

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

Barry saw Harriet.

and

It was Barry who saw Harriet.

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

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

are shorter than those for this sentence:

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

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

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

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

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

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

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

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

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

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

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

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

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

Table 1
Example of the Experimental Texts

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

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

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

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

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

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

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

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

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

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

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

Experiments 1 and 2

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

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

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

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

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

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

Method

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

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

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

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

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

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

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

Results

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

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

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

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

In Experiment 1, with Test Positions 1 and 2, there were no significant differences between the test positions (both F_1 and $F_2 < 1.3$) and no interactions between test word and test position ($Fs < 1.7$). The only significant effect was for test word, $F_1(2, 70) = 61.4$ and $F_2(2, 118) = 64.0$. The response times for the control test word were slower than for the other test words ($F > 42$). The standard error of the response time means was 6 ms. The only significant effect for error rates was the difference among test words, $F_1(2, 70) = 24.7$ and $F_2(2, 118) = 19.1$; the control test words had more errors than the other test words ($Fs > 13$). The standard error for errors was 1.6%.

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

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

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

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

Procedure Check

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

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

Experiment 3

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

Method

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

Results

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

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

Experiment 4

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

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

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

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

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

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

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

Method

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

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

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

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

Results

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

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

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

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

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

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

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

Experiments 5 and 6

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

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

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

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

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

Method

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

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

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

Results

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

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

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

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

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

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

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

Discussion

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

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

Experiment 7

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

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

Method

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

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

Results and Discussion

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

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

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

Experiments 8 and 9

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

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

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

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

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

Method

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

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

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

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

Results and Discussion

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

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

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

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

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

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

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

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

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

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

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

General Discussion

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

I need a plate. Where do you keep them?

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

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

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

McKoon, G., & Ratcliff, R. (1980). The comprehension processes and memory structures involved in anaphoric reference. *Journal of Verbal Learning and Verbal Behavior, 19,* 668–682.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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