JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR 19, 668–682 (1980)

The Comprehension Processes and Memory Structures Involved in Anaphoric Reference

GAIL MCKOON AND ROGER RATCLIFF

Dartmouth College

The inference processes involved in anaphoric reference were examined in three experiments. The first two experiments used an activation procedure in which the subject read a paragraph sentence by sentence and was then presented with a single test word from the first sentence of the paragraph for recognition. Response time to the test word was speeded both when the test word was a referent of an anaphor mentioned in the last sentence of the paragraph and when the test word was in the same proposition as the referent of the anaphor in the last sentence. This shows that an anaphor activates both its referent and concepts in the same proposition as its referent. The third experiment used a priming technique to show that the referent and a concept in the same proposition as the anaphor are connected in the long-term memory representation of a text. These results are discussed in terms of a simple three-process model of anaphoric inference. Other methodologies used to study inference processing are evaluated and it is concluded that these methods involve considerable problems in experimental design and theoretical interpretation and are also limited in the kinds of information they can provide about inference processing.

This paper is concerned with the investigation of inference processes in reading. As a first step, it is reasonable to study very simple inference processes, such as the processes involved in anaphoric reference. Anaphoric reference in a text involves the reference from a concept such as *vehicle* back to its earlier referent *car*. We shall present a fairly simple notion of the processes involved in these inferences and evaluate methodologies that have been developed for their study. Then we shall present experiments that give new evidence for the processes and their resulting memory structures and also provide a better methodology for the study of inference processes.

The theories of inference that have been proposed involve three component processes (Chafe, 1972; Clark, 1978; Haviland & Clark, 1974; Kintsch & Vipond, 1978): First, the concept to be inferred has to be accessed in memory. For example, if a text mentions *the vehicle*, then a referent mentioned earlier in the text, perhaps *car*, must be found. Second, this concept (*car*) has to be activated (or brought into short-term or working memory). Kintsch and Vipond (1978) propose that not only the concept, but also propositions that contain the concept are brought into working memory. Third, the information that caused the concept to be activated has to be connected to the concept. Propositions about *the vehicle* have to be connected to the concept *car* and, through *car*, to propositions about *car*. The connected structure, the result of the three processes, is the structure that is stored in long-term memory as a representation of the text. This three-process theory will provide the background to the following discussion of the methodology of the study of inference processing.

One method of studying inference processes that has been popular recently is cued recall. The subject reads a list of short texts (usually sentences) and then, at the end of the list or after some delay, is given a

The research reported here was supported by NICHD Grant HD 13318-01 to Gail McKoon. We wish to thank Mike Layton for developing our microcomputer laboratory, Patt McCauley for writing the paragraphs used in the experiments, and Gary Dell for suggesting the control experiment mentioned in the discussion of Experiment 2. Reprint requests should be sent to Gail McKoon, Department of Psychology, Dartmouth College, Hanover, N.H., 03755.

0022-5371/80/060668-15$02.00/0

list of cues (single words) and is asked to recall the sentence that is suggested by each cue. It has been shown that a cue that represents information that could be inferred from a sentence will function as an effective retrieval cue (cf. Anderson & Ortony, 1975; Anderson, Pichert, Goetz, Schallert, Stevens, & Trollip, 1976; Barclay, Bransford, Franks, McCarrell, & Nitsch, 1974; Paris & Lindauer, 1976; Till, 1977). For example, given the sentence *The container held the apples*, the cue *basket* will be more effective as a retrieval cue than *bottle* (Anderson & Ortony, 1975).

Although cued recall studies demonstrate that subjects *can* make the inferences represented by the cues, no conclusion can be drawn about *when* the subjects make the inferences. It could be that the inferences are made during reading or it could be that they are made at the time of the cued recall test. In other words, the inference process studied in cued recall tasks could be either an encoding or a retrieval phenomenon (see Tulving, 1976).

In general, the authors of cued recall studies argue that the inferences reflected in their cued recall tasks are an encoding phenomenon (cf. Anderson, Goetz, Pichert, & Halff, 1977; Anderson & Ortony, 1975; Till, 1977). However, several recent experiments suggest that they are in fact a retrieval phenomenon. A study by Singer (Note 1) indicates that subjects use the recall cues to "work backwards" to the sentences to be recalled. In one of his experiments, Singer compared the effectiveness of two recall cues that differed in forward and backward association to the action expressed by the to-be-recalled sentence. For example, the association from cue *ladle* to the action *stirring soup* is strong, while the reverse association (*stirring soup* to *ladle*) is weak. On the other hand, the association from *spoon* to *stirring soup* is weak, while the reverse association is strong. Results showed that *ladle* was a better recall cue than *spoon*, suggesting that subjects generate backward associations from *ladle* in order to recall *stirring soup*. This study supports the view that the inferences demonstrated in cued recall experiments are retrieval phenomena. Further support for this view was obtained by Corbett and Dosher (1978). The to-be-remembered sentences in their experiment included sentences like *The worker pounded the nail with the rock*, *The worker pounded the nail*, and *The worker pounded the nail with the hammer*. *Hammer* was equally effective as a cue for all three sentences. This result contradicts the view that an implied instrument is an effective recall cue because it was inferred during encoding, and supports the view that the inferences that have been studied with the cued recall technique are retrieval phenomena.

This discussion of cued recall as a method for studying inference has demonstrated two points. First, cued recall may be tapping either an encoding or a retrieval phenomenon. It cannot, therefore, be used to study inferences made during encoding unless convincing evidence is provided that it is encoding that is being studied. In fact, and this is the second point, experimental evidence suggests that cued recall taps retrieval processes rather than encoding processes.

Another measure that has become widely used in the study of inference is reading time (Clark & Sengal, 1979; Garrod & Sanford, 1977; Haberlandt & Bingham, 1978; Haviland & Clark, 1974; Lesgold, Roth, & Curtis, 1979; Sanford & Garrod, 1980; Singer, 1979; Yekovich & Walker, 1978; Keenan, Note 2). In a typical experiment, subjects are presented with a text one sentence at a time. Subjects initiate the presentation of each sentence by pressing a key. Reading time for a sentence is defined as the time between the key press to initiate that sentence and the key press to initiate the next sentence. It is assumed, first, that the reading time for a sentence is the time to comprehend the sentence, and, second, that longer reading times represent greater amounts of processing required for comprehension. Thus, reading time is assumed to give an "on-line" indication of inference processes; when an inference is required, reading time increases. For example, consider the following two pairs of sentences (Haviland & Clark, 1974):

1. John left the beer in the car.
   The beer was too warm to drink.
2. John left the picnic supplies in the car.
   The beer was too warm to drink.

Reading time for the second sentence of 2 is found to be longer than reading time for the second sentence of 1 and this is assumed to be because an inference is required in 2.

There are several problems with the reading time methodology, and they fall into two classes. First, it is difficult (although not impossible) to design experiments that answer theoretical questions unambiguously, that is, experiments where the observed data cannot be attributed to some uninteresting confounding variable. Second, even if an experiment is designed in which there are no confounding variables, there are severe limitations on the kinds of theoretical questions that can be addressed. These two problems will be taken up in order.

First, reading time differences that have been obtained in published studies and attributed to inference processes can instead be attributed to several kinds of confounding variables. Some of these variable are fairly trivial; for example, certain critical words in the target sentence may be repetitions of words in earlier sentences in one condition (where no inference is supposed to be required) but not in another condition (where an inference is supposed to be required). Thus the differences in reading time of the target sentence may be the result of uninteresting repetition effects rather than inference processes. Even if words are not repeated, a similar problem arises if words in the target sentence are semantically related to earlier words in one condition but not in another.

Careful design to avoid repetition effects will not, however, solve all problems with reading time measures. Increases in reading time in conditions that are supposed to involve inference processing may be due, not to inference processing, but instead to subjects' recognitions that such processing is required. That is, increases in reading time may be due to the time it takes the subject to decide that the inference processing is required plus the time it takes to decide not to perform this processing (there are an awful lot of these paragraphs to read and the pay isn't great). In other words, increases in reading times may simple reflect judgements of comprehensibility (or incomprehensibility). Yekovich, Walker, and Blackman (1979) have shown that reading times do, in fact, mirror comprehensibility ratings.

Even if comprehensibility can be shown equal across experimental conditions, there is still another problem with reading time measures. The usual way of performing reading time studies is to hold a target sentence constant and vary the preceding context across conditions (see 1 and 2 above). Clearly, comprehension time for the target independent of the context sentences cannot vary with condition because the words of the target are exactly the same across conditions. But it cannot be assumed that reading time of the context sentences independent of the target does not vary across conditions. If reading time of the context sentences does vary across conditions, then reading time for the target may be increased not because the target requires inference processing, but because of the context sentences. The subject may still be processing the context sentences while, or instead of, processing the target or there may be general sequential effects, with a slow reading time of the context sentences producing a subsequent slow reading time of the target. Such "spillover" effects have been reported by Haberlandt and Bingham (1978) not only from one sentence to the next but from the last sentence of one paragraph to the first sentence of the next, unrelated, paragraph.

With careful experimental design, it would be possible to avoid all of these problems with confounding variables; for example, comprehensibility ratings and reading times for context sentences could both be shown to be equal across experimental conditions. However, the reading time technique is still limited in the kinds of theoretical questions that can be addressed. Reading time can indicate when increased processing is required, but it cannot indicate what that processing is. With respect to inference, reading time may be longer because the subject searches for a to-be-inferred concept but does not find it. Or the concept may be found and activated but not connected to the new propositions as it should be. Even if the proper connections are made, they may not be stored in the long-term memory representation of the text. In general, reading time cannot be used to investigate the component processes of inference.

By changing the reading time technique somewhat, an experimenter can be sure that subjects are completing the inference processes. For example, after reading each sentence, subjects can be required to judge whether or not that sentence of a text contradicts preceding sentences (Carpenter & Just, 1977) or required to say aloud the referent of an anaphor (Caramazza, Grober, & Garvey, 1977). Alternatively, the time required to verify that a test sentence is true according to a previously read text can be measured (Anderson & Hastie, 1974; Keenan & Kintsch, 1974; McKoon & Keenan, 1974; Reder, 1979). If response times for sentences that express interferences are equal to response times for sentences that express information explicitly stated in the text, then it can be inferred that subjects have completed the inference processes and stored the result in the long-term memory representation of the text. With neither of these methods, however, can the component processes of inference be investigated.

We shall now present three experiments that use relatively new methodologies to examine the subprocesses of inference processing. These experiments avoid the problems associated with cued recall and reading time procedures. All three experiments use priming or activation techniques. In these techniques, the subject is required to respond as to whether a test word occurred in a previously presented text. A speed-up in response time to the test word is assumed to show that the test word was activated (perhaps brought into working memory) by information that immediately preceded the test word. For example, McKoon and Ratcliff (1980, Note 3; also Ratcliff & McKoon, 1978) have shown that when a test word is immediately preceded by a word that was closely associated to it in a previously studied text, then response time to the test word is speeded. Ratcliff and McKoon (Note 4) have shown that this priming process can be termed automatic in the sense of Posner and Snyder (1975). This is because the size of the priming effect is independent of the probability that two sequentially presented test words are associated and because the onset of priming is of the order of 100 milliseconds, considerably faster than strategic processes (see Ratcliff & McKoon, Note 4, for further discussion).

The priming technique was used to look at two of the processes involved in anaphoric reference: First, the state of activation of concepts was examined with a variation of the priming technique used by Caplan (1972) and successfully applied to inference by Chang (1980). In this technique, a single word is presented for recognition immediately after the text has been read by the subject. Reaction time for the "old" response is measured. The assumption made is that if the word tested was active (or in working memory, Kintsch & Vipond, 1978), then response time to that word will be decreased. For example, suppose that the last sentence of a text mentioned an anaphor (e.g., *vehicle*) and the referent (e.g., *car*) was activated during comprehension of the sentence. Then response time to the test word *car* should be speeded relative to a control condition in which the last sentence did not mention either *car* or *vehicle*. This activation procedure was used in Experiments 1 and 2.

The second part of the anaphoric reference process that was examined was the memory structure that results from anaphoric reference. Memory structure was investigated using the priming technique of McKoon and Ratcliff (1980; Ratcliff and McKoon, 1978). Subjects read texts and then a test list of single words was presented for recognition. It was expected that priming effects would show whether successively tested words were relatively closely connected in memory. For example, if a text stated *the vehicle crashed into the snowman* and this sentence activated the referent *car* so that the connection between *car* and *snowman* was stored in the memory representation of the text, then *car* should prime *snowman* in the test list. This priming procedure was used in Experiment 3, and a similar procedure was used in a fourth experiment designed to rule out alternative interpretations of the results of Experiment 3.

EXPERIMENT 1

The aim of this experiment was to investigate the process of activation in anaphoric reference. We wished to answer the question, is a referent activated (or brought into working memory) by presentation of an appropriate anaphor? Subjects read paragraphs sentence by sentence, and, after reading the last sentence of a paragraph, were presented with a single test word for recognition. Table 1 shows one example of the paragraphs used in this experiment. Three four-sentence versions of the paragraph are shown. Three two-sentence versions were also used in the experiment. The two-sentence versions were the same as the four-sentence versions except that the middle two sentences were omitted. The critical test word for this paragraph was *burglar*, whichever of the six versions was read. We expected that, when the final sentence mentioned *criminal*, response time to *burglar* would be speeded relative to the condition in which the final sentence mentioned *cat*. This priming effect would be taken as evidence for the activation of the concept *burglar* by the word *criminal*.

Method

Subjects. The subjects were 36 Dartmouth undergraduates who participated in the experiment for extra credit in an introductory psychology course.

Materials. Sixty paragraphs like the one shown in Table 1 were written. The first sentence of each paragraph contained the critical word (e.g., *burglar*) as the first noun. The second and third sentences did not repeat this critical word. There were three possible fourth sentences for each

TABLE 1

AN EXAMPLE OF THE PARAGRAPHS USED IN EXPERIMENTS 1, 2, AND 3

Critical word: *burglar*

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 4a: The burglar slipped away from the streetlamp.
Sentence 4b: The criminal slipped away from the streetlamp.
Sentence 4c: A cat slipped away from the streetlamp.

Sentence 4a mentions the critical word.
Sentence 4b mentions the category of the critical word.
Sentence 4c mentions a word unrelated to the critical word.

paragraph. In one, the critical word was the first noun. In a second, the category to which the critical word belonged was the first noun (e.g., *criminal*). In the third, the first noun was a word not related to the critical word (e.g., *cat*). Except for the first noun and sometimes the verb, the three fourth sentences were identical in wording. The test word for each paragraph was the critical word (e.g., *burglar*).

Sixty more paragraphs were written to serve as fillers. There were 14 two-sentence paragraphs with test words which were not in the paragraphs (negative test words), 16 three-sentence paragraphs, 6 with negative and 10 with positive test words, 14 four-sentence paragraphs with negative test words, and 16 five-sentence paragraphs, 6 with negative test words and 10 with positive test words.

Procedure. Subjects were tested individually in one 50-minute session. Presentation of all materials was controlled by a microcomputer driven by Dartmouth's time-sharing computer system.

Each subject read 8 practice paragraphs, the 60 experimental paragraphs, and the 60 filler paragraphs. Presentation of each paragraph began with an instruction to the subject to press the space bar on a CRT keyboard to initiate the paragraph. Then the first sentence of the paragraph appeared on the screen. The subject was instructed to read it carefully and press the space bar when he was sure he understood it. When the space bar was pressed, the first sentence disappeared from the screen and the second sentence appeared. The subject continued in this way through the sentences of the paragraph. When the subject pressed the space bar after reading the final sentence of the paragraph, a row of asterisks with the test word below it appeared on the screen. The subject was instructed to respond to the test word by pressing either a "yes" or a "no" key (the "/" and "z" keys, respectively, on the CRT keyboard), according to whether or not the test word appeared in the paragraph. Subjects were instructed to respond as quickly and accurately as possible. The test word remained on the screen until the response was made; then it disappeared from the screen and the instruction to press the space bar to begin the next paragraph appeared. Because of the variable number of sentences in the filler paragraphs, the subject did not know which was the final sentence until after he had finished reading it and pressed the space bar. A different random order of presentation of paragraphs was used for each two subjects.

In an effort to persuade subjects to read the paragraphs carefully, a sentence verification test was given after every ten paragraphs. One sentence was chosen from each paragraph; if it was to be a false test item, one word was changed to alter its meaning. Half of these test items were false, half were true.

Design. There were six experimental conditions; the experimental paragraphs were either two or four sentences long and the final sentence mentioned the critical word, the category of the critical word, or an unrelated word. The first sentence of a paragraph was always the same (mentioning the critical word), whether the paragraph was short or long, and the test word was always the critical word. The six experimental conditions were crossed with six groups of subjects (six per group) and six sets of paragraphs (six per set) in a Latin-square design.

Results and Discussion

The results are shown in Table 2. All analyses were based on mean response times for each subject or each item in each condition.

Mean response times to the test word were significantly affected by whether the final sentence mentioned the critical word, the category of the critical word, or an unrelated word, $F(2,70) = 74.5, p < .001$, with subjects as the random variable, and $F(2,118) = 46.6, p < .001$, with materials as the random variable. The main effect of

TABLE 2
RESULTS FROM EXPERIMENT 1

	Final sentence reading time[a] (msec)	Test Word (burglar)	
		Response time[b] (msec)	Proportion of errors
Final sentence			
burglar	2275	641	.11
criminal	2346	730	.10
cat	2529	758	.12

[a] Average standard error was 120 msec.
[b] Average standard error was 24 msec.

length of the paragraph was not significant, $F < 1.4$ in both subjects and materials analyses. However, length did affect response times when the final sentence mentioned the critical word; with short paragraphs, mean response time was 662 milliseconds and with long paragraphs, 620 milliseconds. The interaction between length and final sentence was significant, $F(2,70) = 4.2, p < .02$, with subjects as the random variable, and $F(1,118) = 6.7, p < .02$, with materials as the random variable. Length also affected error rate; there were 8% errors wtih short paragraphs and 14% errors with long paragraphs.

If the category of the critical word served to activate the critical word or bring it into working memory, then the response time to the critical word should have been faster when the final sentence mentioned the category (*criminal*) than when it mentioned an unrelated word (*cat*). Comparison of response times in these two conditions showed the expected difference, $F(1,70) = 7.9, p < .01$, with subjects as the random variable, and $F(1,118) = 6.7, p < .02$, with materials as the random variable. Thus, it appears that the anaphor did activate its referent.

Mean reading times for the final sentences (shown in Table 2) depended on whether the final sentence mentioned the critical word, the category, or the unrelated word, $F(2,70) = 11.0, p < .001$, with subjects as the random variable, and $F(2,118) = 7.6, p < .001$, with materials as the random variable. This result may be due simply to the different words (*burglar, criminal,* and *cat*) in the different conditions. On the other hand, it may be that the reading time for the *cat* sentence is slower because of the introduction of a new concept that had not previously appeared in the text. Support for this hypothesis is provided by a study by Kintsch, Kozminsky, Streby, McKoon, and Keenan (1975) where it was found that the reading time for a text increased with the number of concepts in the text. These two explanations of the differences in reading times for the different versions of the final sentences could be discriminated by measuring the reading times of the sentences individually, without any preceding context sentences.

Length of the paragraph also affected reading times, with marginal significance, $F(1,35) = 3.1, p = .08$ (subjects analysis), and $F(1,59) = 4.1, p < .05$ (materials analysis). The longer paragraphs were read faster (2322 milliseconds) than the short paragraphs (2444 milliseconds), which suggests subjects read each succeeding sentence more quickly as they moved through the paragraphs. The interaction between final sentence and length was not significant.

The main result of Experiment 1 is that presentation of an anaphor served to activate its referent. The measure of activation was the amount of priming given by the anaphor in the final sentence to the referent test word. The priming effect was larger when the anaphor was the referent word itself than when it was the category; this is probably due simply to repetition of the word.

Experiment 2

The first experiment showed that an anaphor produced activation of its referent. In the text processing model of Kintsch and Vipond (1978), it is assumed that an anaphor activates not only its referent but also propositions containing the referent. In the example in Table 1, the first sentence of the paragraph expresses the proposition (SURVEY, BURGLAR, GARAGE). Then, if the final sentence mentions *burglar* or *criminal,* not only *burglar* but also *garage* should be activated. Experiment 2 was designed to test this hypothesis.

Method

In every respect except the test word of the experimental paragraphs, Experiment 2 was the same as Experiment 1. For the experimental paragraphs, the test word was a noun from the first sentence of the paragraph that was in the same proposition as the critical word. For the paragraph in Table 1, the critical word was *burglar* (the test word in Experiment 1); the test word in Experiment 2 was *garage*. This word was not repeated in any other sentence of the paragraph.

Results and Discussion

The results are shown in Table 3. All analyses were based on mean response times for each subject of each item in each condition.

Mean response times to the test word (*garage*) were significantly affected by whether the final sentence mentioned the critical word (*burglar*), the category (*criminal*), or an unrelated word (*cat*), $F(2,70) = 6.0$, $p < .01$, with subjects as the random variable, and $F(2,118) = 4.4$, $p < .02$, with materials as the random variable. Neither the length of the paragraph nor the interaction of length and final sentence were significant (all $F < 1.4$). However, length did appear to affect error rate; there were 12% errors with short paragraphs and 17% errors with long paragraphs.

This experiment was designed to test several predictions. First, if the final sentence mentioned the critical word (*burglar*), then old propositions about the critical word should be activated; that is, the test word *garage* should be activated. Therefore, response time to *garage* should be faster when the final sentence mentioned the critical word than when it mentioned some unrelated word (*cat*). Second, a final sentence mentioning the category of the critical word should produce the same result. These predictions describe the data exactly; mean response times for critical word and category final sentences were the same and faster than mean response time for the unrelated word final sentences ($F(1,70) = 9.3$, $p < .01$, subjects analysis, and $F(1,118) = 7.2$, $p < .01$, materials analysis).

TABLE 3
RESULTS FROM EXPERIMENT 2

	Final sentence reading time[a] (msec)	Test word (garage): Response time[b] (msec)	Test word (garage): Proportion of errors
Final sentence			
burglar	2514	797	.13
criminal	2510	796	.15
cat	2753	832	.15

[a] Average standard error was 169 msec.
[b] Average standard error was 28 msec.

In general, the results for reading times of the final sentences followed those obtained in Experiment 1. Which version of the final sentence was read affected reading time, $F(2,70) = 9.1, p < .001$, subjects analysis, and $F(2,118) = 2.2, p = .12$, materials analysis. The final sentence of a longer paragraph was read faster (2749 milliseconds versus 2436 milliseconds), $F(1,35) = 8.7, p < .01$, subjects analysis, and $F(1,59) = 18.8, p < .001$, materials analysis. Final sentence and length did not interact, all $F < 1$.

One of the problems for reading time measures mentioned in the introduction may be a problem for the recognition latency measures used in Experiments 1 and 2. A spillover effect may be present in these experiments because the sentence that immediately preceded the test word was different in different experimental conditions. Response times to the test word were correlated with reading times for the final sentences; if the final sentence contained a word repeated from earlier in the text or an anaphor of a word mentioned earlier in the text, then reading time and response time to the test word were both faster than if the final sentence mentioned a new word. To rule out the possibility that spillover effects were responsible for the results of Experiments 1 and 2, we performed a simple control experiment. It was the same as Experiment 1, except that only the long versions of the paragraphs were used and the test word was from one of the middle two sentences of the paragraph (e.g., *curb*, in Table 1). We found that response time for the test word was the same regardless of whether the final sentence mentioned a repeated word, an anaphor, or a new word. Thus we can conclude that spillover effects were not responsible for the results of Experiments 1 and 2.

Experiment 1 showed that the referent of an anaphor is activated upon processing of a sentence containing the anaphor. Experiment 2 goes one step further and shows that a concept in the same proposition as the referent is also activated. It should be noted that the process of activating concepts in the propositions of the referent is a process that could not be investigated using other methodologies such as cued recall and reading time.

Experiment 3

The third experiment was designed to investigate the structure in memory that results from the processes of anaphoric reference. Proposed theories (Clark, 1978; Kintsch & Vipond, 1978) assume that the propositions of the referent and the propositions of the anaphor are connected together and stored in long-term memory during the comprehension process. For example, in the paragraph in Table 1, the final sentence says that someone slipped away from the streetlamp. When that someone is either *burglar* or *criminal*, then the information that *the burglar slipped away from the streetlamp* should be stored in long-term memory, connected to the information that *the burglar surveyed the garage*. To find out if information is indeed connected in this way, we used a priming technique in Experiment 3. A study–test procedure was used. On each trial, subjects studied two paragraphs and then were tested for recognition of single words. The extent to which two concepts were connected in the memory representation of a paragraph was measured by the amount of priming between them when they were presented sequentially in the test list. For example, *burglar* should prime *streetlamp* more when the last sentence mentions either *burglar* or *criminal* than when it mentions *cat*. However, *streetlamp* should be primed to some extent by *burglar* even when the last sentence mentions *cat*, because *burglar* and *streetlamp* are in the same paragraph, although far apart. Included in Experiment 3 was a condition where *streetlamp* was unprimed, that is, the immediately preceding word in the test list was a word from the other paragraph in the study list. Because the length of the paragraphs had not proved an interesting variable in Experiments 1 and 2, it was not included in Experiment 3; only the short versions of the paragraphs were used.

Method

Subjects. The subjects were 20 Dartmouth undergraduates who participated in the experiment for extra credit in an introductory psychology course.

Materials. The 60 experimental paragraphs of Experiment 1 and 2 were used. There were three positive test words for each paragraph. One was the target word, a noun in the final sentence of the paragraph that was in the same proposition as the critical word when the final sentence mentioned the critical word; for the example in Table 1, the target word was *streetlamp.* The second positive word was the critical word (e.g., *burglar*), and the third was a noun from the first sentence of the paragraph (not the critical word). For each paragraph there were also two words that did not appear in the paragraph; these were to be used as negative test words.

Procedure. A study–test recognition memory procedure was used. Each subject received 3 trials for practice and 30 experimental trials.

The study list for each trial was made up of two paragraphs chosen randomly without replacement from the 60 paragraphs. The test list for each trial consisted of 10 words, 6 positive and 4 negative. The test list was constructed in the following manner: First, the target words (one for each studied paragraph) were placed in randomly chosen positions in the list, but not in positions 1 or 2. Then, if the word was to be primed by the critical word of the same paragraph, this critical word was placed in the immediately preceding test position. If the word was to be unprimed, a word from the other studied paragraph was placed in the immediately preceding test position. Finally, the remaining positive test words and the negative test words were placed randomly in the remaining positions of the test list. Restrictions were that a word from one of the studied paragraphs could not precede the target word of that paragraph by fewer than three positions and that no word could appear more than once in the test list.

Subjects began each trial by pressing the space bar of the CRT keyboard. The first paragraph was then presented for 8 seconds, then it disappeared from the screen, then the second paragraph was presented for 8 seconds, and then it disappeared from the screen. Then the test list began immediately. The test words were presented one at a time. Each remained on the screen until the subject made a response, pressing a "yes" key if the word had appeared in either of the studied paragraphs, a "no" key otherwise. Subjects were instructed to respond as quickly and accurately as possible. If the response to a test word were correct, then the next test word appeared after 150 milliseconds. If the response were incorrect, then the word *ERROR* appeared on the screen for 2 seconds before the next test word appeared.

A sentence recognition test like that used in Experiments 1 and 2 was given after every five trials.

Design. There were four experimental conditions. The target was primed by the critical word (*streetlamp* primed by *burglar*) when the final sentence mentioned the critical word, the category of the critical word, or an unrelated word; or the target was unprimed (preceded by a word from the other studied paragraph). If the target were unprimed, then the final sentence was chosen randomly from the three versions. These four conditions were combined with four groups of subjects (5 per group) and four sets of paragraphs (15 per set) in a Latin-square design. A different random order of presentation of materials was used for every two subjects.

Results and Discussion

All response times longer than 2500 milliseconds were eliminated from the analyses. Only correct responses preceded by correct responses were included in the analyses in order to be as sure as possible that both the priming and target words were in memory. All analyses were based on mean response times for each subject or each item in each condition.

Mean response time for the target word was significantly affected by experimental condition, $F(3,57) = 12.4$, $p < .001$, with subjects as the random variable, and $F(3,168) = 10.2$, $p < .001$, with materials as the random variable. In the final sentence that mentions *burglar*, *burglar* and *streetlamp* are in the same proposition, so *streetlamp* should be maximally primed. The mean response time to *streetlamp* in this condition was 648 milliseconds (12% errors). If the final sentence that mentions the category *criminal* serves to activate *burglar*, and if the new propositions of the final sentence are connected to *burglar*, then *streetlamp* should be primed as much when the final sentence mentions *criminal* as when it mentions *burglar*. The mean response time to the target word when the final sentence mentioned the category was 644 milliseconds (9% errors). If the final sentence mentions an unrelated word, then the critical word and the target word should not be directly connected in the memory representation. Thus, mean response time in this condition should be slower than in the conditions where the final sentence mentions either the critical word or the category. This result was found; mean response time to the target word in the condition where the final sentence mentioned an unrelated word was 686 milliseconds (11% errors). This was significantly slower than the category final sentence condition, $F(1,57) = 5.4$, $p < .03$ (subjects analysis), and $F(1,168) = 5.4$, $p < .03$ (materials analysis). Finally, response times should be slowest of all in the condition when the target was unprimed, that is, preceded in the test list by a word from the other paragraph; mean response time in this condition was 741 milliseconds (18% errors). Average standard error of the means over all four conditions was 30 milliseconds.

Experiment 4

We would like to interpret the results of Experiment 3 in terms of inference processes. The anaphor *criminal* is inferred to be a reference to the *burglar* and so the information about *slipping away from the streetlamp* is connected to the *burglar*. Thus, *burglar* primes *streetlamp* equally whether *criminal* or *burglar* is mentioned with *streetlamp* in the final sentence of the paragraph. However, there is an alternative interpretation of the results of Experiment 3. According to this interpretation, the anaphor *criminal* does not lead to a connection between *burglar* and *streetlamp*. Instead, the priming results in the condition where the final sentence mentions *criminal* are explained by assuming that presentation of *burglar* in the test list primes the concept *criminal* through preexperimental semantic association. *Criminal*, in turn, primes *streetlamp*. Thus, *burglar* primes *streetlamp* even though the two concepts are not directly connected in the memory representation of the paragraph.

Experiment 4 was designed to rule out this interpretation by demonstrating that preexperimental semantic association alone, without anaphoric reference, does not lead to the pattern of priming effects found in Experiment 3. The sentences of the paragraphs of Experiment 3 were presented to subjects in such a way that the category names (e.g., *criminal*) would not be perceived as anaphors. This was accomplished by mixing the sentences of the paragraphs into lists of other, unrelated sentences. On each trial of the experiment, the subject studied a list of such sentences and was then tested for recognition of words from the sentences. There were two experimental conditions: a paragraph was represented in a study list either by the first sentence and the final sentence that mentioned the category name or by the first sentence and the final sentence that mentioned the unrelated word. For example, the sentences for the *burglar* paragraph in the first condition were *A burglar surveyed the garage set back from the street* and *A criminal slipped away from the streetlamp* and in the second condition, *A burglar surveyed the garage set back from the street* and *A cat slipped away from the streetlamp*. In the test list for both conditions, *streetlamp* was immediately preceded by *burglar*. If *burglar* primes *criminal* by preexperimental semantic association, then response time for *streetlamp* should be faster in the first condition than in the second. On the other hand, if the priming results in Experiment 3 depended on anaphoric reference, then response time for *streetlamp* should be the same in the two conditions of Experiment 4.

Method

Subjects. The subjects were 10 Dartmouth undergraduates from the same population as Experiment 3.

Materials. The 60 paragraphs of the previous three experiments were used. For each paragraph, the experimental design involved the first sentence and two of the possible fourth (final) sentences, the fourth sentence that mentioned the category name and the fourth sentence that mentioned an unrelated word (see Table 1). The second and third sentences of a paragraph were used as filler sentences in the following way. The 60 paragraphs were divided into two sets. The first and fourth sentences of one set were combined with the second and third sentences of the other set, and vice versa. Thus, the first and fourth sentences were combined with second and third sentences unrelated in meaning.

For the first and fourth sentences of a paragraph, there were two positive test words, the critical word of the first sentence (e.g., *burglar*) and the noun of the fourth sentences that was used in Experiment 3 (e.g., *streetlamp*). For each of the second and third sentences, there was one positive test word, a noun.

Procedure. A study–test procedure similar to that of Experiment 3 was used. Each subject received 2 trials for practice and 30 experimental trials.

The study list for each trial consisted of eight sentences, the first and fourth sentences of two paragraphs, and the second and third sentences of two other paragraphs. These eight sentences were presented in random order with the restriction that two sentences of the same paragraph were separated by at least three other sentences. The test list for each trial consisted of 12 words, 8 positive and 4 negative. The test list was constructed in the following manner. First, for each fourth sentence, the target word (e.g., *streetlamp*) was placed in a randomly chosen position in the list, except in positions 1 or 2. Then, the priming word, the critical word of the first sentence (e.g., *burglar*), was placed in the immediately preceding test position. Finally, the positive test words of the second and third sentences and four negative test words were placed randomly in the remaining test positions.

Subjects began each trial by pressing the space bar of the CRT keyboard. Then the eight sentences of the study list were presented one at a time for 4 seconds each, and then the test list began immediately. Presentation of the test words was the same as in Experiment 3.

Design. There were two experimental conditions; the fourth sentence of a paragraph was presented in either the category name version (e.g., *criminal*) or the unrelated word version (e.g., *cat*). The first sentence of a paragraph was always presented, and the target word of the fourth sentence (e.g., *streetlamp*) was always immediately preceded in the test list by the critical word from the first sentence (e.g., *burglar*). The two experimental conditions were combined with two groups of subjects (5 per group) and two sets of paragraphs (30 per set) in a Latin-square design. Order of presentation of materials was rerandomized after every second subject.

Results and Discussion

The data were analyzed as in Experiment 3. Mean response time for the target test word (e.g., *streetlamp*) in the first condition where the fourth sentence mentioned the category name (e.g., *criminal*) was 590 milliseconds (12% errors) and in the second condition where the fourth sentence mentioned the unrelated word (e.g., *cat*), 595 milliseconds (11% errors). Average standard error of these means was 29 milliseconds. It should be noted that the number of observations per experimental condition was the same in this experiment as in Experiment 3, so the power of this experiment is the same as that of Experiment 3.

These results rule out the alternative interpretation of Experiment 3. The preexperimental semantic association between *burglar* and *criminal* does not lead to priming between *burglar* and *streetlamp* when the word *criminal* is not perceived as an anaphor for *burglar*. Thus, the results of Experiment 3 demonstrate that *criminal* does function as an anaphor, connecting information about the *burglar* mentioned in the first sentence to information about the *criminal* mentioned in the fourth sentence.

General Discussion

In this paper, we have presented data that lend strong support to a model of simple inference processes (Clark, 1978; Kintsch & Vipond, 1978). The model assumes that inference involves three component processes. First, a concept to be inferred has to be accessed in memory. Second, the concept, along with its associated propositions, has to be activated (i.e., brought into working memory). Third, the information responsible for the activation has to be connected to the concept and its associated propositions and then the result stored in long-term memory. Experiment 1 examined the second component; the results showed that an anaphor activated its referent and thus supports the findings of Chang (1980). Experiment 2 also examined the second component and showed that an anaphor activates concepts in propositions associated with the referent. Experiment 3 showed the operation of the third component. Information that caused activation of a referent was connected to the referent in long-term memory.

The experiments reported in this paper used relatively new methodologies to investigate the processes of anaphoric reference. Both methodologies involve testing a single word for recognition. In Experiments 1 and 2, subjects were presented with a text which they read sentence by sentence. Immediately after the last sentence, a test word, which had appeared in an earlier sentence of the text, was presented for recognition. When the last sentence mentioned an anaphor of the test word or an anaphor of a word closely associated in the text to the test word, then the response to the test word was speeded. This result is interpreted as showing activation of the referent of the anaphor during comprehension of the last sentence. In Experiment 3, subjects studied two paragraphs and then were presented with a test list of single words for recognition. A response to a test word was speeded (a priming effect) when the test word was preceded in the test list by another test word that was closely associated in the preceding text. Results showed that this priming effect was equally large whether the two words were in the same sentence of the text or in different sentences and connected by inference processes.

With these priming techniques, it is relatively easy to avoid problems in interpretation of results. One problem that might arise with the activation technique of Experiments 1 and 2 is a spillover effect; processing time for the sentence immediately preceding the test word may differ across experimental conditions, and this processing time difference may lead to differences in response time to the test word. However, a spillover effect can easily be detected by measuring response time for a test word that should not vary in activation level across experimental conditions. With the primary technique used in Experiment 3, it is important to show that any priming effects obtained reflect the representation in memory of the text and not preexperimental semantic association. Experiment 4 exemplifies such a demonstration (see also McKoon & Ratcliff, 1980).

We argue that priming methods avoid the problems of other currently popular methods. For example, with cued recall, inference processes cannot be studied "on-line." Experimental results that could be attributed to inferences made during encoding can just as well be attributed to inferences made at retrieval. In fact, recent evidence supports the latter view (Corbett & Dosher, 1978; Singer, Note 1). To study inference processes "on-line," experimenters have used another technique, measuring the time a subject takes to read individual sentences. The assumption is that comprehension of a sentence that requires an inference will take longer than comprehension of a sentence that does not require an inference. To show unambiguously that increases in reading time are due to inference processes and not to uninteresting confounding variables is difficult. If a sentence that is supposed to require an inference does not repeat words from earlier in the text while a control sentence supposed not to require an inference does repeat words, then differences in reading time may be due to simple repetition effects. A more common problem is that texts designed to require inference processing are less comprehensible than texts designed to require no inference processing. In this case, differences in reading time may be due to the time it takes subjects to judge that a text is difficult to comprehend and perhaps decide not to try. There is still a further problem in the design of reading time experiments, and that is what can be called "spillover" or sequential effects. In the usual reading time experiment, a target sentence is held constant across experimental conditions and preceding context sentences are varied. Differences in reading times for the context sentences can carry over to affect reading times for the target sentences (Haberlandt & Bingham, 1978). Finally, even if these design problems are solved, reading time is a measure with limited usefulness because it can only indicate when increased processing is required and cannot indicate the components of that processing.

In this paper, we have introduced techniques that can avoid these problems. The results of the experiments support theories of inference processing, but more than that, illustrate techniques that allow the investigation of the component processes involved in inference.

References

Anderson, J., & Hastie, R. Individuation and reference in memory: Proper names and definite descriptions. *Cognitive Psychology*, 1974, 6, 495–514.

Anderson, R. C., Goetz, E. T., Pichert, J. W., & Halff, H. M. Two faces of the conceptual peg hypothesis. *Journal of Experimental Psychology: Human Learning and Memory*, 1977, 3, 142–149.

Anderson, R. C., & Ortony, A. On putting apples into bottles—a problem of polysemy. *Cognitive Psychology*, 1975, 7, 167–180.

Anderson, R. C., Pichert, J. W., Goetz, E. T., Schallert, D. L., Stevens, K. V., & Trollip, S. R. Instantiation of general terms. *Journal of Verbal Learning and Verbal Behavior*, 1976, 15, 667–679.

Barclay, J., Bransford, J., Franks, J., McCarrell, N., & Nitsch, K. Comprehension and semantic flexibility. *Journal of Verbal Learning and Verbal Behavior*, 1974, 13, 471–481.

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

Caramazza, A., Grober, E., & Garvey, C. Comprehension of anaphoric pronouns. *Journal of Verbal Learning and Verbal Behavior*, 1977, 16, 601–609.

Carpenter, P., & Just, M. Reading comprehension as eyes see it. In M. Just & P. Carpenter, (Eds.), *Cognitive processes in comprehension*. Hillsdale, N.J.: Erlbaum, 1977.

Chafe, W. L. Discourse structure and human knowledge. In J. B. Carrol & R. O. Freedle (Eds.), *Language comprehension and the acquisition of knowledge*. Washington, D.C.: Winston, 1972.

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

CLARK, H. Inferences in comprehension. In D. LaBerge & S. Samuels (Eds.) *Basic processes in reading, perception and comprehension*. Hillsdale, N.J.: Erlbaum, 1978.

CLARK, H., & SENGAL, C. In search of referents for nouns and pronouns. *Memory and Cognition*, 1979, 7, 35–41.

CORBETT, A., & DOSHER, B. Instrument inferences in sentence encoding. *Journal of Verbal Learning and Verbal Behavior*, 1978, 17, 479–491.

GARROD, S., & SANFORD, A. Interpreting anaphoric relations: the integration of semantic information while reading. *Journal of Verbal Learning and Verbal Behavior*, 1977, 16, 77–90.

HABERLANDT, K., & BINGHAM, G. Verbs contribute to the coherence of brief narratives: reading related and unrelated sentence triples. *Journal of Verbal Learning and Verbal Behavior*, 1978, 17, 419–425.

HAVILAND, S., & CLARK, H. What's New? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior*, 1974, 13, 512–521.

KEENAN, J. M., & KINTSCH, W. The identification of explicitly presented information. In W. Kintsch (Ed.), *The representation of meaning in memory*. Hillsdale, N.J.: Erlbaum, 1974.

KINTSCH, W., KOZMINSKY, L., STREBY, W., MCKOON, G., & KEENAN, J. Comprehension and recall of text as a function of content variables. *Journal of Verbal Learning and Verbal Behavior*, 1975, 14, 196–214.

KINTSCH, W., & VIPOND, D. Reading comprehension and readability in educational practice and psychological theory. In L. G. Nilsson (Ed.), *Memory: Processes and problems*. Hillsdale, N.J.: Erlbaum, 1978.

LESGOLD, A., ROTH, S., & CURTIS, M. Foregrounding effects in discourse comprehension. *Journal of Verbal Learning and Verbal Behavior*, 1979, 18, 291–308.

MCKOON, G., & KEENAN, J. M. Response latencies to explicit and implicit statements. In W. Kintsch (Ed.), *The representation of meaning in memory*. Hillsdale, N.J.: Erlbaum, 1974.

MCKOON, G., & RATCLIFF, R. Priming in item recognition: The organization of propositions in memory for text. *Journal of Learning and Verbal Behavior*, 1980, 19, 369–386.

PARIS, S. G., & LINDAUER, B. K. The role of inference in children's comprehension and memory for sentences. *Cognitive Psychology*, 1976, 8, 217–224.

POSNER, M., & SNYDER, C. Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition*. Hillsdale, N.J.: Erlbaum, 1975.

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

REDER, L. The role of elaborations in memory for prose. *Cognitive Psychology*, 1979, 11, 221–234.

SANFORD, A. J., & GARROD, S. Memory and attention in text comprehension: The problem of reference. In R. S. Nickerson (Ed.), *Attention and performance VIII*. Hillsdale, N.J.: 1980.

SINGER, M. Processes of inference during sentence encoding. *Memory and Cognition*, 1979, 7, 192–200.

TILL, R. E. Sentence memory prompted with inferential recall cues. *Journal of Experimental Psychology: Human Learning and Memory*, 1977, 3, 129–141.

TULVING, E. Ecphoric processes in recall and recognition. In J. Brown (Ed.), *Recall and recognition*. London: Wiley, 1976.

YEKOVICH, F., WALKER, C., & BLACKMAN, H. The role of presupposed and focal information in integrating sentences. *Journal of Verbal Learning and Verbal Behavior*, 1979, 18, 535–548.

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

REFERENCE NOTES

1. SINGER, M. *The role of explicit and implicit recall cues*. Presented in a symposium titled, Implication and Inference in Language Comprehension, at the Annual Meeting of the American Psychological Association, Toronto, August 1978.
2. KEENAN, J. M. *Inferring causal connections in prose comprehension*. Invited paper in a symposium titled, Implication and Inference in Language Comprehension, at the Annual Meeting of the American Psychological Association, Toronto, August 1978.
3. MCKOON, G., & RATCLIFF, R. *The comprehension processes and memory structures involved in instrumental inference*. Submitted for publication.
4. RATCLIFF, R., & MCKOON, G. *Automatic and strategic priming in recognition*. Submitted for publication.

(Received April 15, 1980)