

Similarity Information versus Relational Information: Differences in the Time Course of Retrieval

ROGER RATCLIFF AND GAIL MCKOON

Northwestern University

Two experiments are reported that examine the time course of retrieval in a sentence matching procedure. Subjects learned lists of active and passive sentences and were tested with sentences in active or passive, correct or incorrect versions; for example, if "John hit Bill" was a studied sentence, "Bill hit John" would be an incorrect active test sentence. A response signal procedure was used so that accuracy could be measured as a function of time. The data show that sentences containing words from studied sentences are discriminable early in processing from sentences containing all new words, but discrimination of correct from incorrect versions of studied sentences occurs only later in processing (after 600-700 ms). These results demonstrate that different kinds of information are available at different points during the time course of retrieval and so suggest that modifications are required of models that provide only a unitary value for the amount of match between a test probe and information in memory. Early in processing, the growth of accuracy can be explained by a simple model that assumes independent contributions to total amount of match for each of the content words of a sentence, but this independent processing model cannot account for discrimination later in processing. Several, more general, memory models are examined with respect to their abilities to produce independent item information early in processing and relational information later in processing. © 1989 Academic Press, Inc.

It is reasonable to suppose that all of the different kinds of information about the answer to a question do not become available at the same point in the time course of retrieval. Data that speak to this issue are fundamentally important because they constrain theories of retrieval. For example, current theories of memory and retrieval allow information to become available in two ways, via a fast matching process that gives an overall goodness of match (for example, in a recognition procedure) and via a slower item-to-item serial retrieval process (e.g., Gillund & Shiffrin, 1984; Murdock, 1982, 1983). The aim of this article is to add to the literature a set of results that demonstrate another aspect of the time course

This research was supported by NSF Grant BNS 8510361 to Roger Ratcliff, NIH Grant HD18812 to Gail McKoon, and NSF Grant BNS 85 16350 to Gail McKoon. We thank Barbara Doshier and Richard Shiffrin for their useful comments on this article. Reprint requests should be sent to Roger Ratcliff, Psychology Department, Northwestern University, Evanston, IL 60208.

of retrieval. Specifically, evidence is presented that similarity information is available earlier during the course of processing than relational information.

The procedure used in the experiments is the response signal procedure. When presented with a test item, subjects are required to respond at one of several experimenter-determined time signals or deadlines (Corbett & Wickelgren, 1978; Doshier, 1976, 1982, 1984; Ratcliff, 1981; Ratcliff & McKoon, 1982; Reed, 1973; 1976; Wickelgren, 1977). Typically, subjects are presented with a test question and then when the signal is presented, they are required to respond within 200–300 ms. The dependent measure is accuracy as a function of time. When the signal is presented immediately after the test question, accuracy is expected to be at chance. At a relatively long delay between test question and signal, accuracy should asymptote.

There are several examples in the literature where the response signal procedure has been used to show differences in the availability of different kinds of information. In the clearest cases, two kinds of information are pitted against each other, for example, when the similarity of two concepts is high but the relationship between them dictates a negative response. Ratcliff and McKoon (1982) had subjects answer questions like “is a robin a bird” (“yes”) and “is a bird a robin” (“no,” because not all birds are robins). For the negative questions (“is a bird a robin”), probability of an incorrect “yes” response rose over the 600 ms of processing and then dropped, producing a nonmonotonic function. This result was interpreted as showing that information about the semantic relationship between the two concepts (“bird” and “robin”) was available early in processing but that information about the ordered relationship expressed by the question was available only later in processing.

Examining a related issue, Doshier (1984) taught subjects pairs of words such as “king–pear” and “paper–queen.” At test time, subjects were required to judge whether the words in a test pair had been studied together or not. If a test pair was rearranged with respect to a study pair so that the test pair was semantically related (e.g., “king–queen”), then the probability of responding positively (in error) increased early in processing, decreasing toward correct negative responses only later in processing. Doshier interpreted this as showing the initial involvement and later suppression of semantic information in retrieval processing.

The studies presented in this article examined the availability of information that discriminates relationally correct test sentences from highly similar but relationally incorrect test sentences. Subjects studied short sentences presented in either active or passive form. They were required to respond whether test sentences were the same in meaning as studied sentences. For example, if “John hit Bill” was studied, then the

correct responses to "John hit Bill" and "Bill was hit by John" were "yes" and the correct responses to "Bill hit John" and "John was hit by Bill" were "no." These test sentences are all highly similar to the studied sentence: There were also test sentences that were dissimilar to study sentences, sentences in which the content words had not appeared in any studied sentence (correct response, "no"). Given these different kinds of test sentences, the point in the time course of processing at which relational meaning ("who hit whom") becomes available can be measured and compared to the point at which similarity information becomes available.

Anderson (1983) in his ACT* model has explicitly examined the processes of matching test sentences against studied sentences with respect to words in correct order versus words in reversed order. He assumes that information about "connectivity" among concepts (whether the words were in the same sentence) becomes available initially, and then later in processing, productions are applied that discriminate correct from incorrect versions of the sentences. This model, as well as other memory models, will be considered in detail later in this article.

EXPERIMENT 1

There is an important design issue concerning repetition of materials that arises in applying the response signal procedure. With some materials, it is quite reasonable to repeat items between sessions, because information gained from one presentation will not affect performance on a subsequent presentation. However, with our pool of sentences, considerable learning can take place. We first ran a multisession procedure with the same set of sentences repeated in each session and found that subjects developed strategies for encoding study sentences in anticipation of potential test sentences. To avoid this problem with repetition of materials, we changed to a procedure in which we collected data from only one session per subject. This session was preceded by two practice sessions on other materials.

In the first and second practice sessions, subjects were taught the response signal procedure with sentences they knew to be true or false with respect to general knowledge (e.g., "a mother is a father"). In the second half of the second practice session, they were switched to practice study test procedure in which, for each trial, four sentences were presented for study and five sentences were tested. In the third session, data were collected with the study-test procedure and a new set of sentences.

The conditions of the experiment are shown in Table 1. Both active and passive study and test sentences were used so that word order did not determine the correct response. To summarize the table, test sentences were related to study sentences in that they used the same content words

and were the same in meaning (Conditions 1, 2, 5, and 6), they used the same content words but were not the same in meaning (Conditions 3, 4, 7, and 8), or they used different words altogether (Conditions 9 and 10). All of the conditions in which the content words were the same involve high similarity between study sentence and test sentence. Only the relational information discriminates which of these require positive and which negative responses.

Method

Subjects. Each of the 37 subjects participated in three 1-h sessions. The first two sessions gave practice on the response signal procedure, and the data reported in the results section below were collected only in the third session. Some of the subjects participated in exchange for credit in a psychology course, and some were paid \$5.00 per session.

Materials. For the practice test lists, sentences were used that were true or false according to general knowledge (these were taken from sentences used by Ratcliff & McKoon, 1982). There were 600 different sentences, half true and half false, that averaged four words in length.

For the study-test trials, in which subjects had to study sentences for later test, there were 360 sets of sentences. Each set was made up from two noun phrases and one verb, combined into the four possible active and passive sentences as shown in Table 1. Some of the nouns were proper names and so were one word in length; others were an article plus a noun. No noun phrases or verbs were used in more than one sentence set.

There were also 120 sentences used for test sentences that contained no words from study sentences and so required a negative response.

Procedure. Stimuli were presented on the screens of CRT terminals, and subjects made their responses on the CRT keyboards. The terminals were controlled by a real-time microcomputer system.

In the first of the three sessions, subjects practiced the response signal procedure on three separate test lists. The first list was made up of 200 true/false sentences from general knowledge. At the beginning of the list, an instruction was displayed on the CRT screen to press the space bar of the keyboard. When the space bar was pressed, there was a 250-ms pause, and then the test items were presented one at a time. Each item began with a warning signal, a row of "+" signs displayed for 250 ms. Then the "+"s were erased and a test sentence appeared immediately below where the "+"s had been. At a variable time after the sentence appeared, it was erased and a row of asterisks was displayed immediately below where the sentence had been. The lag from sentence to asterisks was 50, 150, 250, 400, 800, or 2000 ms. The asterisks were the cue to give a response, pressing the "?" key for *true* and the "Z" key for *false*. After the response, the asterisks were erased and, on the line below where they had been, the response time (from presentation of asterisks to key press) was displayed. The time remained on the screen for 550 ms; then the screen was erased and the warning signal for the next test item appeared. After every 20 test items, subjects were given feedback on their performance. They were told their average speed and number of errors on the fastest two lags (combined) and their speed and number of errors on the slowest two lags (combined). This information was displayed for 10 s, and then the instruction to press the space bar to begin more tests was presented.

The second test list was identical to the first, except that on the first list, the fastest two lags were not used in the first 20 test items. In the third list, items were presented in the same way as in the first two lists, first two lists, but there were 400 true/false general knowledge

Subjects were instructed to respond as quickly as they could when the asterisks appeared, at least within 300 ms. They were told that, although they should try to do their best, speed was much more important than accuracy. They were told that performance at the shortest lags was expected to be at chance, but at the longest lags it should be close to perfect.

The second session of practice began with a list of 200 true/false general knowledge test sentences (200 of the last 400 given in session 1), presented in the same way as the lists in the first session. Then subjects were given instructions about the study-test procedure that would be used in the remainder of the experiment and a practice list of 30 study-test trials.

Each study-test trial began with the instruction to press the space bar of the keyboard, and then four sentences were given for study and five sentences were tested. The study sentences were presented one at a time for 4.5 s each. After the fourth one, the words *TEST TEST TEST* were displayed for 1.5 s as a warning that the test list was about to begin. Then each test item was presented in the same way as in the practice sessions, with warning signal (" + "s), sentence, asterisks, and response time (displayed for only 400 ms instead of 550 ms). After the fifth test sentence, the only feedback given was accuracy at the longest two lags (combined). This measure of accuracy was accumulated across trials, so that at the end of each trial, subjects could evaluate their performance up to that point in the experiment.

With respect to speed and accuracy, subjects were given the same instructions as for the practice on general knowledge test sentences. They were also told about the different kinds of test sentences they would be given and told to respond according to whether the meaning of a test sentence matched the meaning of a studied sentence, and not according to whether the wording matched exactly.

In the third session, the data to be reported in the results section below were collected. Subjects were given 60 study-test trials, in the same way as the practice in the previous session. They were told that they would be paid an extra \$2.00 for their participation if their speed on the shortest lags averaged less than 300 ms and their accuracy on the longest lags was higher than 85% correct, over the whole session.

Design. Eight of the conditions in the experiment were made up of all possible combinations of active and passive study and test sentences, as shown in Table 1. The correct response for half of these was "yes," and for the other half, "no." In the other two conditions of the experiment, the test sentences were new in that none of the nouns and

TABLE 1
Conditions in Experiments 1 and 2

Study	Test	Correct response	1 (p)	2 (p)
1. AA+ Helen attracted Jeff	Helen attracted Jeff	Yes	.5	.5
2. AP+ Helen attracted Jeff	Jeff was attracted by Helen	Yes	.5	.5
3. AA- Helen attracted Jeff	Jeff attracted Helen	No	.5	.25
4. AP- Helen attracted Jeff	Helen was attracted by Jeff	No	.5	.25
5. PA+ Jeff was attracted by Helen	Helen attracted Jeff	Yes	.5	.5
6. PP+ Jeff was attracted by Helen	Jeff was attracted by Helen	Yes	.5	.5
7. PA- Jeff was attracted by Helen	Jeff attracted Helen	No	.5	.25
8. PP- Jeff was attracted by Helen	Helen was attracted by Jeff	No	.5	.25
9. —	Andrew accosted Mary	No	.5	.5
10. —	Mary was accosted by Andrew	No	.5	.5

Note. AA+ refers to an active study sentence and a matching (+) active test sentence. P refers to passive sentences. 1: This is the probability that any one test sentence appeared in this condition in Experiment 1, multiplied by 5. (The total of the numbers in the column is 5, representing the five test sentences on each trial.) 2: This is the probability that any one test sentence appeared in this condition in Experiment 2, multiplied by 6. (The total of the numbers in the column is 4, representing four of the six test sentences on each trial.)

verbs contained in them had been presented in any study list. In one of these conditions, the sentence was active; in the other, passive. The 10 conditions just described were combined with the six response signal lags to give a total of 60 experimental conditions. Sentence sets for the study lists and conditions to be used in the test lists were chosen randomly for each trial, and order of presentation was random. No sentence was repeated in a test list, and no sentence set was used more than once in the experiment.

Results

The main results are shown in Figs. 1 through 4. d' values are computed by pooling responses across subjects rather than computing d' s for each subject and then averaging. Note that the time values represent the average response times added to the lag from presentation of the test sentence to presentation of the response signal. Figures 1 and 2 show the d' values for Conditions 1 through 8 in Table 1 scaled against the appropriate new, dissimilar negative condition (passive tests scaled against the passive negative, Condition 10, and active tests scaled against the active negative, Condition 9). The results are extremely clear cut. d' rises for both relationally correct and relationally incorrect versions of the test sentences at the same rate up to about 700 ms, at which point the correct and incorrect versions become discriminable. For active test sentences (Fig. 1), there is arguably a nonmonotonic lag function for the negative test sentences. For passive test sentences (Fig. 2), d' begins a little lower than for the active test sentences and the d' function for the negative test sentences is just about flat as a function of lag after 700 ms. Figures 3 and 4 show the probabilities of a "yes" response as a function of time. Essentially they show the same patterns as Figs. 1 and 2, with the data from relationally correct and incorrect versions of test sentences splitting apart at about 700 ms. The figures also show that the general increase in d' over

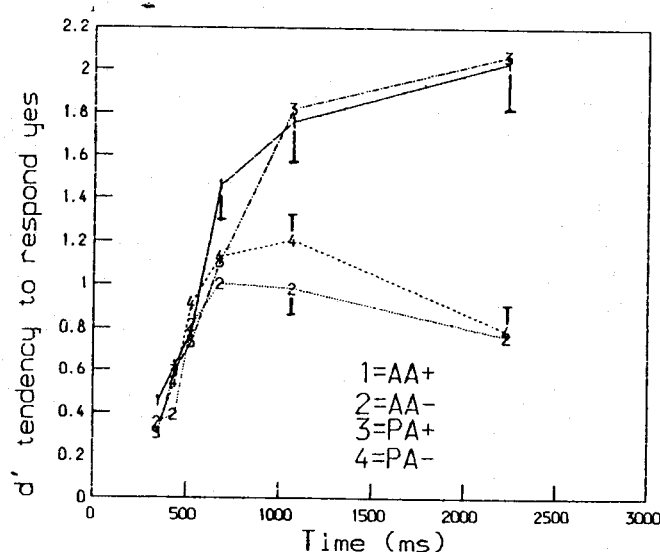


FIG. 1. d' as a function of time for the active test sentences in Experiment 1.

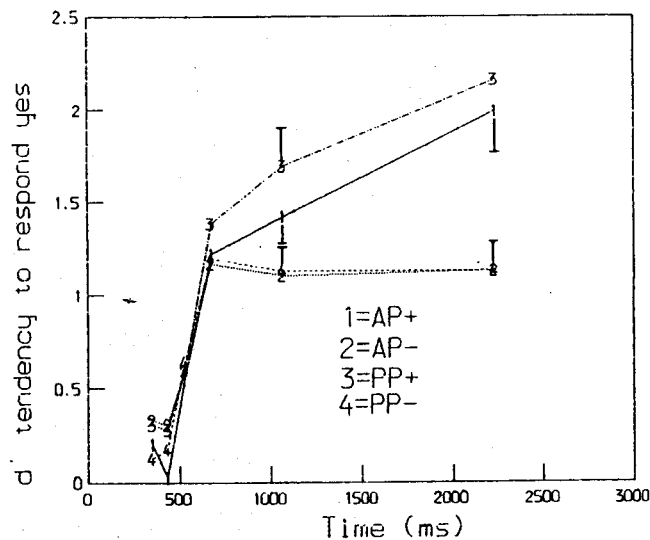


FIG. 2. d' as a function of time for the passive test sentences in Experiment 1.

time is attributable to a rapid increase in accuracy for the new sentences in Conditions 9 and 10.

In order to examine the variability in the data points in Figs. 1 through 4, we calculated standard errors in the probabilities of the responses at each lag for each condition (see also Ratcliff & McKoon, 1982). Standard errors in probabilities can be calculated from the standard formula for the standard error in binomial probability, $\sqrt{p(1-p)/N}$. There are typically 130–160 observations per condition so that the following equations show typical standard errors for several probability values: $p = .5 \pm .040$, $.6 \pm .039$, $.7 \pm .036$, Figs. 3 and 4 (note that for the negative sentences in Conditions 9 and 10, the standard error in p is the same as that in $1 - p$). For d' , typical values are $.3 \pm .08$, $.6 \pm .09$, $.8 \pm .10$, $1.5 \pm .15$, and

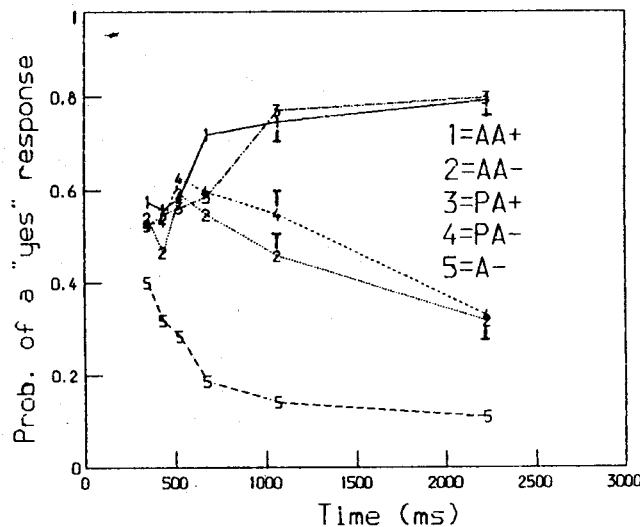


FIG. 3. Probability of a "yes" response for the active test sentences in Experiment 1.

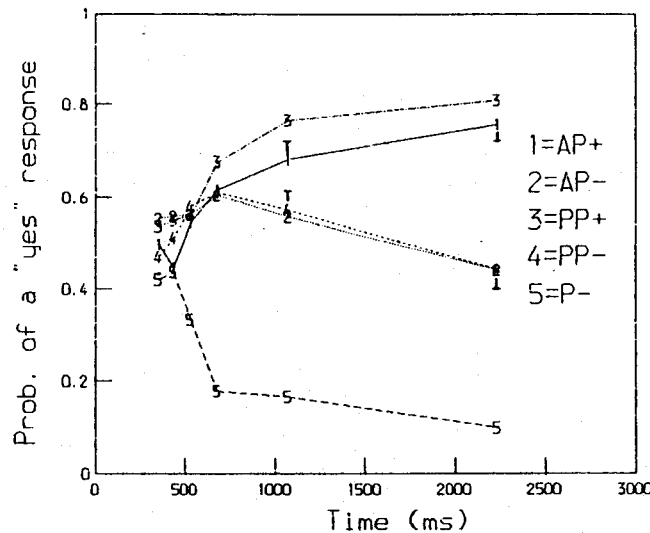


FIG. 4. Probability of a "yes" response for the passive test sentences in Experiment 1.

$2.0 \pm .2$ (but these values depend on the precise hit and false alarm rates; for example, for an N of 100, the standard error in a d' of 0 with hit and false alarm rates of .5 is .13, and for hit and false alarm rates of .95, the standard error is .26). Some typical standard errors for the values obtained in the experiment are shown as error bars on the figures (each bar represents 1 standard error) and these can be used to determine the significance of differences among conditions.

In Figs. 1 through 4, the curves do not differ significantly until after the point at which they separate (about 700 ms). After that point, the positive and negative curves appear significantly different, mostly by more than 2 standard errors. Before that point, when the curves overlap, the differences are well within 2 standard errors.

In sum, the curves in Figures 1 through 4 clearly show early discrimination between sentences with studied words and sentences with non-studied words and only late discrimination between correct and incorrect versions of studied sentences. In addition to these results, there are several other points to make about the data. One issue concerns the mental representation of the study sentences. Subjects could have transformed study sentences into a common form, for example, active. If this occurred, then the late discrimination for the active test sentences should occur at an earlier time than the late discrimination for the passive test sentences; however, this does not appear in the data. If, instead, the sentences were encoded in their surface form (active study sentences into an active representation and passive study sentences into a passive representation), then the late discrimination should be delayed when a transformation was required (e.g., from active study sentence to passive test sentence) relative to no transformation between study and test. This too

is not observed in the data. A third possibility is that individual subjects adopt different strategies, and these average out. The fastest processes would begin to produce divergence at about 700 ms, and slower processes would add to this divergence later. This possibility cannot be distinguished from the idea that there is a common representation for the sentences that is neither active nor passive; either of these last two possibilities is consistent with the data.

The last point to make about the data concerns the effect of word order. The materials were designed specifically so that word order was not predictive of the response, and there was little evidence that word order made a difference (although, perhaps the AP+ condition is slightly more difficult than PP+ (Figs. 2 and 4), and the PA- condition more difficult than AA- (Figs. 1 and 3)). This does not mean that word order is never important (see Ratcliff, 1987), only that it contributed little in this experiment.

EXPERIMENT 2

In Experiment 1, the sentences were reversible in that a change in the order of the subject and object produced a sentence that was sensible. In Experiment 2, irreversible sentences were added to the conditions of Experiment 1. For these sentences, reversing subject and object would produce a sentence that was not meaningful. These sentences allowed examination of the availability and possible use of this different kind of meaning information. Also added were test sentences which were the same as study sentences except that either the first noun or the verb was replaced by a new word (these were labeled the "single replace" conditions). These test sentences were designed to distinguish hypotheses about serial processing of words in test sentences (e.g., left to right) from hypotheses about parallel and independent processing of each word.

Method

Subjects. Each of 27 subjects participated in three 1-h sessions. They were either paid or given course credit, as in Experiment 1.

Materials. Sixty new sets of sentences were created such that the subjects and objects could not be reversed; an example is shown in Table 2. The other sentences were the same as those used in Experiment 1. No nouns or verbs from the sentence sets of Experiment 1 were used in the new sets created for Experiment 2.

There were also lists of nouns and verbs that did not appear in any sentences, 240 nouns and 120 verbs. These were used to create lures, by replacing the first noun or the verb of a studied sentence by one of these new words.

Procedure. The procedure was the same as in Experiment 1, except that there were five sentences in each study list and six sentences in each test list. One study sentence was always one of the irreversible sentences, and it was always used for one of the conditions in the test list. Subjects were given the same instructions, plus a description of the lures created by replacing one word of a studied sentence.

TABLE 2
Conditions in Experiment 2

Study	Test	Correct response	1 (p)
Replacement negatives			
11. Helen attracted Jeff	Betty attracted Jeff	No	.25
12. Helen attracted Jeff	Helen saw Jeff	No	.25
13. Helen was attracted by Jeff	Betty was attracted by Jeff	No	.25
14. Helen was attracted by Jeff	Helen was seen by Jeff	No	.25
Irreversible sentences			
15. Tony sewed the jacket	Tony sewed the jacket	Yes	.25
16. The jacket was sewn by Tony	The jacket was sewn by Tony	Yes	.25
17. Tony sewed the jacket	The jacket sewed Tony	No	.25
18. The jacket was sewn by Tony	Tony was sewn by the jacket	No	.25

Note. Conditions 1–10 in Table 1 were also included in this experiment. 1: This is the probability that any one test sentence appeared in this condition, multiplied by 6. (The total of the numbers in the column is 2, representing two of the six test sentences on each trial.)

Design. There were 18 different conditions in the experiment. The first 10 were the same as those in Experiment 1. In another four conditions, one of the words of a study sentence was replaced by a word that had not appeared in any study sentence. These test sentences appeared in the same form as in the study list; active in two conditions and passive in the other two. In two of these conditions, the replaced item was the first noun phrase and in the other two, it was the verb. The last four test conditions used the irreversible sentences. In two conditions, one active and one passive, the test sentence was exactly the same as the studied sentence. In the other two conditions (again, one active and one passive), the noun phrases were switched. Examples of all conditions are shown in Table 2, with the probability of presentation of each condition. Test items were constructed according to these probabilities and then assigned one of the six response signal lags randomly.

Results

The results are shown in Figs. 5 through 7. Standard errors were computed as in Experiment 1, except that unlike Experiment 1, the number of observations differed across conditions.

Figure 5 shows the pattern of d' as a function of time for Conditions 1 through 10, positive and negative active and passive tests, collapsed over study sentence type (as in Experiment 1, Figs. 1 and 2, differences due to study sentence were small). The results replicate those of Experiment 1, showing discrimination according to relational information only after 700 ms.

Figure 6 shows the results from Conditions 15 through 18 in Table 2, that is, for irreversible sentences as a function of positive and negative test type and active and passive test type. The results show discrimination according to relational information beginning at about 500 ms, earlier in processing than the 700 ms for the reversible sentences. Apparently, the

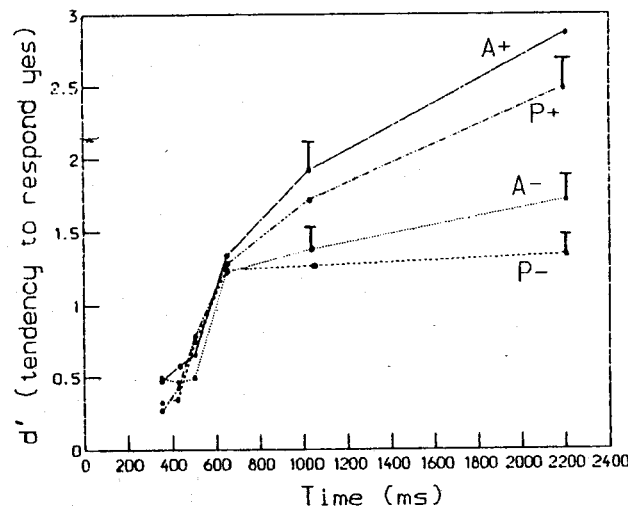


FIG. 5. d' as a function of time for the active and passive test sentences in Experiment 2 averaged over study sentence condition (this replicates the results in Figs. 1 and 2).

meaning of the irreversible sentences provided information earlier in processing than the meaning of the reversible sentences.

Serial processing. Figure 7 shows the conditions in which the first noun or the verb is replaced by a new word. The trends are extremely noisy but there is clearly no evidence for a left to right evaluation process. Such a trend would appear in a time intercept difference between the noun and the verb conditions.

Parallel processing. The data from the single replace conditions can be used to test a simple parallel model of processing, whereby the three content words of a test sentence are matched against memory independently and in parallel. The decision rule for this model is that if one or

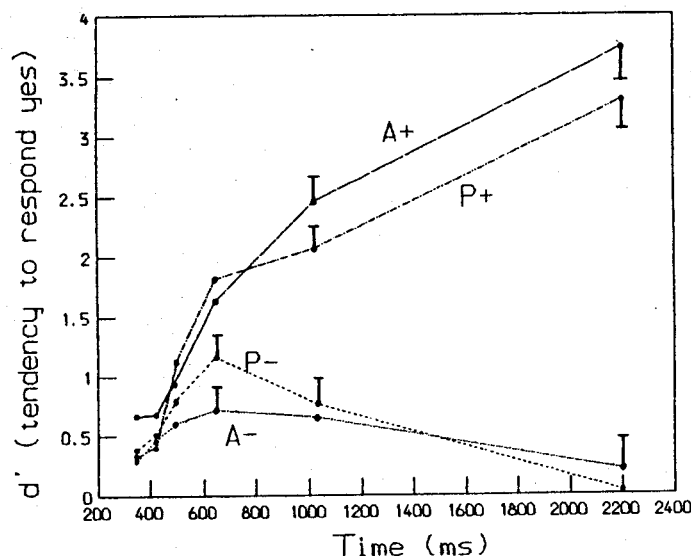


FIG. 6. d' as a function of time for the nonreversible sentences in Experiment 2.

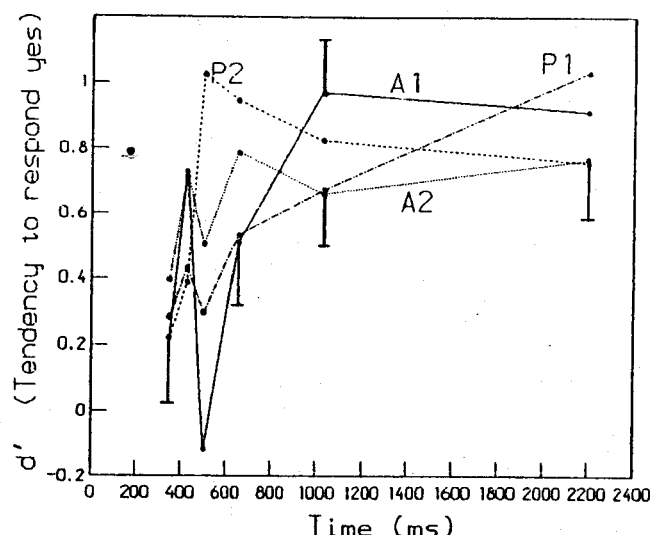


FIG. 7. d' as a function of time for test sentences with the first noun (A1 and P1) or the verb (A2 and P2) replaced by a new word in Experiment 2.

more of the three comparisons produce a mismatch, then the response is negative; otherwise, the response is positive. Assuming parallel and independent comparisons, positive responses for the test sentences with three studied ("old") words represent $\Pr(\text{yes}|\text{three old}) = p_o^3$, and so the data from this condition can be used to determine the probability of a match for a single studied word by calculating the cube root of the probability of a positive response. Similarly, the probability of a positive response for a single "new" word (never studied) can be calculated from the cube root of the probability of a positive response to a test sentence with three new words. $\Pr(\text{yes}|\text{three new}) = p_n^3$. To calculate these probabilities, we used the AA+ and PP+ conditions, combined, for match probability and the AA- and PP- conditions combined for the mismatch probability (these conditions involved no transformation from study to test). From the values of p_n and p_o obtained in this way, the probability of a "yes" response can be predicted for the conditions with one new word, the four single replace conditions: $\Pr(\text{yes}|\text{two old and one new}) = p_o^2 p_n$. Table 3 shows the predictions and data, and the fit is very good for the first four lags with divergence at the last two lags. Thus, early in processing, the results show performance consistent with a model that assumes parallel independent comparisons based on single item information alone. Later in processing, the predicted and empirical functions diverge and this requires that another source of information, i.e., relational, becomes available.

DISCUSSION

From the experiments presented here, it is clear that this example of the

TABLE 3
Predictions and Fit for the Parallel Matching Model

Time (ms)	All old data Pr(yes)	All new data Pr(yes)	One new theory Pr(yes)	One new data Pr(yes)	d'_{theory}	d'_{data}
350	.50	.35	.44	.46	.23	.29
425	.54	.31	.45	.53	.37	.57
500	.58	.28	.46	.44	.48	.43
650	.64	.15	.39	.37	.76	.71
1025	.77	.09	.38	.30	1.04	.82
2200	.79	.04	.29	.16	1.20	.76

Note. d' is computed from the "one new" condition scaled against the "all new" condition, and Pr(yes) is the probability of a "yes" response.

examination of the time course of processing provides data to constrain models of retrieval. This situation parallels past work in recognition memory: the use of measures such as reaction time, in addition to accuracy, has placed additional constraints on models of retrieval processes (e.g., search models for recognition memory for single words). In the same way, time course data from sentence matching procedures can provide constraints on models of the processes underlying sentence matching, and in the future we hope will constrain theories of the processes underlying question answering. However, most current models of retrieval make few predictions about the time course of availability of different kinds of information.

The empirical findings from these experiments provide a clear and interpretable pattern. First, information discriminating test sentences similar (via shared words) to studied sentences from test sentences dissimilar to studied sentences was available quite early in processing. Performance on test sentences with new words (words never studied in any sentences) separated from performance on test sentences with studied words at about 400 ms in both Experiments 1 and 2. Second, for test sentences that contained studied words, information that discriminated a correct (same meaning as a studied sentence) from an incorrect (reversed meaning) version was available only later in processing, at about 700 ms. Third, in Experiment 2, when the studied sentences were not reversible, then information that allowed discrimination between correct and incorrect relations was available by about 500 ms. In other words, information about whether a sentence was pragmatically sensible was available earlier in processing than information about whether one or the other of two possible relationships was the one that was actually studied. (However, we believe that if subjects studied reversed versions of the nonreversible sentences mixed in with sensible sentences, then the same results would

be obtained as for the reversible sentences, because whether a sentence was sensible or not would no longer be a cue to the correct response).

These results are interpretable within the framework of a simple model. At short lags (before correct versions of a studied sentence can be discriminated from incorrect versions), each of the content words of a test sentence provides independent item information toward a decision. For each of the three content words, the probability that it was in a studied sentence is calculated, and the three probabilities are multiplied; the product determines the probability of a "yes" response. This model does not apply after correct versus incorrect versions of studied sentences can be discriminated because then relational information provides extra evidence for the decision. Thus, in interpreting theories of memory below, one of the criteria will be whether they can provide independent item information for each content word of a sentence or a prediction that mimics this, and the second criterion will be whether they can account for the delayed availability of relational information.

The first model to be examined qualitatively is Anderson's (1983) ACT*, a general cognitive theory that specifically deals with the processes underlying sentence matching. ACT* has two major components, a declarative memory and a set of productions that operate on that declarative memory. Retrieval takes place through interactions among the elements (concepts) in the test probe and elements in memory, via a spreading activation process. Activation spreads from all the elements in the probe, and decisions about the quality of match between the probe and memory are carried out by productions that monitor states of activation in declarative memory. For a positive probe sentence, at short response signal lags, the concept nodes representing the sentence in memory will be highly activated and productions that judge connectedness will be available to fire and respond positively. Responses for negative test sentences are generated by a lack of activation using an inhibitory link between the positive production and a negative production, and this enables the negative production to fire and produce a negative response (Anderson, 1983, p. 109). It is unclear whether this could mimic independent parallel processes as described above; detailed fitting of the model would be needed.

For ACT* to discriminate relationally correct from relationally incorrect sentences, additional productions determine the truth value of a test sentence relative to a studied sentence (see Anderson, 1983, p. 116). To account for the late availability of relational information, these productions would have to take more time than the productions that are based on recency of individual words or connectivity among concepts (for the same conclusions, see Anderson, 1983, p. 74).

Although the results presented in this article are consistent with ACT*,

Dosher (1983) has presented data from a reaction time study using similar materials that does cause problems for ACT*. Subjects studied active and passive versions of reversible sentences and were tested with both positive and negative test sentences. Two kinds of negative test sentences were of interest: one in which the only incorrect word of the test sentence was in the correct role but from another study sentence (e.g., an object replacing an object) and one in which the only incorrect word was in an incorrect role (e.g., an object replacing an agent). Results showed that subjects were generally faster and more accurate at rejecting the latter (incorrect role) than the former (correct role). Dosher argues that if responses could be based purely on connectedness, then there should be no differences between these conditions. The connectedness of the two kinds of test sentences would be the same, and lead to a negative response, so that other relations such as the role of the words in the sentence would not affect processing. However, the fact that role affects reaction time and accuracy suggests that some modifications to ACT* are required. But, despite possible modifications, ACT* does contain mechanisms to give different times of availability for different kinds of information.

Unlike ACT*, other current models of retrieval do not have explicit mechanisms to account for differences in the time of availability of item information and relational information. For example, the models of Gillund and Shiffrin (1984) and Murdock (1982) focus primarily on encoding, representation, and retrieval in relatively simple experimental paradigms with simple materials such as single words. The aim of these models is to account for quantitative effects in these paradigms (not a simple or trivial task). To deal with the results from the experiments presented here, it would be necessary to add mechanisms that are currently outside the scope of the models.

The Gillund and Shiffrin (1984) model uses a cue dependent retrieval structure to represent the strength between an item as a retrieval cue and all items in memory as targets. For recognition, the familiarity of the cue is assessed by obtaining the sum of the strengths of that cue across items. For a cue consisting of more than one item, the familiarity is obtained by multiplying cue strengths for an item together and then summing over items. For recall, there are two stages: first, an item image is sampled (the probability is given by the familiarity of that item divided by the total familiarity) and then an attempt is made to retrieve it (the probability of successful retrieval is given by the exponential transformation of the total strength of the cue set). The Gillund and Shiffrin model could account for the results in this article by assuming that early responses are based on familiarity. This would qualitatively mimic the parallel independent model described above because the three conditions (all old words, one new

word, or all new words) would produce a monotonic sequence of familiarity values. But, whether the detailed predictions fitted the data quantitatively would require explicit fits of the models. In parallel with the computation of familiarity, the recall process (which used the computed value of familiarity) would attempt to retrieve the proposition that contains the words presented in the test cue (see Raaijmakers & Shiffrin, 1981). Upon successful retrieval, information would become available to discriminate whether the test sentence was in correct or incorrect order. This processing would require the Gillund and Shiffrin model to contain mechanisms to transform between active and passive versions of sentences and to produce propositional as opposed to surface form representations. Mechanisms would also be required to match retrieved propositions against test sentences, and produce a positive or negative response based on order and relationship among concepts.

In contrast to the Gillund and Shiffrin (1984) model, Murdock's (1982, 1983) model requires that information be stored in at least two forms. Each item is represented as a vector of features, and all vectors (single items and convolutions of more than one item to represent combinations of items) are stored in a common memory vector. To account for the separation of old from new sentences, a decision could be based on memory for the individual items (summed together in the memory vector) and these could be combined as in the parallel independent process model, i.e., an exact mimicking or implementation of that model. To store triples and maintain the order of items in those triples, items would be stored in a chained representation (e.g., storing A, B, and C in that order required the convolutions $A*B$, and $(A*B)*C$ along with A, B, and C stored as individual items; Murdock, 1983). At retrieval, successive correlations would read out the individual items in order, and this could account for the relatively late availability of relational information. However, in the model as currently formulated there is no mechanism for storing propositions as opposed to surface forms (e.g., to relate active and passive). In a vector representation, it is not easy to see how this could be done without using ad hoc assumptions. Also, as with Gillund and Shiffrin model, it would be necessary to add a process that could match a recovered triple against a test triple.

The Gillund and Shiffrin and Murdock models are only two examples. While Anderson's ACT* (1983) model is a notable exception, the general conclusion is that current memory models deal only with a unitary matching process and so cannot account for the finding that different kinds of information become available at different points in time. The models also do not allow for transformations of linguistic forms. The evidence presented in this article provides a basis for testing hypotheses about the behavior of such transformations. The data provide clear estimates of the

time at which independent item information becomes available compared to the time at which information about relations between the items become available.

REFERENCES

- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge: Harvard University Press.
- Corbett, A. T., & Wickelgren, W. A. (1978). Semantic memory retrieval: Analysis by speed-accuracy tradeoff functions. *Quarterly Journal of Experimental Psychology*, 30, 1-15.
- Doshier, B. A. (1976). The retrieval of sentences from memory: A speed-accuracy study. *Cognitive Psychology*, 8, 291-310.
- Doshier, B. A. (1982). Effect of sentence size and network distance on retrieval speed. *Journal of Experimental Psychology: Learning Memory, and Cognition*, 8, 173-207.
- Doshier, B. A. (1983). Relational information in sentence memory. *Journal of Psycholinguistic Research*, 12, 377-396.
- Doshier, B. A. (1984). Discriminating preexperimental (semantic) from learned (episodic) associations: A speed-accuracy study. *Cognitive Psychology*, 16, 519-555.
- Gillund, G., & Shrifin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, 19, 1-65.
- Murdock, B. B. (1982). A theory for the storage and retrieval of item and association information. *Psychological Review*, 89, 609-626.
- Murdock, B. B. (1983). A distributed memory model for serial-order information. *Psychological Review*, 90, 316-338.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85, 59-108.
- Ratcliff, R. (1981). A theory of order relations in perceptual matching. *Psychological Review*, 88, 552-572.
- Ratcliff, R. (1987). Order information and distributed memory models. In *Proceedings of the Ninth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.
- Ratcliff, R. (1988). Continuous vs discrete information processing: Modeling the accumulation of partial information. *Psychological Review*, 95, 238-255.
- Ratcliff, R., & McKoon, G. (1982). Speed and accuracy in the processing of false statements about semantic information. *Journal of Experimental Psychology: Human Learning and Memory*, 8, 16-36.
- Reed, A. V. (1973). Speed-accuracy trade-off in recognition memory. *Science*, 181, 574-576.
- Reed, A. V. (1976). List length and the time course of recognition in human memory. *Memory & Cognition*, 4, 16-30.
- Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, 41, 67-85.
- (Accepted July 6, 1988)