

Speed and Accuracy in the Processing of False Statements About Semantic Information

Roger Ratcliff and Gail McKoon
Yale University

Data from eight experiments on semantic verification are presented. Two procedures were used, a standard reaction time procedure and a response signal procedure. Data from the response signal procedure showed that for false category-member statements (e.g., *all birds are robins*), there was an increasing tendency to respond yes early in processing, replaced later in processing by an increasing tendency to respond no. For statements involving antonym relationships (*all mothers are fathers*), data from the response signal procedure showed that there is no greater tendency to respond yes than for anomalous statements (*a chair is a wall*). Results in the standard reaction time procedure *could not* be predicted from the results in the response signal procedure. This suggests that simple models of the semantic verification task that assume a single yes/no dimension on which discrimination is made are not correct. Some suggestions are made as to the kinds of properties that would be required for an adequate model.

Research in the area of semantic memory has focused on the structure of semantic information and the processes by which subjects verify sentences expressing semantic information. Models of the processes of verification have been developed and the constructs of the models translated into experimental variables that are supposed to affect the time required to verify sentences. These variables have usually involved differences in the information to be processed, such as differences in semantic relatedness or production frequency. For example, in the model of Rips, Shoben, and Smith (1973), the amount of overlap in semantic features between the concepts of a sentence (e.g., between *bird* and *robin* in the sentence *a robin is a bird*) is assumed to affect processing time. This construct is measured by

asking subjects to rate the semantic relatedness of the two concepts. The obtained value of semantic relatedness is then used to predict response time for verification of a statement about the concepts. This is a research strategy that concentrates on differences in the information to be processed. As such it has several problems: For example, the constructs of a model may not be measured correctly (McCloskey & Glucksberg, 1979), or there may exist other highly correlated or more important constructs (Glass, Holyoak, & O'Dell, 1974; Smith, Shoben, & Rips, 1974).

The experiments presented in this article reflect a different strategy. The interest is not primarily in variables that are concerned with differences in the information to be processed. In fact, such variables are held constant. Rather, the aim is to examine the course of processing over time. This is done in the first four experiments with a response signal procedure. In the second four experiments, results from ordinary reaction time experiments are obtained for comparison with the results from the response signal experiments. The experiments are particularly directed toward investigating the falsification of negative statements (e.g., *a mother*

This research was supported by National Institute of Mental Health Grant 32087 to Roger Ratcliff and by National Institute of Child Health and Human Development Grant 13318. We would like to thank Mike Layton for developing the microcomputer system used in the experiments, Andy Beran for conducting the experiments, and Patt McCauley for developing the materials.

Requests for reprints should be sent to Roger Ratcliff, Department of Psychology, Yale University, Box 11A Yale Station, New Haven, Connecticut 06520.

is a father), because falsification data have proven important in evaluating models of semantic verification.

In the first part of the introduction, a brief review of models of semantic processing is presented. In this review, we concentrate on the ways information is processed over time and on the ways negative statements are processed. A more complete review of the area has been provided by Smith (1978). Our review is intended to be sufficient to provide the basis for evaluation of the models in light of the data presented in the eight experiments.

Review of Models

Recent research on semantic memory blossomed as a result of Collins and Quillian's (1969) test of Quillian's (1967) theory of semantic memory. In Quillian's theory, concepts such as *robin*, *bird*, *animal*, and *thing* are stored in a hierarchy with *thing* as the root node and other concepts branching off; for example, *animal* is one link from *thing*, *bird* one link from *animal*, and *robin* one link from *bird*. Processing is assumed to proceed by means of activation spreading along the links. In Collins and Quillian's experiments, subjects were asked to verify statements such as *a robin is a bird* or *a robin is an animal*. The prediction made was that the time to verify a statement would be a linear function of the distance between the concepts in the memory representation. This prediction was confirmed and further experiments followed.

Collins and Quillian's (1969) model does not adequately account for data from falsification judgments. In this model, falsifying a statement involving two unrelated concepts (e.g., *a rock is a thought*) requires very long processing time, but in fact such statements are rejected rapidly. Also, Conrad (1972) found that, in the studies of Collins and Quillian, hierarchical distance in the semantic network was confounded with semantic relatedness. Partly in response to these problems, Rips et al. (1973) and Smith et al. (1974) developed a feature comparison model.

In the Smith et al. model, concepts are represented by bundles of semantic features,

and a statement such as *a robin is a bird* is verified by assessing the relative amount of feature overlap between the two noun concepts in the statement. Processing time is determined by a two-stage decision mechanism. In the first stage, the amount of overlap is computed by a parallel comparison process; if the amount of overlap exceeds one criterion or is less than a second criterion, then a response is made. If the amount of overlap lies between the two criteria, then the second stage is executed. This model predicts that the more feature overlap between two concepts, the faster and more accurate positive responses are and the slower and less accurate negative responses are. Thus, verifying *a robin is a bird* should be faster than verifying *a penguin is a bird* and responding negatively to *a bird is a robin* should be slower than responding negatively to *a bird is a penguin*.

Holyoak and Glass (1975) tested the Smith et al. (1974) model and found that predictions of the model were contradicted with respect to false statements about antonyms (e.g., *a mother is a father*). They found that these statements, in which the two concept nouns are highly similar, were actually rejected faster than statements in which the nouns are less similar. The best predictor of response time was found to be the frequency with which one concept is produced by subjects in response to the other. Holyoak and Glass (1975) also found that in statements such as *all fruits are oranges*, reaction time was determined by how closely a possible counterexample (e.g., *apples*) was associated with the subject noun.

These results led Glass and Holyoak (1975) to develop a model they called the marker search model. This model, like that of Collins and Quillian, assumes that concepts are stored in a semantic network, but it also assumes that there are shortcut paths between some concepts. Processing proceeds by search through the network (whether the search is serial or parallel is unspecified). The search is ordered so that some types of paths are searched before others. The order is specified by production frequency, measured by a sentence completion procedure. For true statements, the greater the production frequency, the shorter the search and

the faster a true response. For example, if *wings* is produced frequently for *all birds have* —, then true responses for the statement *all birds have wings* will be fast. For false statements, a false response rests on finding a contradiction or counterexample. For a contradiction (e.g., *all mothers are fathers*), the greater the production frequency (of *fathers* as a response to *false: all mothers are* —), the faster the response time. For a counterexample (e.g., *all birds are robins*), the greater the production frequency of some other example (e.g., *canaries* as a response to *false: all birds are* —), the faster the response time. These assumptions are sufficient to explain the data for verifying true statements and for rejecting false statements of the types used by Holyoak and Glass (1975) (contradiction and counterexample statements). But the model requires ad hoc assumptions to explain fast falsification of anomalous statements (e.g., *a rod is a thief*).

McCloskey and Glucksberg (1979) recently presented results that they argue contradict both the feature comparison model of Smith et al. and the marker search model of Holyoak and Glass. McCloskey and Glucksberg performed several experiments in which the nature of the statements in the test list was varied. For example, one list might contain only true statements in which the concepts were highly semantically related and false statements in which the concepts were unrelated; another list might contain both highly related and unrelated true and false statements. For statements that appeared in both lists, verification times were longer (and accuracy poorer) in the second case—that is, when highly related false statements were included in the same list with unrelated true statements. From these results, McCloskey and Glucksberg argued that the processing of a particular statement depends not only on the information in that statement but also on the context of other statements in which the information is processed.

To account for their results, McCloskey and Glucksberg (1979) developed a feature comparison model that uses a Bayesian sequential sampling scheme to predict mean reaction time and accuracy. During the

course of processing, features are sampled sequentially; each time a feature is sampled, the total proportions of positive and negative evidence are calculated using Bayes's theorem. When the probability of obtaining those proportions is either larger than a positive criterion or smaller than a negative criterion, a response is made. This model, like the Smith et al. model, cannot account for the fast negative responses obtained by Holyoak and Glass (1975) for statements like *a mother is a father*. There are also other problems with the model proposed by McCloskey and Glucksberg. First, the model predicts that accuracy does not reach asymptote as the boundary criteria are moved apart. This means that subjects could achieve an arbitrarily high level of accuracy. Second, in the mathematical model, computations of accuracy and reaction time are incorrect; the feature comparisons that are assumed to terminate at one of the boundaries include comparisons that would in fact have terminated earlier at the opposite boundary.

The final model to be reviewed in the semantic memory area is an updated version of the Collins and Quillian (1969) model. Collins and Loftus (1975) proposed that there are two stages in the verification process. In the first stage, the path between two concepts in a semantic net is activated by a spreading activation process. Activation spreads in parallel from a node and serially between nodes. The greater the strength of a link, the more activation is passed down it. In the second stage, the path is evaluated. The evaluation process consists of several different subprocesses, some contributing positive evidence (e.g., through a superordinate connection or a match on defining properties) and some negative evidence (e.g., through counterexamples, contradictions, or negative superordinate connections). Positive and negative evidence cancel each other, and the total amount of evidence is accumulated in a Bayesian decision process.

Because the model incorporates many of the processes proposed in other models (e.g., Glass & Holyoak, 1975; Smith et al. 1974), it is capable of dealing with many of the empirical results reported in the literature (such as relatedness effects and fast negatives). Problems with the model involve the

spreading activation process: First, it seems that the spreading activation process is redundant. The evaluation process is necessarily complex, involving partial matching of patterns, coordinating multiple parallel processes, and so on. Finding pathways between nodes would be relatively trivial and could probably be integrated into the evaluation process. Second, Ratcliff and McKoon (1981) attempted to measure the rate of spread of activation and found that only a very short time is required for activation to spread from concept to concept through the memory representation of a paragraph. In their experiments, subjects studied paragraphs and then were required to recognize whether target words had appeared in the paragraphs. Each target was preceded by a prime word that was either near to the target in the structure of the paragraph or far from it. The onset asynchrony of the prime and target was varied (from 50 to 350 msec), and it was found that near and far primes facilitated response times for the targets equally quickly (by 100 msec). In other words, it was not true that activation required more time to spread to the target from a far prime than from a near prime. Thus, we suggest that the spreading activation process cannot be used to account for different response times in different kinds of semantic judgments; response time differences must be accounted for by the evaluation process.

All of the models that have been reviewed were specifically designed for the semantic verification task. They include assumptions about the organization of information in semantic memory as well as assumptions about retrieval processes. We shall use one more model, a model that makes assumptions about retrieval dynamics but not about memory organization, to examine the experimental data. This is a random walk (diffusion) model, of the class of sequential sampling models (Audley & Pike, 1965; Laming, 1968; Link, 1975; Ratcliff, 1978; Stone, 1960). The specific formulation presented by Ratcliff (1978) is used because it deals with the relationship between reaction time and accuracy and also accounts for data from time course studies. In the area of semantic memory, this model is similar to the

models of Collins and Loftus (1975) and McCloskey and Glucksberg (1979) that assume a sequential sampling process for retrieval. The diffusion model is used to compare the results of time course studies with the results of standard reaction time studies in order to assess the ability of sequential sampling models to relate these complementary measures of processing.

Empirical Investigations

The aim of the experiments presented in this article is to provide data that are more comprehensive than the data usually presented in the semantic memory area. Most of the experiments in this area have been concerned only with the statistic mean reaction time. There are several reasons we should go beyond this single statistic. First, there is the problem of mimicking. Townsend (1972) demonstrated that it is possible to mimic serial processing models with parallel processing models at the level of mean reaction time. Thus, network models, which usually assume serial processing from node to node, can be mimicked by feature models, which usually incorporate parallel processing assumptions. In fact, Hollan (1975) pointed out that network models and feature models are formally isomorphic. Rips, Smith, and Shoben (1975) argued in reply that even though the representations of information may be isomorphic, the processing mechanisms are usually quite different. But it is still likely that, with a minimum of ad hoc assumptions, a serial search model could be mimicked by a parallel feature model, and vice versa, at the level of mean reaction time.

The second reason to look further than mean reaction time is that error rates may either covary with mean reaction time or, sometimes, show a speed-accuracy trade-off (e.g., Glass, Holyoak, & Kiger, 1979, Table 2). Some models are able to make predictions about the relationship between accuracy and reaction time (e.g., McCloskey & Glucksberg, 1979; Smith et al., 1974), but others are mute on the subject (Collins & Quillian, 1969; Glass & Holyoak, 1975).

Finally, the third reason to consider more than just mean reaction time is that there exists an alternative way of assessing the

dynamics of processing; deadline or response signal procedures can be used to measure accuracy as a function of time—that is, to trace out the time course of processing (Reed, 1973, 1976). These procedures can be applied to the semantic verification task by presenting the subject with a sentence to verify and then requiring the subject to respond at one of a number of experimenter-determined times. Corbett and Wickelgren (1978) have used the response signal technique to investigate the effect of category dominance on semantic memory retrieval. They found that dominance (the strength of association from a category to an instance) affected only the asymptotic accuracy of a response and not the time at which accuracy rose above chance or the rate of approach to the asymptote. These results show that the response signal procedure can be used effectively in the semantic memory area.

The response signal procedure was used in Experiments 1–4 of this article to examine the time course of processing in the semantic verification task. In Experiments 5–8, a standard reaction time procedure was used in order to provide data for comparison with the response signal data. In both the response signal and standard experiments, two factors were varied. The first was the form of the verification statement: “is a ____ a ____” or “all ____ are ____.” It was thought that a response to the “all” version would require the subject to be more certain. The second factor was whether the parts of the statement (“is a ____” and “a ____”) were presented simultaneously or sequentially. This manipulation was designed to examine the effect of reading time of the statement on processing.

In all eight experiments, the same materials were used; examples are shown in Table 1. We were particularly interested in the response times for false statements because in previous work they have been decisive in the rejection of some models and the development of others. There were three types of false statements; the first type was labeled *opposites*. The two concepts in these statements were very similar and closely associated; most were antonyms, although some were simply different members of a category. Corresponding to the opposites were

Table 1
Examples of Statements Used in the Experiments

Statement type	Example
Opposite	A mother is a father. A lion is a tiger.
Synonym	A carpet is a rug. A cellar is a basement.
Category-member	A color is purple. A bird is a robin.
Member-category	A jeep is a vehicle. A physicist is a scientist.
Anomalous	A problem is a swallow. A captain is a sandwich.
Description	A razor is sharp. A banana is yellow.

synonyms, true statements in which the concepts were as closely associated as the concepts in the opposite statements. The second type of false statements was *category-member* statements; the first concept was the name of a category and the second was a member of that category. The response to *is a bird a robin* should be no because there are birds that are not robins. Corresponding to these false statements were *member-category* statements. The third type of false statement was *anomalous* statements, statements in which the concepts were not related to each other. Finally, to avoid a preponderance of false statements, there were *description* true statements.

In performing response signal experiments, it has been usual to test individual subjects for many sessions. The choice to be made in our experiments was whether to test subjects for many sessions and thus repeat materials or to test subjects for only one session and use the set of materials only once per subject. If materials were repeated, subjects could make use of their memory for their previous response, thus confounding semantic effects with response information. On the other hand, testing for only one session would require averaging data across subjects because of the small number of observations per subject per condition. Such averaging would produce a slowing in the

initial rise of the response signal curve and a slight slowing of the rate. We decided that the problems associated with averaging would be considerably less important than those associated with repeating materials.

Experiment 1

In Experiment 1, the response signal procedure was used to trace the course of processing over time. The first part of a statement was presented for 500 msec, and then the second part of the statement was presented, followed by a signal to respond. For example, *is a mother* was presented for 500 msec, and then *a father* was presented. The signal to respond (a row of asterisks) was given at 50, 150, 300, 600, 1,000, or 1,500 msec after the second part of the statement. Subjects were instructed to respond immediately after the signal. They were given two sessions of practice before the experimental session to ensure that they would be able to do this. Subjects were also instructed and given training in interpreting the *is a* question; for example, they were instructed that the answer to *is a bird a robin?* is no because all birds are not robins.

Method

Subjects. The subjects were 21 Dartmouth College undergraduates who each participated in 3 1-hr. sessions for \$3 per session.

Materials. There were 240 anomalous statements and 120 statements of each of the other types. These statements were chosen from a larger pool of statements according to semantic association ratings. The ratings were collected by asking 40 subjects to decide for each statement how closely related the meanings of the two concepts were on a scale of 1 to 7. Opposite and synonym statements were chosen to have equal average ratings, 6.59 and 6.51, respectively, as were member-category and category-member statements, 5.98 and 5.96, respectively. These semantic association ratings were balanced so that semantic relatedness would not be confounded with sentence type (e.g., subjects could not use relatedness in deciding between opposites and synonyms). The description statements (included in the experiment to increase the number of true statements) averaged 5.72, and the anomalous statements, 1.28. The statements that were not included in the experimental design were used for practice sessions.

Procedure. Stimulus presentation and data collection were controlled by a microcomputer interfaced to Dartmouth's time-sharing system. Stimuli were displayed on a cathode-ray tube (CRT) screen, and subjects responded by pressing keys on the CRT's keyboard (the ? key for *yes* and the Z key for *no*).

There were two kinds of practice trials, which all subjects received. In the first kind, a study-test procedure was used. On each trial, four pairs of words were presented for study and then four test pairs were presented. For each test pair, a subject had to decide whether the two words of the pair had appeared together as a study pair. The first word of a test pair was displayed for 500 msec and then withdrawn; the second word was presented; and then the signal to respond was given. The response signal was presented at a variable lag after the second word (50, 150, 300, 600, 1,000, or 1,500 msec). The subject was instructed to respond as quickly as possible when he or she saw the response signal. Fast responses were encouraged by displaying a subject's response time immediately after each response; subjects were told to try to respond in less than 300 msec.

In the second kind of practice, a semantic verification procedure was used. Statements were of the same types as used in the final, nonpractice session. The first part of the statement was presented for 500 msec, then the second part of the statement, then the response signal, and then, after the subject's response, the time from signal to response. This sequence of events was exactly the same as that used in the final, experimental session.

Each subject participated in three sessions. In the first session, a subject was given three blocks of 30 study-test trials. After each block, a subject's mean accuracy and response time for each response signal lag were calculated and shown to the subject. The subject was encouraged to respond as quickly as possible, even though accuracy was very bad at short lags. In the second session, there were one block of 30 study-test trials and two blocks of semantic verification statements. Each block of verification statements was made up of 16 sets of 30 statements. The subject initiated each set by pressing the space bar of the CRT terminal. Again, a subject's performance was assessed and described to him or her after each block. In the third session, there were 2 sets of 30 verification statements for practice and 28 sets of 30 statements for the experiment itself.

For the experimental materials (the statements used in the final 28 sets of 30 statements), a new random order of presentation of items was used for every two subjects. The response signal lag for each item was chosen randomly from the six possible lags.

Results

The results from this experiment are shown in Figure 1 and Table 2. The course of processing over time is shown in the speed-accuracy curves of Figure 1. The speed-accuracy curves were constructed by calculating d' measures in a standard d' analysis. The d' measures were calculated from group data; the average over subjects had to be used because in some conditions there were zero yes responses for some subjects. The proportion of yes responses for the anomalous statements was the false alarm rate for the other five types of statements.

Table 2
Results From Experiment 1

Response signal lag (in msec)	Response time for yes responses (in msec)	Response time for no responses (in msec)	Probability of a yes response	d'
Opposite				
50	267	270	.262	-.02
150	227	227	.185	.04
300	171	202	.103	.00
600	170	178	.057	.08
1,000	184	167	.031	-.13
1,500	182	165	.020	-.30
Synonym				
50	294	262	.468	.54
150	233	216	.602	1.19
300	199	195	.753	1.96
600	172	197	.913	2.98
1,000	165	188	.929	3.22
1,500	149	183	.942	3.30
Category-member				
50	295	.276	.359	.25
150	240	231	.406	.71
300	204	226	.464	1.19
600	179	205	.402	1.38
1,000	179	178	.311	1.24
1,500	174	174	.235	1.02
Member-category				
50	298	264	.504	.62
150	238	216	.602	1.19
300	206	215	.789	2.08
600	166	193	.873	2.79
1,000	166	183	.918	3.15
1,500	156	190	.918	3.15
Anomalous				
50	277	285	.270	—
150	208	237	.176	—
300	197	200	.102	—
600	187	181	.048	—
1,000	183	170	.041	—
1,500	178	168	.038	—
Description				
50	298	268	.420	.41
150	243	230	.572	1.12
300	196	201	.759	1.98
600	173	183	.887	2.85
1,000	165	178	.909	3.09
1,500	163	188	.936	3.26

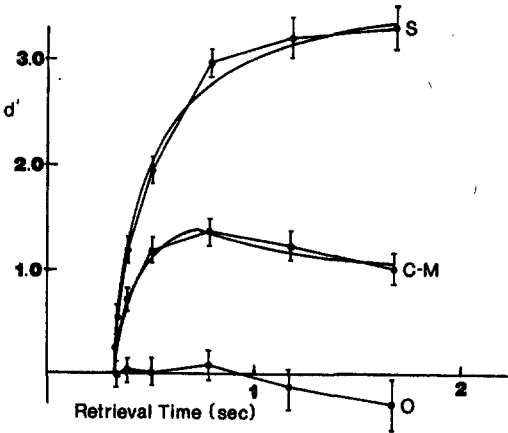


Figure 1. d' as a function of retrieval time for Experiment 1 for synonym (S), category-member (C-M), and opposite (O) statements, with fits of the diffusion model (continuous lines). (Error bars represent 97.5% confidence intervals.)

For each of these five types of statements, the proportion of yes responses served as the hit rate. Thus the ordinate in Figure 1 represents the tendency to respond yes scaled against the anomalous condition. The abscissa represents the total retrieval time computed by adding the mean response latency at each response signal lag to the lag.

The anomalous statements were chosen to scale the other statements against for several reasons. First, scaling all of the other statements against the same single statement type allows comparisons between the other types of statements directly. Second, the anomalous statements are as discriminable from true statements as false statements could possibly be. Of all the types of statements used in the experiments, only in the anomalous statements is there no semantic relation between the two concept nouns. Finally, it is relatively easy to generate large numbers of anomalous statements.

Reed (1973, 1976) and Corbett and Wickelgren (1978) have suggested that the correct, statistically unbiased measure for accuracy is d_T , which is calculated from confidence ratings. The confidence ratings allow the slope of the operating characteristic to be calculated, and from this the statistically unbiased estimate d_T of d' can be

computed. However, the choice we made at the outset was not to repeat materials (because of the possibility the subjects would be able to make use of information in memory about the previous response they made to an item), so there were not enough materials per subject per signal lag to allow meaningful estimates of d' for individual subjects. Thus, unbiased estimates of d' cannot be calculated. This is not a problem, however, because in the results that follow, the absolute level of the d' curves is not vital to any conclusions; rather, relative levels are important and these should have equivalent amounts of bias.

In Figure 1, the confidence intervals shown are 97.5% confidence intervals. They are calculated from a d' value obtained from hit rate plus one standard deviation and false alarm rate minus one standard deviation and from a d' value obtained from hit rate minus one standard deviation and false alarm rate plus one standard deviation. The standard deviations were calculated from $SD = [p(1-p)/N]^{1/2}$, where p is the proportion of hits (or false alarms) and N is the total number of responses.

The first result to note from Figure 1 is that the d' function for category-member statements is nonmonotonic. The curve rises up to about 600 to 800 msec and then falls as retrieval time increases further. This effect is even more pronounced in the proportion of yes responses, as can be seen in Table 2. The difference between d' at 600 msec and d' at 1,500 msec is .36, which is a difference greater than three standard deviations.

The nonmonotonicity immediately suggests that there are two components involved in the processing of the category-member statements, one component providing information that biases toward a yes response and a second component entering processing later and adding a bias toward a no response. The first component might involve the overall similarity of the two concept nouns in the statements, and the second component might involve the exact relationship specified by the statement for the two nouns.

Two components would seem to be compatible with the Smith et al. (1974) model. The first stage of their model, the feature

matching process, would produce evidence indicating a positive response, but then the second stage would indicate a negative response. However, this account is not completely adequate. The first stage should give the same amount of positive evidence for the category-member statements as for the member-category statements because the concepts in these statements are equally semantically related. Thus, the initial parts of the speed-accuracy curves should be the same for the two types of statements. In fact, they are not the same; the category-member curve rises to a lower asymptote in the initial part of its curve than the member-category curve does. This is demonstrated by the fact that points on the category-member curve are always below points on the member-category curve. (Another demonstration is presented in the next section, in which the random walk diffusion model is fit to the curves.) The only way the difference in the initial parts of the curves could be accommodated by the Smith et al. model would be to include in the model an assumption that the onset of the second stage is extremely variable, so variable that the first stage occasionally terminates in almost zero time. With this assumption, the category-member curve would lie below the member-category curve. Other kinds of models, for example, a two-stage parallel processing model, might be able to account for the different initial asymptotes without such an assumption.

In the Glass and Holyoak model (1975; see also Holyoak & Glass, 1975), a negative decision for category-member statements is made by generating a member of the category and then finding a contradiction between the category member in the statement and the generated member. To account for nonmonotonicity in the speed-accuracy curve, it would be necessary to assume a second, parallel process that produces positive evidence over the first part of the time course of processing (and has a different onset time than the first process). A similar mechanism could be added to the Collins and Loftus (1975) model, but it is difficult to see how the McCloskey and Glucksberg (1979) model could account for nonmonotonicity.

The second result is that, at any response signal lag, opposites are just as accurate as anomalous statements (except perhaps at the longest lag). This result holds for subsets of items with equal Kucera and Francis (1967) word frequencies. The result indicates that, through the course of processing, there is just as much evidence toward a negative response for an opposite statement as for an anomalous statement.

These data present problems for the feature-matching models (McCloskey & Glucksberg, 1979; Smith et al., 1974). Both of these models predict that opposite statements would produce more yes responses at short lags than anomalous statements because the concepts in the opposite statements are very similar and so there would be considerable feature overlap. At longer lags, the second stage of the Smith et al. model would produce no responses (and a nonmonotonic speed-accuracy curve), but the McCloskey and Glucksberg model would continue to produce yes responses.

In contrast to the feature-matching models, the network models have no difficulty with the result that responses to opposite statements are very accurate even at short response signal lags. A similar result has been found by Holyoak and Glass (1975) and Glass et al. (1974) with a standard reaction time procedure, and our result extends their result to cover the course of processing over time. It is exactly these kinds of results that the Holyoak and Glass (1975) model was designed to explain, and Collins and Loftus (1975) have incorporated mechanisms like those of the Holyoak and Glass model into their model. However, these network models would predict that anomalous statements would be less accurate than opposite statements, and this is not the case in our data.

For the positive statements (synonyms, member-category, and description), results are much as would be expected from the study of Corbett and Wickelgren (1978). The d' curves rise to an asymptote monotonically, and all reach about the same asymptote. To prevent clutter on the graph, only the curve for synonyms is shown in Figure 1 (although the data are presented in full in Table 2).

The Random Walk Diffusion Model

All of the current models of semantic memory in their present form have been contradicted by one or more aspects of the results of Experiment 1. A random walk model, however, without assumptions about semantic structure, can be used to interpret the retrieval dynamics shown in the time course curves (Figure 2) and to provide estimates of various parameters for the different conditions such as asymptotic accuracy, the time at which accuracy begins to rise above zero, and the rate of approach to the asymptote. In a random walk model, it is assumed that evidence toward a response cannot be assessed instantaneously and so must be sampled continuously. The process of evidence accumulation is represented by the random walk process. Positive and negative evidence are accumulated continuously, and positive and negative evidence cancel each other.

For the response signal procedure, processing of all statements is assumed to begin with the same starting amount of evidence. As processing proceeds, evidence about a particular statement accumulates either positively or negatively. If the rate at which evidence is accumulated (the drift rate) is positive, then as time increases evidence tends to accumulate toward the positive side of the starting point. If the drift is negative, evidence tends to accumulate toward the negative side of the starting point. The decision rule proposed by Ratcliff (1978) for the continuous version of the random walk (the diffusion process) is that if the process is on the positive side of the starting point when the response signal occurs, then respond yes, or, if the process is on the negative side, respond no.

To model a nonmonotonic d' function, it is necessary to assume that the drift rate changes at some time during processing. In the fits presented later, the assumption is made that the change in drift occurs at some single point in time. If it were assumed instead that drift rate changed over some distribution of time values, fits of the model to data would not be altered significantly. (For discussion of other possibilities, such as continuous changes in drift over time, see Ratcliff, 1980.)

To model the response signal procedure, it is necessary to assume that the drift rate is subject to two sources of variability: The first is variability in the drift rate itself during the course of a single comparison. The second is variability among the items of a condition; responses to different items in the same condition will vary in difficulty and so vary in drift rate. For diffusion models of the recognition process, it has been shown that both these sources of variability are necessary to fit time course data and reaction time data (Ratcliff, 1978; Reed, 1976). If variability within a trial were zero, then the d' curve would rise to asymptote immediately when any evidence became available. If there were no variability in drift between trials, then accuracy would not reach asymptote, but instead would increase as a function of the square root of time (or as a function of the square root of the number of steps in the discrete random walk).

The distribution of drift rates for nominally identical items is represented by a normal distribution (n) with mean u and variance η , $n(u, \eta)$, where u represents the drift rate toward a positive response. The quantity s^2 represents the variability in the drift rate itself. To calculate the asymptotic d' value for a particular type of item, $[n(u, \eta)]$ scaled against another type of item with a drift rate of v , $[n(v, \eta)]$, the quantity $d'_{asy} = (u - v)/\eta$ is formed. Then the formula for the time course curves for d' as a function of time can be written:

$$d' = \frac{d'_{asy}}{\{1 + s^2/[\eta^2(t - t_{ER})]\}^{1/2}}$$

For the nonmonotonic curves, it is assumed that the drift rate changes at some time t_1 , so the formulae for d' are

$$d' = \frac{d'_1}{\{1 + s^2/[\eta^2(t - t_{ER})]\}^{1/2}}, \quad t \leq t_1,$$

$$d' = \frac{d'_2 + (d'_1 - d'_2)(t_1 - t_{ER})/(t - t_{ER})}{\{1 + s^2/[\eta^2(t - t_{ER})]\}^{1/2}}, \quad t \geq t_1,$$

where d'_1 is the asymptotic d' for $t \leq t_1$ and d'_2 is the asymptotic d' for $t \geq t_1$.

Figure 1 shows fits of the diffusion model

to the data from Experiment 1. The fits are achieved by means of a minimization routine: Theoretical values of d' are calculated using the above equations, and the function that is the sum of squares of experimental minus theoretical values of d' is minimized by varying the theoretical parameter values. Eight parameters are used: asymptotic d' values for the three positive statements, d'_1, d'_2 , an encoding and response time parameter t_{ER} , the time at which new information becomes available for the category-member statement t_1 , and the variance parameter s^2/η^2 (assumed to be the same for all five item types). The variance ratio parameter was estimated to be 447 msec, and the encoding and response parameter, 332 msec. The estimate for the time at which new information becomes available for the category-member statements is 696 msec, and the estimates for the asymptotic d' s are $d'_1 = 2.15$ and $d'_2 = .91$. The asymptotic d' values for the positive statement curves are estimated to be very nearly equal (3.91, 3.80, 3.81, for synonyms, member-category, and description, respectively). The fits of the diffusion model to the data are good, falling largely within the 97.5% confidence intervals shown in Figure 1.

Up to this point, the diffusion model has been presented as a description of the data (i.e., simply as an empirical fit of the data). However, the fact that the fits of the diffusion model to the data are good shows that a model of the sequential sampling type can describe the time course of processing of semantic statements in the response signal paradigm. A diffusion model is particularly attractive because it can be used to account for several measures of performance, including mean reaction time, accuracy, response time distributions, and response signal curves. The good fit of the model to the data of Figure 1 allows parameter estimates and comparisons of parameters (e.g., comparison of the asymptote of the member-category statements and the asymptote of the first process for the category-member statements).

It is appropriate to make some comments about other methods of fitting response signal curves. The data presented above could also be fit by exponential functions (e.g.,

Corbett & Wickelgren, 1978; Doshier, 1979). These functions would provide empirical summaries of the data. However, any interpretations of the data (e.g., does Condition 1 have a faster rise time than Condition 2?) would depend on the exponential assumption. Different models would give different interpretations (e.g., equal rise times to different asymptotes in the exponential would be fitted by different rise times to different asymptotes using a ramp function). In general, the time intercept and rate parameter are model dependent, and conclusions based on these parameters should be considered carefully.

The goodness of fit of the diffusion model to the data of Figure 1 could be measured with a statistic such as R^2 (Reed, 1973, 1976). However, such a measure must be viewed with caution. It is necessary to determine how much of the variance in the data is due to noise and how much should be accounted for by a model. If models are to be discriminated on the basis of very high R^2 values, then noise in the data may be determining the choice of model. It would also be possible to determine whether different parameters were significantly different from each other. (This has not been done in empirical work reported so far using the response signal method.) Estimates of variance in parameters could be obtained by Monte Carlo methods.

For the experiments reported in this article, the fits of the diffusion model are sufficient to describe the data, and the fits would not be significantly improved with other models. Furthermore, none of the major conclusions depends on the exact form of the diffusion model, and none of these conclusions would be altered by better fits.

Experiments 2, 3, and 4

In Experiment 1, the time course of processing of semantic information was examined. The results showed that the time course curve was nonmonotonic for category-member statements and that accuracy for opposite statements was equal to accuracy for anomalous statements at all but the longest response signal lags.

The next three experiments were designed to replicate the results of Experiment 1 and to address two further questions. First, what happens if the whole statement is presented at once? In Experiment 1, the first part of a statement was presented 500 msec before the second part. Presenting the first concept early may have aided processing (as in a priming effect) for some of the types of statements; for example, for opposites, high associates may have been generated when the first concept was presented early, resulting in faster response times than would otherwise have been the case. Second, what happens if the statement is changed from *is a* _____ *a* _____ to *all* _____ *are* _____? In many experiments using the semantic verification task, the choice of the form of the statement has not been considered. However, the results of Holyoak and Glass (1978) suggest that the quantifiers *all*, *many*, *some*, *a few*, and *none* form a unidimensional scale with *all* requiring the most certainty for a response and *none* the least. Thus, we can predict that responses to the *all* statements will be more accurate than responses to the *is a* statements. This effect may interact with the time course curves to reduce the nonmonotonicity of the category-member statements.

Experiment 2

Method

In Experiment 2, the form of the statements was *is a* _____ *a* _____, and the parts of the statement were presented simultaneously. (In the study-test practice, the two test items were also presented simultaneously.) There were 23 subjects and the response signal lags were 100, 250, 450, 700, 1,000, and 1,500 msec. In all other respects, the experiment was the same as Experiment 1.

Results

The results for Experiment 2 are shown in Figure 2 and Table 3. The main results of Experiment 1 are replicated. The d' curve for category-member statements shows a significant nonmonotonicity, and opposite statements are just as accurate as anomalous statements except at longer response signal lags.

Table 3
Results From Experiment 2

Response signal lag	Response time for yes responses (in msec)	Response time for no responses (in msec)	Probability of a yes response	d'
Opposite				
100	284	286	.394	.00
250	261	260	.440	.01
450	247	253	.376	-.03
700	236	224	.229	-.16
1,000	250	221	.139	-.06
1,500	265	205	.114	-.17
Synonym				
100	282	298	.412	.05
250	265	259	.425	-.30
450	248	250	.481	.26
700	236	254	.646	.95
1,000	216	260	.721	1.60
1,500	206	261	.781	1.81
Category-member				
100	280	285	.483	.21
250	263	268	.419	-.02
450	249	260	.495	.27
700	240	250	.534	.67
1,000	236	235	.492	.99
1,500	226	220	.299	.52
Member-category				
100	276	285	.424	.09
250	277	262	.426	-.02
450	245	259	.487	.28
700	229	256	.675	1.03
1,000	215	260	.741	1.65
1,500	210	247	.800	1.88
Anomalous				
100	284	285	.391	—
250	270	265	.436	—
450	256	255	.390	—
700	248	236	.281	—
1,000	243	228	.155	—
1,500	243	215	.151	—
Description				
100	282	284	.388	.00
250	249	260	.440	.02
450	251	250	.496	.29
700	234	242	.646	.96
1,000	213	245	.717	1.58
1,500	205	262	.759	1.74

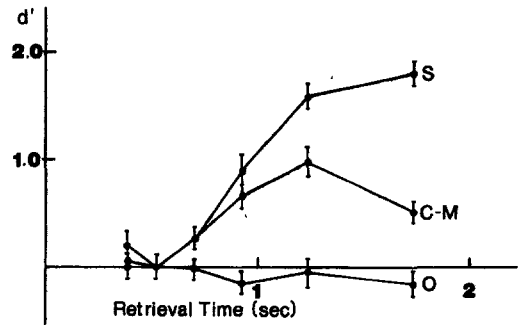


Figure 2. d' as a function of retrieval time for Experiment 2 for synonym (S), category-member (C-M), and opposite (O) statements. (Error bars represent 97.5% confidence intervals.)

There are two main differences between the curves shown in Figures 1 and 2. First, the curves in Experiment 2 begin to rise above chance at about 600–700 msec, much later than the curves of Experiment 1 do. This 300–400 msec difference can be attributed to the increased reading time required in Experiment 2, in which the whole statement was presented simultaneously. Second, in Experiment 2, the asymptotic value reached by the synonym curve (and also the member-category and description curves) is considerably lower than in Experiment 1.

Fits of the diffusion model were not made for this data because it is difficult to estimate where the curves first start to rise above zero. This is because averaging across subjects whose performance begins to rise above zero at different points leads to an S-shaped d' function. In Experiment 1, on the other hand, the variability in reading time was much reduced by presenting the first part of the statement 500 msec before the second part.

Experiment 3

Method

In Experiment 3, the form of the statements was *all _____ are _____*. The parts of the statement were presented sequentially; *all _____ are* was presented for 500 msec, and then the second concept was presented. There were 14 subjects and the response signal lags were 50, 150, 350, 700, 1,250, and 2,000 msec.

Table 4
Results From Experiment 3

Response signal lag	Response time for yes responses (in msec)	Response time for no responses (in msec)	Probability of a yes response	d'
Opposite				
50	308	290	.284	.03
150	249	253	.243	-.02
350	243	239	.139	-.04
700	230	220	.089	.01
1,250	244	211	.043	-.32
2,000	348	204	.022	-.46
Synonym				
50	329	282	.378	.29
150	260	245	.409	.44
350	247	248	.611	1.32
700	228	256	.823	2.27
1,250	215	222	.883	2.60
2,000	214	230	.915	2.96
Category-member				
50	303	297	.302	.09
150	285	252	.341	.26
350	245	261	.312	.54
700	247	226	.175	.44
1,250	228	213	.099	.12
2,000	254	212	.070	.12
Member-category				
50	313	283	.369	.26
150	272	242	.413	.45
350	243	247	.601	1.30
700	228	237	.832	2.32
1,250	215	241	.886	2.62
2,000	209	216	.874	2.74
Anomalous				
50	302	299	.276	—
150	248	253	.250	—
350	263	239	.149	—
700	242	226	.084	—
1,250	250	212	.077	—
2,000	256	212	.056	—
Description				
50	295	278	.335	.17
150	284	239	.383	.38
350	248	238	.599	1.30
700	217	232	.820	2.29
1,250	217	226	.864	2.50
2,000	219	211	.896	2.85

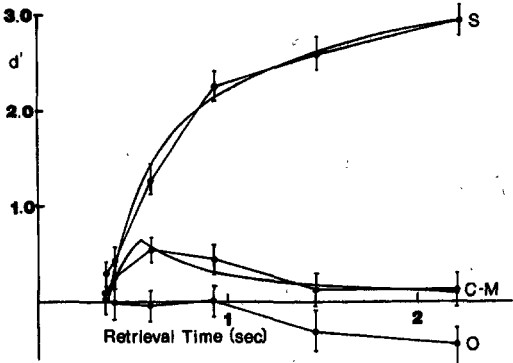


Figure 3. d' as a function of retrieval time for Experiment 3 for synonym (S), category-member (C-M), and opposite (O) statements, with fits of the diffusion model (continuous lines). (Error bars represent 97.5% confidence intervals.)

Results

The results for Experiment 3 are shown in Figure 3 and Table 4. These results replicate the effects found in Experiments 1 and 2. First, opposite and anomalous statements are equally accurate at short response signal lags, with opposite statements becoming a little more accurate at long lags. Second, the category-member curve shows nonmonotonicity, although the size of the effect is attenuated. The curve rises to half the maximum height of the curves in Experiments 1 and 2 and begins to fall at an earlier point in time to a considerably lower asymptote. This pattern is shown by the fits of the diffusion model. In Experiment 3, the time at which secondary information becomes available for the category-member statements is estimated at 503 msec, compared with 696 msec in Experiment 1. The asymptotic d' values are $d'_1 = 2.56$ and $d'_2 = -.05$, compared with 2.15 and .91 in Experiment 1. This pattern of results fulfills the expectation that the *all* version of the statements would lead to greater accuracy for the category-member statements relative to the anomalous statements and indicates that the improvement occurs mainly in the second part of processing (d'_2).

The parameter estimates for positive responses are much the same as for Experiment 1; $d' = 3.65$, 3.61, and 3.60 for syn-

onyms, member-category, and description, respectively. The variance ratio estimate is about double that for Experiment 1 (1,051 msec). This could be simply a result of different subjects in the two experiments (individual differences in this parameter are usually quite large; see Reed, 1976). Since the rate parameters are averaged over many subjects in the two experiments, however, the difference in the variance ratio estimate may be due to the difference in experimental manipulations in the two experiments. The response and encoding time parameter was estimated at 354 msec.

Experiment 4

Method

This experiment completes the manipulations of Experiments 1, 2, and 3. The parts of a statement were presented simultaneously, and the form of a statement was *all* — are —. There were 15 subjects, and the response signal lags were 100, 300, 550, 900, 1,400, and 2,100 msec.

Results

The results for this experiment are shown in Figure 4 and Table 5 and are similar to those found in the first three experiments. Opposite statements are just as accurate as anomalous statements, except at the longest lag, where opposites are slightly more accurate. For the category-member statements, nonmonotonicity is difficult to detect because the curve does not rise very high. This is consistent with the result of Experiment 3 and the hypothesis that the *all* form

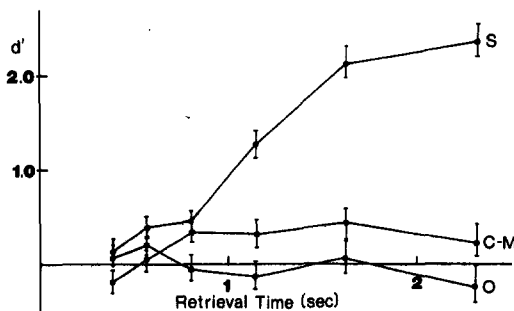


Figure 4. d' as a function of retrieval time for Experiment 4 for synonym (S), category-member (C-M), and opposite (O) statements. (Error bars represent 97.5% confidence intervals.)

Table 5
Results From Experiment 4

Response signal lag	Response time for yes responses (in msec)	Response time for no responses (in msec)	Probability of a yes response	d'
Opposite				
100	280	280	.347	.06
300	264	259	.361	.20
550	279	253	.225	-.06
900	260	236	.138	-.13
1,400	237	212	.106	.06
2,100	212	215	.053	-.24
Synonym				
100	287	279	.372	.13
300	268	260	.429	.39
550	266	251	.393	.45
900	238	252	.619	1.28
1,400	219	249	.793	2.14
2,100	210	238	.831	2.37
Category-member				
100	284	276	.265	-.20
300	270	260	.314	.05
550	268	262	.359	.34
900	272	250	.261	.31
1,400	236	221	.185	.44
2,100	229	215	.123	.22
Member-category				
100	282	281	.333	.03
300	279	261	.377	.26
550	263	259	.384	.43
900	232	256	.647	1.36
1,400	226	239	.813	2.22
2,100	214	235	.858	2.48
Anomalous				
100	281	284	.325	—
300	263	260	.286	—
550	260	253	.236	—
900	245	240	.166	—
1,400	245	221	.092	—
2,100	237	215	.077	—
Description				
100	283	276	.360	.10
300	256	253	.343	.16
550	262	259	.434	.56
900	242	242	.664	1.40
1,400	227	244	.738	1.97
2,100	212	233	.809	2.28

of the statements gives greater accuracy than the *is a* form. The three positive curves once again fall nearly on top of each other and so, once again, only the curve for synonyms is shown. As in Experiment 2, it is difficult to estimate the point at which accuracy rises above zero, so fits of the diffusion model were not attempted.

Experiment 5

Experiments 5 through 8 were designed to produce the reaction time data corresponding to the response signal data from Experiments 1 through 4. The correspondence of response signal experiments to reaction time experiments is Experiment 1 to Experiment 6, Experiment 2 to Experiment 5, Experiment 3 to Experiment 8, and Experiment 4 to Experiment 7. The aim was to collect data using the same materials with both procedures in order to investigate constraints on models that would not be obvious from either procedure alone. This point can be illustrated by using the example of a sequential sampling model (e.g., McCloskey & Glucksberg, 1979) and the opposite and anomalous statements.

For a sequential sampling model, the data for the opposite and anomalous statements in a reaction time experiment are under two constraints. The first constraint comes from the model itself and is independent of the response signal results. This constraint is that responses to opposite and anomalous statements cannot show a speed-accuracy trade-off. In other words, responses to one of the types of statements can be slower and less accurate than responses to the other type of statement, or the responses can be equal in speed and accuracy. But responses to one of the types cannot be slower and more accurate than responses to the other type (as long as information begins to accumulate at the same point in time for the two types and the functions are monotonic). The only way one type of statement could be slower and more accurate than another in a sequential sampling model is for the subject to change the decision criteria, that is, to demand more evidence for one type of statement before responding. But subjects cannot do this for

one type of statement independent of the other type of statement, because they have no way of knowing which type of statement is which until evidence has accumulated.

The second constraint on the data of a reaction time experiment for a sequential sampling model comes from the response signal data. The result that accuracy is equal for the two types of statements implies (for the sequential sampling models) that accuracy and response time must be equal in the reaction time procedure. This is because, if the items are of equal difficulty in the response signal procedure, they must be of equal difficulty in the reaction time procedure. According to a sequential sampling model, items of equal difficulty must be equal in response speed and accuracy. Thus, for a sequential sampling model of the type proposed by McCloskey and Glucksberg (1979), the only result possible with the reaction time procedure is equal response times and accuracies for the opposite and anomalous statements except at long reaction times, at which opposites may be a little faster and more accurate.

It might seem that sequential sampling models could allow other results by invoking a rechecking process. A mandatory rechecking process for one type of statement would lead to slower and more accurate responses for that type of statement. The problem with a rechecking process is that it adds considerable complexity to otherwise simple models. In a sequential sampling model, processing no longer would involve only the accumulation of a quantity of homogenous evidence; rather the quality of the evidence would be used to decide whether rechecking is necessary. This would be completely contrary to the spirit of sequential sampling models because not only is a yes-no decision made, but also the type of statement is discovered, and this would necessitate the inclusion of other processes. Similarly, in the Smith et al. (1974) model, the quality (rather than the quantity) of Stage 1 evidence would have to be used to decide whether Stage 2 processing should proceed. This would, of course, completely change the model.

A secondary aim of Experiment 5 was to

provide sufficient data to allow examination of response time distributions (see Ratcliff, 1979; Ratcliff & Murdock, 1976) in order to place further constraints on models of retrieval processes.

Method

Subjects. The subjects were 37 Dartmouth College undergraduates who each participated in a 1-hr. session for extra credit in an introductory psychology course.

Materials. The statements used in this experiment were the same as those used in the final, experimental sessions of Experiments 1-4.

Procedure. The statements were presented in the form *a* — *a* —, and the parts of the statement were presented simultaneously. The statements were presented in a different random order for every two subjects. They were divided into 28 sets of 30 statements each (preceded by 2 sets for practice). A subject initiated each set by pressing the space bar on the CRT's keyboard. Each statement was displayed until the subject made a response; then there was a 500-msec pause, and then the next statement was presented. Subjects were encouraged to respond quickly and accurately.

Results

Table 6 shows mean response time and accuracy for each type of statement. Figure 5 shows group response time distributions (Ratcliff, 1979) for the anomalous, opposite, category-member, and synonym conditions. The distributions for the other positive conditions are very similar to the synonym distribution, so they are not shown.

The results of most interest are those for opposite and anomalous statements. The data show a speed-accuracy trade-off; op-

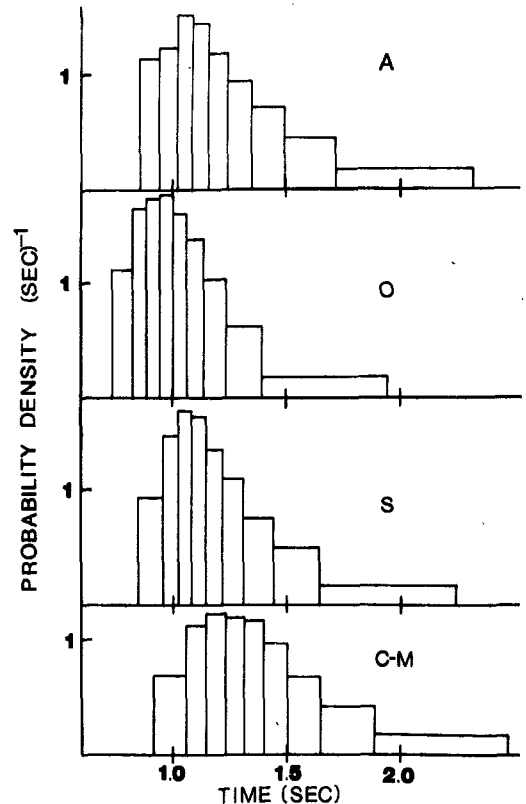


Figure 5. Group reaction time distributions for Experiment 5 for anomalous (A), opposite (O), synonym (S), and category-member (C-M) statements.

posites are faster and less accurate than anomalous statements. Even though the accuracy difference is small (.058 vs. .036), it is highly significant ($z = 5.5$, $p < .000001$), as is the reaction time difference (204 msec, $SE = 9$ msec).

There are three possible explanations of this trade-off that can be ruled out. First, the trade-off is not due to high error rates for just a few items; the same items were used in the response signal experiments in which opposite and anomalous statements were equally accurate. Second, the trade-off is not due to high error rates for just a few subjects; inspection shows that the majority of subjects show the trade-off. Third, it could be that the higher accuracy for anomalous statements is the result of an optionally activated rechecking process (e.g., the second stage of the Smith et al., 1974, model). Such

Table 6
Results From Experiment 5

Statement type	Response time for yes responses (in msec)	Response time for no responses (in msec)	Probability of yes responses
Opposite	1,135	1,118	.058
Synonym	1,288	1,283	.822
Category-member	1,282	1,477	.367
Member-category	1,298	1,278	.861
Anomalous	1,274	1,322	.036
Description	1,226	1,252	.836

Note. Standard error in correct reaction time is ± 7 msec.

a probability mixture of processes would be apparent in the distributions of response times. The fastest responses in the distributions for opposite and anomalous statements would be equally fast, but there would be more slow responses in the anomalous distribution (i.e., there would be an elongated tail). As can be seen in Figure 5, the distributions do not show this pattern. There is no overlap between the fastest responses for the two conditions; the fastest responses for the anomalous condition are 100 msec slower than the fastest responses for the opposite condition, and in fact the whole distribution for the anomalous condition is shifted and skewed relative to the opposite condition.

As discussed earlier, for some models the response signal curves constrain the reaction time and accuracy data such that they should be the same for opposite and anomalous statements because, during the course of retrieval, these statements gave equal values of accuracy (see Figure 2). Therefore the speed-accuracy trade-off for opposite and anomalous statements immediately disconfirms sequential sampling models like that proposed by McCloskey and Glucksberg (1979).

The category-member condition shows a longer mean reaction time and higher error rate than the anomalous condition. This result suggests that the category-member statements are more difficult than the anomalous statements, as would be expected from the response signal time course curves in which accuracy for the category-member statements was always poorer than accuracy for the anomalous statements. The group reaction time distribution for the category-member condition shows a shift and skewing relative to the anomalous condition. This suggests that the slower response times for the category-member statements are not caused by an additional, sometimes activated, rechecking process; such a process would result in a probability mixture of processes and there is no evidence for such a mixture in the response time distributions.

As in the response signal experiments, the positive conditions all show roughly the same general levels of speed and accuracy. How-

ever, accuracy for the synonym and member-category conditions exhibits a large difference (highly significant), whereas response times are almost equal. For the sequential sampling models, this result is inconsistent with the response signal data. In the response signal experiments, when response time was equal for the synonym and member-category conditions, accuracy was also equal. Therefore this result, like the results for opposite and anomalous statements, show that for some models the response signal and reaction time results are inconsistent.

Group reaction time distributions for the positive conditions are very similar to each other, and so only the distribution for the synonym condition is shown.

Before we present further discussion of these results, the results from the reaction time analogues of Experiments 1, 3, and 4 are presented.

Experiment 6

Method

In Experiment 6, the form of the statements was *is a — a —*, and the first part of the statement was presented for 500 msec prior to the second part. There were 11 subjects. In all other respects, the experiment was like Experiment 5.

Results

The results are shown in Table 7. As in Experiment 5, there is a speed-accuracy trade-off between the anomalous and op-

Table 7
Results From Experiment 6

Statement type	Response time for yes responses (in msec)	Response time for no responses (in msec)	Probability of yes responses
Opposite	706	756	.142
Synonym	826	889	.791
Category-member	798	1,032	.338
Member-category	832	847	.814
Anomalous	762	842	.136
Description	826	848	.795

Note. Standard error in correct reaction time is ± 11 msec.

posite conditions, and, in the category-member condition, accuracy is poorer and response time longer than in the anomalous condition. However, unlike in Experiment 5, accuracy and reaction time for the synonym and member-category conditions are compatible.

Experiment 7

Method

In Experiment 7, the form of the statement was *all — are —*, and the parts of the statement were presented simultaneously. There were 11 subjects.

Results

Results from Experiment 7 are shown in Table 8. Once again there is a speed-accuracy trade-off between the anomalous and opposite conditions. In this experiment, as in the last one, positive responses do not show significant speed-accuracy trade-offs.

In Experiments 3 and 4, we found that the accuracy of the category-member statements relative to the anomalous statements was significantly increased by changing the statement from the *is a* form to the *all* form. In Experiment 7, we found that the effect of *all* is to speed the responses to the category-member statements so that they become just as fast as responses to the anomalous statements, whereas accuracy is significantly lower (.140 errors vs. .020 errors). It was argued before that the speed-accuracy trade-off for the opposite and

Table 8
Results From Experiment 7

Statement type	Response time for yes responses (in msec)	Response time for no responses (in msec)	Probability of yes responses
Opposite	1,347	1,337	.032
Synonym	1,433	1,495	.816
Category-member	1,434	1,434	.140
Member-category	1,449	1,415	.848
Anomalous	1,887	1,479	.020
Description	1,378	1,459	.838

Note. Standard error in correct reaction time is ± 12 msec.

Table 9
Results From Experiment 8

Statement type	Response time for yes responses (in msec)	Response time for no responses (in msec)	Probability of yes responses
Opposite	1,301	940	.038
Synonym	1,050	1,550	.892
Category-member	1,103	1,103	.089
Member-category	1,054	1,349	.886
Anomalous	1,611	1,098	.033
Description	1,008	1,392	.832

Note. Standard error in correct reaction time is ± 12 msec.

anomalous conditions was inconsistent with sequential sampling models of the form proposed by McCloskey and Glucksberg (1979) because those models predict that accuracy and reaction time should be correlated. The results for the category-member and anomalous statements provide further evidence that these models are inadequate.

Experiment 8

Method

In Experiment 8, the form of the statement was *all — are —*, and the first part of the statement was presented 500 msec prior to the second part. There were 9 subjects.

Results

The results for Experiment 8 are shown in Table 9 and replicate the results from Experiment 7. There is a speed-accuracy trade-off for the opposite and anomalous conditions, and the category-member condition gives reaction times almost the same as for the anomalous condition, but accuracy is much poorer. For positive responses, the description condition shows a speed-accuracy trade-off relative to the category-member and synonym conditions.

General Discussion

In the experiments of this article we have presented data that provide new insights into the processes underlying semantic verifica-

tion. After describing the major findings, we offer suggestions for the kind of processing system required to deal with these findings.

First, the data from the response signal experiments show a nonmonotonic function for category-member statements (e.g., *a bird is a robin*). Early in processing, there is an increasing tendency to respond yes. This tendency is replaced later in processing by an increasing tendency to respond no. This suggests that, at some time in processing, either new information becomes available or a new stage of processing is encountered. These two kinds of information or processing can be tentatively identified with similarity information and relational information, respectively.

The nonmonotonic function for the category-member statements is inconsistent with the models of McCloskey and Glucksberg (1979), Glass and Holyoak (1975), and Collins and Loftus (1975). It offers some support to the two-stage assumption of the Smith et al. model (1974), but does not support the exact predictions of the model because the early parts of the category-member and member-category curves are different.

The second result of note is that opposite statements are just as accurate as anomalous statements at all except the longest retrieval times (at which opposites are slightly more accurate). This is true despite the fact that the relatedness of the two concepts in the opposite statements was high. This result is inconsistent with the models of McCloskey and Glucksberg (1979) and Smith et al. (1974). High accuracy on opposite statements is consistent with the Collins and Loftus (1975) and Glass and Holyoak (1975) models, which were specifically designed to account for such data but equally high accuracy on *anomalous* statements is not consistent with these models.

The third important result comes from comparison of the reaction time data with the response signal data. For opposite and anomalous statements, there is a speed-accuracy trade-off; responses to opposite statements are faster and less accurate than responses to anomalous statements. This result is inconsistent with the McCloskey and

Glucksberg (1979) and Collins and Loftus (1975) models, which both incorporate a sequential sampling decision mechanism. With this mechanism, accuracy and response time must be correlated across experimental conditions, unless the subject can know which condition is being tested and adjust criteria. This was not possible in the experiments presented here. A speed-accuracy trade-off is also inconsistent with the Smith et al. (1974) model. The Glass and Holyoak (1975) model does not make strong predictions about data for anomalous statements but would predict that longer reaction times would be correlated with lower accuracy.

To summarize, the Smith et al. (1974) model has problems in dealing with the highly accurate opposite statements and with the speed-accuracy trade-off. The McCloskey and Glucksberg (1979) model has problems with all three of the major findings (the nonmonotonic d' function for category-member statements, the accuracy on opposite statements, and the speed-accuracy trade-off). The Glass and Holyoak (1975) model has problems with the nonmonotonic d' function, the speed-accuracy trade-off, and equal accuracy for opposite and anomalous statements. Finally, the Collins and Loftus (1975) model has problems with the speed-accuracy trade-off, equal accuracy for opposite and anomalous statements, and, in an unelaborated form, with the nonmonotonic d' function.

The diffusion model fitted to the speed-accuracy curves in Experiments 1 and 3 has the same problems as the other sequential sampling models in dealing with the speed-accuracy trade-offs found in Experiments 5-8. However, a multiple counter diffusion model (like that described below), in which evidence was accumulated mainly in two counters, would approximate the two-dimensional model, and thus the fits presented in Experiments 1 and 3 would still be valid. If more than two counters were accumulating significant amounts of evidence over the same period, the fits would not be quantitatively accurate but would still capture qualitative aspects of the data.

In one or another respect, the results of the experiments presented here disconfirm

all of the current models of semantic memory. However, the results do suggest one property that a model for the verification task should possess, and that is that different statements are assessed by different mechanisms. This is necessary to allow the possibility of speed-accuracy trade-offs for different statements. This suggestion is similar to that made by Collins and Loftus (1975) in their model, except that instead of evidence being either positive or negative for an overall yes or no decision, evidence is either positive or negative for a type of statement.

There is currently one theoretical scheme that fits this suggestion, the logogen model (Morton, 1969). The logogen model was developed to account for data from word recognition studies, but it can be easily modified for semantic verification studies. In this model, different counters accumulate evidence for different types of statements, and a response is initiated when one of the counters reaches a criterion amount of evidence. In the model proposed by Morton, the criterion is absolute—it does not matter whether one of the other counters has accumulated a lot or a little evidence. The problem with this is that, as the criterion is increased (e.g., with “respond accurately” instructions), the reaction time distribution becomes more normal. This is the opposite of what is observed empirically.

Another possibility is that the criterion is relative; for a response to be initiated, the evidence in one counter must exceed the maximum of the other counters by some criterial amount (see Audley & Pike, 1965, for further discussion). In the two-counter case, the relative criterion model is a random walk or diffusion process. In the multicounter case, the model is a multidimensional generalization of the random walk. Thus, the model has many of the properties of the models of Collins and Loftus (1975) and McCloskey and Glucksberg (1979). This system could produce speed-accuracy trade-offs (e.g., with the opposite and anomalous statements) by allowing different criteria or counter starting points for the different types of statements. The model produces reasonable predictions about the shape of reaction

time distributions; as mean reaction time increases, the distribution becomes more skewed. The relative criterion model is also capable of producing reasonable predictions about the shapes of response signal curves. Thus the model has an adequate retrieval mechanism and can capture most aspects of the data presented in this article, but it has little to say about the representation of semantic information. However, the representation would have to allow different kinds of information to become available at different times during the course of processing in order to give nonmonotonic d' curves.

In conclusion, the challenge for research in the area of semantic memory is clear: Models must be developed that are tightly constrained but that accommodate the complexity of the processing involved in making judgments about semantic information.

References

- Audley, R. J., & Pike, A. R. Some alternative stochastic models of choice. *British Journal of Mathematical and Statistical Psychology*, 1965, 18, 207-225.
- Collins, A. M., & Loftus, E. F. A spreading-activation theory of semantic processing. *Psychological Review*, 1975, 82, 407-428.
- Collins, A. M., & Quillian, M. R. Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 1969, 8, 240-247.
- Conrad, C. Cognitive economy in semantic memory. *Journal of Experimental Psychology*, 1972, 92, 149-154.
- Corbett, A. T., & Wickelgren, W. A. Semantic memory retrieval: Analysis by speed accuracy tradeoff functions. *Quarterly Journal of Experimental Psychology*, 1978, 30, 1-15.
- Doshier, B. A. Empirical approaches to information processing: Speed-accuracy tradeoff functions or reaction time—A reply. *Acta Psychologica*, 1979, 43, 347-359.
- Glass, A. L., & Holyoak, K. J. Alternative conceptions of semantic memory. *Cognition*, 1975, 3, 313-339.
- Glass, A. L., Holyoak, K. J., & Kiger, J. I. Role of antonymy relations in semantic judgments. *Journal of Experimental Psychology: Human Learning and Memory*, 1979, 5, 598-606.
- Glass, A. L., Holyoak, K. J., & O'Dell, C. Production frequency and the verification of quantified statements. *Journal of Verbal Learning and Verbal Behavior*, 1974, 13, 237-254.
- Hollan, J. D. Features and semantic memory: Set-theoretic or network model? *Psychological Review*, 1975, 82, 154-155.
- Holyoak, K. J., & Glass, A. L. The role of contradictions and counterexamples in the rejection of false sen-

- tences. *Journal of Verbal Learning and Verbal Behavior*, 1975, 14, 215-239.
- Holyoak, K. J., & Glass, A. L. Recognition confusions among quantifiers. *Journal of Verbal Learning and Verbal Behavior*, 1978, 17, 249-264.
- Kucera, H., & Francis, W. N. *Computational analysis of present-day English*. Providence, R.I.: Brown University Press, 1967.
- Laming, D. *Information theory of choice reaction time*. New York: Wiley, 1968.
- Link, S. W. The relative judgment theory of choice reaction time. *Journal of Mathematical Psychology*, 1975, 12, 114-135.
- McCloskey, M., & Glucksberg, S. Decision processes in verifying category membership statements: Implications for models of semantic memory. *Cognitive Psychology*, 1979, 11, 1-37.
- Morton, J. The interaction of information in word recognition. *Psychological Review*, 1969, 76, 165-178.
- Quillian, R. Word concepts: A theory and simulation of some basic semantic capabilities. *Behavioral Science*, 1967, 12, 410-430.
- Ratcliff, R. A theory of memory retrieval. *Psychological Review*, 1978, 85, 59-108.
- Ratcliff, R. Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, 1979, 86, 446-461.
- Ratcliff, R. A note on modeling accumulation of information when the rate of accumulation changes over time. *Journal of Mathematical Psychology*, 1980, 21, 178-184.
- Ratcliff, R., & McKoon, G. Does activation really spread? *Psychological Review* 1981, 88, 454-462.
- Ratcliff, R., & Murdock, B. B., Jr. Retrieval processes in recognition memory. *Psychological Review*, 1976, 83, 190-214.
- Reed, A. V. Speed-accuracy trade-off in recognition memory. *Science*, 1973, 181, 574-576.
- Reed, A. V. List length and the time course of recognition in immediate memory. *Memory & Cognition*, 1976, 4, 16-30.
- Rips, L. J., Shoben, E. J., & Smith, E. E. Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, 1973, 12, 1-20.
- Rips, L. J., Smith, E. E., & Shoben, E. J. Set-theoretic and network models, reconsidered: A comment on Hollan's "Features and semantic memory." *Psychological Review*, 1975, 82, 156-157.
- Smith, E. E. Theories of semantic memory. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes* (Vol. 6). Hillsdale, N.J.: Erlbaum, 1978.
- Smith, E. E., Shoben, E. J., & Rips, L. J. Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, 1974, 81, 214-241.
- Stone, M. Models for choice reaction time. *Psychometrika*, 1960, 25, 251-260.
- Townsend, J. T. Some results concerning the "identifiability" of parallel and serial processes. *British Journal of Mathematical and Statistical Psychology*, 1972, 25, 168-199.

Received November 30, 1980

Revision received June 1, 1981 ■