1976

The processing of false propositions/

Robert Frederick Lorch
University of Massachusetts Amherst

THE PROCESSING OF FALSE PROPOSITIONS

A Thesis Presented

By

ROBERT FREDERICK LORCH, JR.

Submitted to the Graduate School of the University of Massachusetts in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

August 1976

Psychology

THE PROCESSING OF FALSE PROPOSITIONS

A Thesis Presented

By

ROBERT FREDERICK LORCH, JR.

Approved as to style and content by:

Jerome L. Myers, Chairperson of Committee

Daniel R. Anderson, Member

Alexander Pollatsek, Member

Bonnie R. Strickland, Department head
Department of Psychology

ABSTRACT

Two hypotheses (Holyoak and Glass, 1975) concerning how particular types of false statements are recognized as false were investigated. Experiment 1 examined two factors affecting the time subjects need to reject false, universal affirmative propositions in which the subject- and predicate-concept represent instances from the same semantic category (e.g., "All skunks are beavers"). The more accessible in memory was the fact that the subject and predicate had the same immediate superordinate, the faster subjects were to reject the statement in a sentence-verification task. Also, the more similar the two instances were, the slower the subjects were to decide that the proposition was false.

Experiment 2 examined two factors affecting reaction time to reject false, universal affirmative propositions in which the predicate is an instance in the category designated by the subject-term (e.g., "All animals are panthers"). The more accessible a counterexample to the proposition was (e.g., "Some animals are dogs"), the faster subjects were to reject the statement. Also, the more typical the predicate was rated as an instance of the subject-category, the slower subjects were to reject the statement. These findings were interpreted as posing difficulties for the models of semantic memory proposed by Smith, Shoben, and Rips (1974) and Holyoak and Glass (1975), while being consistent with some proposals of Collins and Quillian (1969). The implications of the results were considered with respect to some important issues concerning semantic memory structure and process.

TABLE OF CONTENTS

LIST OF TABLES

INTRODUCTION

Since Collins and Quillian's (1969) seminal paper, semantic memory has been a prominent topic of research in cognitive psychology. Much of the research which attempts to elucidate semantic memory organization and process has concerned itself with the question of how subjects verify true statements. However, investigators have recently begun to look more closely at how false statements are processed. There are two reasons for this development. First, an adequate theory of semantic memory must provide an account of how false assertions are recognized as such. Second, detailed examination of how false statements are processed may provide information about basic memory structure and process which verification studies fail to supply.

In a typical semantic memory task, the subject's task is to judge the truth value of a simple statement as quickly as possible. For example, the sentence "All canaries are birds" might be presented and the subject's latency to respond "true" would be the dependent variable of interest. In most studies, a positive response to a stimulus sentence probably depends upon the recognition that the asserted relationship between the subject and predicate -- "canaries" and "birds" -- is the actual relationship between them. (An exception to this statement is the work of Collins and Quillian (1969; 1970b; 1972b).) The rejection of a false statement, on the other hand, probably entails the use of information not presented in the stimulus.

It may sometimes be the case that directly-stored negative information is used to reject a false proposition. For example, the assertion that "A dog is a cat" may be rejected by finding the information in memory that "A dog is <u>not</u> a cat". However, direct storage of negative information is probably an infrequent occurrence. More often, rejections probably result from logical inferences based on less direct information concerning the assertion being evaluated. For instance, a subject knows that both "dog" and "cat" occupy the same level in the same semantic hierarchy -- both are common instances of the category "animals". That information is one potential basis for the inference that a dog is not a cat. If the statement "All animals are dogs" is presented, it may be rejected in at least two ways. The subject may recognize that the statement reverses the actual relationship between "animals" and "dogs"; or a counterexample to the statement might be retrieved from memory, e.g., "Some animals are cats". Another rejection strategy might depend on inductive reasoning. For example, a subject who does not have stored the fact that "A whale is a mammal" may reason that a whale is like a fish and fish are not mammals, therefore a whale is not a mammal. (Such a strategy is probably time-consuming and error-prone.) Still another potential basis for inferring that a proposition is false is the failure to find the proposition, or any relevant information with which to evaluate the proposition, in memory. If the proposed instance, e.g., "whale", is not found after an exhaustive search of the designated category, "mammal", then a negative decision results. This strategy is the implicit basis for <u>all</u> negative decisions in some category-search models of long-term memory (e.g., Landauer and Meyer, 1972). As such, it represents a rather special rejection strategy.

It is easy to understand the difficulties in trying to explain how false statements are processed. However, not only is the issue an interesting one in itself, it encompasses other important issues concerning semantic memory: What is stored in memory?; How is information organized?; Is the search process involved in information retrieval directed or undirected (Anderson and Bower, 1972)?; How is relevant information distinguished from irrelevant information during the evaluation of a proposition? The present study focuses on two possible bases subjects use for rejecting false statements in an effort to gather some information regarding the questions which have been raised.

A REVIEW OF THE LITERATURE

The theoretical context within which the present research evolves includes a number of important issues specific to the question of how false statements are processed and to the more general topic of semantic memory structure and process. However, two issues are of central importance. These issues are: (1) are "false" responses based upon something other than a failure to find a "match" in memory with the stimulus proposition?; and (2) if "false" responses are based on the positive retrieval of some sort of information, what sort of information is it and how is it accessed? The literature review follows the organization provided by these two issues. Four theories of semantic memory which respond to these questions in different ways are considered.

First, the exhaustive search hypothesis represents the simplest response to the above questions. It states that false propositions are rejected when no "match" for the stimulus proposition can be found in memory.

Next, the feature comparison model remains within the set-theoretic tradition of the exhaustive search model in maintaining an emphasis on a simple match-mismatch process of evaluation of stimulus propositions. Thus, like the exhaustive search hypothesis, it answers the first question above in the negative. However, the feature comparison model conceptualizes the processes leading to the rejection of a false statement as quite different from those hypothesized by the exhaustive search model. The source of the differences lies in the unique system of information representation hypothesized by the feature comparison model.

The final two theories to be reviewed contend that "false" responses are based upon something other than the failure to find some simple "match" in memory. Rather, the theories of Holyoak and Glass, and Collins and Quillian hypothesize that many false propositions are rejected on the basis of the retrieval and evaluation of indirect relationships between their subject- and predicate-concepts. The theories differ in how they conceptualize the retrieval process and in the types of information they claim is considered during the evaluation of a proposition.

The following review is divided into four subsections corresponding to the four theories noted above. The organization of each subsection is basically the same: A theory is presented; then the evidence supporting it is examined; finally, some of its weaknesses are noted. Throughout the section, the concern is with results from variations on the same basic paradigm. The sentence-verification paradigm involves presentation of a logically quantified or unquantified proposition which the subject must evaluate and respond "true" or "false" to as rapidly as possible. A prominent variation of this task involves presentation of just two nouns, always in the same relative positions (e.g., one above the other) on each trial; the subject's task being to determine whether, for example, the bottom noun is a member of the category designated by the top noun. Reaction time is the dependent variable in both the sentence-verification and the categorization paradigm.

The Exhaustive Search Hypothesis

Although there are a number of bases on which a proposition might be rejected, not all theorists agree that the processing of false statements is complex. In particular, a frequent hypothesis of category-search models (e.g., Landauer and Meyer, 1972) is that negative decisions result after an exhaustive search of the designated category fails to locate the target instance. That is, no "match" can be found in memory for the stimulus proposition. A direct prediction of this "exhaustive search hypothesis" is that negative RT should increase monotonically with increases in the size of the search set. For example, according to the exhaustive search hypothesis, it should take longer to respond that a "table" is not an "animal", than to respond that a "table" is not a "dog".

The issue of the effect on reaction time (RT) of semantic category size in a sentence-verification task was a prominent one during the first few years after Collins and Quillian's (1969) paper. In general, the results of this research support the exhaustive search hypothesis. Small, but consistent category-size effects on negative RT have been reported by several investigators (Landauer and Freedman, 1968; Meyer, 1969; 1970; 1973a; 1973b; Meyer and Ellis, 1970; Wilkins, 1970). However, interpretations of these findings are made ambiguous by methodological difficulties consistently encountered in this literature. Typically, no tests of the generality of category-size effects to the population from which stimulus materials were selected are provided (Clark, 1973). Further, usually no controls are provided for possible systematic differences in word-frequency as a function of category-size; or for possible differences in associative relationships of the instances with respect to the large and small categories. Both of these factors are known to affect RT and are frequently found to covary with category-size (e.g., Collins and Quillian, 1970a; Wilkins, 1971; Loftus and Freedman, 1972; Anderson and Reder, 1974). (Landauer and Meyer, 1972, have discussed the difficulties of the category-size research in some detail.) Thus evidence that subjects reject false propositions on the basis of not being able to find confirming information in memory is unconvincing.

There are stronger reasons for rejecting the exhaustive search hypothesis as the sole basis for a negative decision. There is direct evidence (Anderson and Reder, 1974; Glass, Holyoak and O'Dell, 1974; Holyoak and Glass, 1975) that subjects do use other bases for rejecting false propositions. Further, there is evidence that the category-size effect disappears or actually reverses for particular category comparisons (e.g., when instances are categorized with respect to the category "mammals" versus "animals"), or when the semantic relatedness of the instance and category and category-size are manipulated orthogonally (Collins and Quillian, 1970a; Rips, Shoben, and Smith, 1973; Smith, Shoben, and Rips, 1974). Thus, the exhaustive search hypothesis seems insufficient as the sole explanation of how false propositions are processed.

Smith, Shoben, and Rips: A Feature Comparison Model

Smith, Shoben, and Rips (1974; Smith, Rips, and Shoben, 1974) have proposed a model which was originally formulated to explain performance in categorization tasks, but which has been generalized to account for results from a wide range of semantic memory experiments. Although within the set-theoretic approach, the theory does not assume that an exhaustive search process is the basis for all negative responses. However, it does emphasize a match-mismatch process of information evaluation and the memory structure it hypothesizes is designed to implement such a process in a straightforward manner.

The most important aspect of the feature comparison model is the representational system for semantic information it hypothesizes. It is proposed that concepts are represented in memory as a set of semantic dimensions. Each dimension is associated with a weight which indicates how essential that dimension is to the meaning of the lexical item. These weights form the basis for a distinction made between "defining" and "characteristic" features. When an item is retrieved from memory, it is not the entire set of dimensions but "points" on the relevant dimensions which are sampled. For each concept, there is a distribution of possible values on any one dimension corresponding to one's subjective impressions of the relative frequencies of the dimension values for that item: The momentary (random variable) values on the relevant semantic dimensions of a concept are called "features". These structural assumptions are the task-invariants of the model; the basic processing assumptions follow.

A two-stage process model is proposed to account for categorization performance. When a proposition is presented, the defining and characteristic features of both the subject and predicate concepts are retrieved and compared during the first stage of processing.[1] For example, if the proposition is "A canary is a bird", the attributes of both bird (e.g., feathered, flies, wings, beak, etc.) and canary (e.g., feathered, flies, wings, yellow, etc.) are retrieved and compared. The holistic comparison process yields a measure, x, of the overall similarity of the two concepts. This measure may be thought of as representing the semantic "overlap" between the two concepts; or the proportion of all predicate features shared by the subject concept. If x is greater than some high criterion of dissimilarity, a fast positive decision results; if x falls below a criterion of dissimilarity, a fast negative decision results; if x falls between the two similarity criteria, the second stage of processing is entered.

In the second stage of processing, the defining features of the subject (canary) and predicate (bird) of the proposition are isolated and compared. Defining features are essential to the meaning of a concept, while characteristic features are not. For example, a defining feature of all birds is that they have wings; however, it is only characteristic of birds that they can fly. If (a) each defining dimension of the predicate is also a defining dimension of the subject concept, and (b) the values on the

---

[1]Note the difference between the exhaustive search model and the feature comparison model: the former states that a long-term memory representation is compared with the stimulus proposition during the match-mismatch evaluation; while the latter states that the comparison is between two long-term memory representations -- the subject- and predicate-concepts.

dimensions (features) of the subject-concept are within the allowable range for the predicate-concept, a positive response results; otherwise, a negative response will be given. The feature-comparison process is hypothesized to be exhaustive for both true and false subject-predicate pairs. Finally, second-stage processing is hypothesized to be error-free; all errors are the result of first-stage processing.

The predictions of the feature comparison model with respect to categorization performance are straightforward. If a positive decision is called for, RT will be shorter the more similar the subject and predicate concepts are because a high proportion of responses will be made on the basis of first-stage processing only. If a negative response is appropriate, greater similarity between the subject and predicate will slow RT because second-stage processing will be more frequent. If the degree of similarity between the subject and predicate contradicts their actual relationship, the probability of a fast (first stage), incorrect response is increased. These predictions are well-supported in the literature. We are concerned with the pattern of negative RTs as a function of subject-predicate similarity.

A number of studies suggest that as subject-predicate similarity increases, negative RT increases (Collins, 1969; 1970a; 1970b; 1972b; Conrad, 1972; Meyer, 1970; Rips, 1975; Rips, Smith, and Shoben, 1973; Schaeffer and Wallace, 1969; 1970; Smith, <u>et. al.</u>, 1974a; Wilkins, 1971). However, these experiments are subject to many of the criticisms made with respect to the category-size effect literature. There frequently were no controls for the possibility that word-frequency and subject-predicate similarity may covary in the stimuli used. Further, no statistical measure of the generalizability of results to items other than those specifically used as stimuli is presented in any of the above experiments. On the other hand, the effects reported are quite large (ranging up to 250 milliseconds) relative to those generally reported for category-size manipulations. Also, the effect of semantic relatedness on positive RT is the best documented in the semantic memory literature; thus, semantic relatedness would be expected to have a prominent influence on the processing of false propositions as well. (The term "semantic relatedness" is used to refer noncommitally to any of several, highly-confounded measures of the strength of the semantic relationship between two concepts. Examples of such measures are similarity and typicality ratings, production frequency, and association norms.)

Smith, et. al. (1974b) have conducted an experiment which responds to most of these criticisms. It is a simple categorization experiment in which stimulus presentation was blocked by category (judgments were made with respect to the same category throughout a block of trials). The relevant manipulation was that negative instances were of three levels of rated-relatedness to the target category. The results were in accord with the model's predictions: slower RTs and higher error rates were found with increasing category-negative instance similarity. However, the size of the effect was small (thirty-six milliseconds) and no measure of the statistical reliability of the effect was reported. A probable reason for these unconvincing results that negative instances were always drawn from the same category for a particular trial block (the category most closely related to the target category). This procedure probably restricted the range of similarity of the negative instances to the target category and may have encouraged a strategy of rejecting negative instances by recognizing that they belonged to the single category from which all negative instances were selected in a particular block of trials (Anderson and Reder, 1974; Glass, et. al., 1974; Holyoak and Glass, 1975).

Glass, et. al. (1974) have presented clear evidence in support of the feature comparison model. They controlled word-frequency, collected objective measures of subject-predicate relatedness for false as well as true propositions, and used statistical procedures allowing generalization of the results to the population of items from which the stimulus materials were selected (Clark, 1973). Their sentence-verification experiment (Experiment 2) manipulated the quantifier ("all", "some", "many", "no"), truth value, and semantic relatedness of the proposition. Highly-related subject-predicate pairs for true propositions were replaced by quantifiers which made them false (e.g., for "Some birds are canaries", "no" might replace "some" to produce a high-relatedness, false proposition). Low-relatedness, false propositions were constructed by rearranging the predicates of the high-relatedness, false propositions to form anomalous statements. For instance, "All chairs are lions" would be an anomalous proposition. As expected, RTs were slower for high- than for low-relatedness false propositions quantified by "some", "all", and "many" (the effect ranged from 144 to 362 milliseconds); but were faster for high- than for low-relatedness false propositions quantified by "no" (eighty-two milliseconds).[2] These findings were replicated (Experiment 3) for property statements (adjectives were used in the predicate position instead of nouns) and with the addition of "few" (which behaved as predicted for a negative quantifier) as a fifth quantifier.

To summarize, there appears to be clear evidence for the similarity/negative RT effect predicted by the feature comparison model (Smith, et. al., 1974a; 1974b). If the comparison of semantic features in a match-mismatch process is the sole basis on which the truth value of a proposition is judged, then the complicated picture of negative information processing drawn in the introduction to this study is incorrect. However, there is recent evidence that at least some of the bases for rejecting false propositions which were suggested in the introduction may in fact be used by subjects. Let us now turn to that evidence.

---

[2]The pattern is reversed for negative quantifiers presumably because subjects evaluate the affirmative, "base" portion of a proposition, then apply the negative quantifier to reverse the decision reached in the first step of evaluation. Thus, for the proposition "No dogs are animals", the subject first determines that a dog is an animal. This retrieved fact matches the root proposition, "Dogs are animals". To the "match" outcome is then applied the negative quantifier with the result of a negative decision. The important observation to be made is that, according to this model (which is adopted by Smith, et. al., 1974b), a negative decision with respect to a negatively-quantified proposition is based on a "match" outcome during the memory search and comparison stages of information-processing. Since high semantic-relatedness is hypothesized to facilitate a "match", RT to negatively-quantified false propositions and positively-quantified true proposition should -- and does -- decrease as semantic-relatedness increases.

Holyoak and Glass: Contradictions and Counterexamples

Holyoak and Glass (1975) do not provide a well-formulated alternative to the exhaustive search and feature comparison models. However, they are explicit in their claim that not all false responses are based on the failure to find a match between either a memory representation and the stimulus proposition (exhaustive search model), or between long-term memory feature-lists representing the subject- and predicate-concepts of a proposition (feature comparison model). They contend that many false responses are based on the positive retrieval of information about the subject- and predicate concepts of a proposition which only indirectly relates the two concepts (e.g., a "dog" and a "horse" are indirectly-related in that they both are "animals").

Holyoak and Glass (1975) hypothesize that items retrieved from memory in response to a particular stimulus word are found in a limited capacity search whose order is determined by the relatedness of the individual items to that word. An important implicit assumption made is that the search is one-way, proceeding from the subject-concept and unaffected by the particular predicate of the complete proposition which is presented in the actual sentence-verification task.[3]

Holyoak and Glass's explanation of how propositions are actually rejected is more detailed. They are in agreement with the position that false propositions are rejected for numerous reasons. They have considered three "strategies" of information-

[3]It is not clear that Holyoak and Glass wish to be held to this conceptualization. However, it is a theoretical issue of some interest and will tentatively be attributed to them as a point of contrast with Collins and Quillian's model of the retrieval process.

evaluation other than simple verification by recognition, (e.g., "A is, indeed, a B"), (Glass, et. al., 1974; Holyoak and Glass, 1975); although nothing in their work indicates that they consider these strategies the only bases for reaching a decision as to the truth-value of a proposition. Of particular interest are two hypothesized bases for a negative decision -- the contradiction and counterexample hypotheses (Glass, et. al., 1974; Holyoak and Glass, 1975). The contradiction hypothesis asserts that false propositions in which the subject and predicate represent concepts from the same level in the same semantic hierarchy (e.g., "dog" and "cat") are rejected on the basis of finding that the actual relationship between the two concepts contradicts the asserted relationship. Thus, RT to make a "false" response is determined by how quickly the information is accessed that the subject and predicate have the same, immediate superordinate. The counter-example hypothesis refers to a second situation in which contradictory information forms the basis for a negative decision. In the case of universal propositions, if the subject of the proposition is a superset of the predicate, or the predicate is one of more than one possible value of an attribute of the subject-concept; then the proposition may be rejected when a counterexample to the assertion is retrieved. For example, the assertion that "All animals are dogs" will probably be rejected on the basis of finding in memory an instance of the category "animal" which is not also the particular predicate, "dog" (e.g., "cat").

The contradiction and counterexample hypotheses make clear an important point of contrast between Holyoak and Glass and Smith, Shoben and Rips. Holyoak and Glass not only contend that many false statements are rejected on the basis of the positive retrieval of information about the subject- and predicate-concepts of the propositions, they further hypothesize that information about the subset/superset relationships the subject and/or predicate concept are involved in is a particularly important basis for a decision about the truth-value of a proposition. This is in contrast to the feature-comparison model, which emphasizes the importance of property information concerning a concept and denies that subset/superset information is utilized during the evaluation of a proposition.

To summarize the Holyoak and Glass (1975) model, information retrieved in an ordered, (self-terminating) search of the subject-concept's associates is compared with the asserted relationship between the subject and predicate. Retrieved information which contradicts the asserted proposition forms the basis for a "false" response.

There are four important sources of support for Holyoak and Glass's contentions. First, Anderson and Reder (1974) have conducted a multiple regression study of factors which might affect RT in a categorization task. The dependent variable of interest here is RT to reject a negative instance, or "instance negation time" (INT). Of twenty-two independent variables entered into a stepwise multiple regression analysis, three significant predictors of INT were identified: (1) RT to judge that the instance, or subject-concept, is a word (presumably a measure of encoding time); (2) time to generate a superordinate of the instance; and (3) the time to recognize that the predicate concept was not a member of another category occupying the same level in the same semantic hierarchy as itself. As Anderson and Reder (1974) note, these findings suggest that the subject generates a category of the instance and determines that that category is disjoint from the presented category. Although Anderson and Reder hypothesize that the latter part of this "strategy" involves direct storage of the disjoint relationship between the actual and the presented category, the findings are also consistent with Holyoak and Glass's contradiction hypothesis.

A study by Millward, Rice, and Corbett (1975) may be interpreted as support for Holyoak and Glass's contention that contradictory information retrieved from memory is used to reject false propositions. Millward, et. al., used a simple categorization paradigm and investigated a number of variables, only one of which is of interest here. Negative instances were selected from five levels of dominance (determined by ordinal rank in production frequency norms) within their respective categories. The finding was that negative RT increased with decreasing instance-dominance, except for a slight decrease in RT from the second-lowest to the lowest level of dominance. On the assumption that instance- and category-dominance are positively-correlated, these findings are consistent with those of Anderson and Reder (1974) and thus with Holyoak and Glass's (1975) contradiction hypothesis.

Two more direct sources of evidence for Holyoak and Glass's ideas are their own studies. Glass, Holyoak, and O'Dell (1974, Experiment 4) investigated the pattern of RTs as a function of the production frequency of the predicate to the quantified subject of a proposition. Two quantifiers, "many" and "few", were used. True propositions were those for which the predicate-completion was one given in a production task to the particular quantifier-subject pair used in the sentence-verification task. False "few" propositions substituted a predicate (property) produced for a "many" completion task; false "many" propositions were completed with predicates produced for a "few" completion blank. For example, if a frequent completion of "Many arrows are ______" was "pointed", subjects in the verification task would be presented the false proposition "Few arrows are pointed". For both true and false propositions, statements with high frequency predicates were processed more quickly than those with low frequency predicates. The results support the contention that false propositions may be rejected on the basis of contradictory information retrieved from memory. Further, on the assumption that production frequency and subject-predicate similarity are positively correlated (Rips, et. al., 1973), the finding that "false" responses were faster for high frequency predicates is contrary to the feature comparison model's predictions.

Holyoak and Glass (1975, Experiment 2) have performed another sentence-verification experiment based on explicit statements of the contradiction and counterexample hypotheses, and more directly concerned with the differential predictions of Smith, et. al.'s (1974a; 1974b) theory and their own. The experiment will be considered in detail because it represents the point of departure for the present research. Only those aspects of the experiment relevant to the present considerations will be discussed.

The experiment may be divided into two sub-experiments according to whether the contradiction or counterexample hypothesis is being examined. First, the contradiction experiment used both universal- and particular affirmative propositions. Stimuli were selected on the basis of production frequency norms (Experiment 1) in which subjects were instructed to give false completions to incomplete propositions of the form: "Quantifier A are _____". For instance, if the stimulus was "All arrows are _____", a frequent, false completion would be "bows". Three levels of predicate frequency were used: high (mean production frequency of thirty-five percent); low (mean production frequency of five percent); and anomalous (four percent). Anomalous predicate completions were defined as those for which the intersection of the subject and predicate concepts was more remote than the immediate superordinate of both. To summarize the design, only subject-predicate pairs corresponding to the contradiction situation were used in a two (quantifier) by three (production frequency of the predicate) factorial design.

Holyoak and Glass expected that the shortest RTs would be to anomalous propositions; while low production frequency predicates should be slowest. Since false production frequency presumably represents a manipulation of the accessibility of subset/superset information and the feature comparison model contends that such information is not utilized during the evaluation of a proposition, Smith, Shoben and Rips predict no effect of false production frequency, per se, on RT. However, false production frequency results and ratings of subject-predicate similarity were found to correlate positively and significantly (r=.32). Since the feature comparison model expects false RT to be slowed as subject-predicate similarity increases, Smith, et. al., predict that RTs should be ordered anomalous to high production frequency predicates, fastest to slowest. The pattern of results for both universal and particular affirmative propositions confirm Holyoak and Glass's predictions. The magnitude of the difference between high and low production frequency predicates was 146 milliseconds for particular affirmative, and 152 milliseconds for universal affirmative propositions.

The second subexperiment was concerned with the counter-example hypothesis; thus, only universal affirmative propositions constitute appropriate stimuli. Stimuli were again selected on the basis of false completions given to incomplete propositions of the form: "All A are ____". The frequency with which particular completions were given constituted one of the two variables examined in this subexperiment; namely, production frequency (high or low). In this case, production frequency was interpreted solely as a manipulation of subject-predicate similarity. The second variable, counterexample frequency (high or low), was assessed on the basis of true completion responses for the same subject-concept, but quantified by "some". As an example, a high frequency sentence selected from the false completion norms might be: "All flowers are roses". The counterexample to this false assertion, which is the most common completion of "Some flowers are _____" after "roses" ("pansies"), is produced relatively infrequently. Thus, production frequency and counterexample frequency were manipulated orthogonally in this subexperiment.

Smith, et. al. (1974b) expect RT to vary as a function of production frequency because production frequency is positively correlated with similarity ratings. However, they predict no effect of counterexample frequency because it represents a manipulation of the accessibility of subset/superset information. Holyoak and Glass predict an effect of counterexample frequency alone because only subset/superset information will form the basis for a negative response according to their hypothesis. In fact, there was no effect of production frequency, but a significant effect of counterexample frequency was found. Low counterexample frequency propositions were an average of approximately 108 milliseconds slower than high counterexample frequency propositions. The main effect of counterexample frequency has been replicated by Holyoak and Glass (1975, Experiment 3); however, a high error rate (forty-three percent in one cell of the design) makes interpretation of the RT results of the study difficult.

While Holyoak and Glass have demonstrated an effect on RT to reject false propositions which cannot be attributed to subject-predicate similarity, some questions may be raised concerning an appropriate interpretation of their findings. Consider the contradiction subexperiment first. Holyoak and Glass interpret the results for high and low production frequency statements in this subexperiment as reflecting a common process: RT differences as a function of production frequency are attributed to differences in the speed with which the predicate concept is accessed in memory from the subject concept. The process of tracing the connection in memory between the two concepts is hypothesized to involve generating the immediate superordinate of the subject of the proposition, then finding that the predicate concept is an instance of the category as well. An examination of the stimuli used to test the contradiction hypothesis calls this conceptualization into question. Although it seems an appropriate description of the situation for low production frequency propositions, the majority of high production frequency propositions may be stored directly in memory as mutually-exclusive concepts. In other words, the fact that "A is not a B" may be stored in memory for high production frequency propositions; whereas this fact must be inferred for low production frequency propositions from the fact that "A" and "B" have the same immediate superordinate. (Some typical, high production frequency subject-predicate pairs are "chairs-tables", "men-women", and "boys-girls"; some typical, low production frequency subject-predicate pairs are "blossoms-grass", "chairs-desks", and "women-babies".) In terms of an interpretation of the contradiction subexperiment, what Holyoak and Glass may in fact have demonstrated are: (1) two different ways in which negative propositions are rejected; and (2) that "inferential" processing takes longer than retrieval of a direct, negative connection between the subject and predicate concepts.

A similar critique may be made with respect to the stimuli used in the counterexample subexperiment. High frequency counterexample predicates appear to be fairly homogeneous across levels of production frequency -- virtually all are instances of the subject-category and are of intermediate or low production frequency, as judged by reference to the Battig and Montague (1969) production frequency norms. However, low frequency counterexample predicates seem quite different as a function of false production frequency. High production frequency, low counterexample frequency predicates are the most typical instances of the subject-category. For example, if the subject-category is "flowers", the high frequency predicate is "roses" and the low frequency counterexample is presumably "pansies". On the other hand, low production frequency, low counterexample frequency predicates represent unusual partitions of the subject-category. For example, two predicates paired with the subject-category "fruits" were "citrus" and "spheres"; for "gems", they were "necklaces" and "earrings"; and for "teachers", they were "friends" and "parents". It is unlikely that many subjects had alternative partitions of these categories directly-stored in memory which could serve as easily accessible counterexamples to statements constructed from these subject-predicate pairs. (For example, what is an alternative partition of "fruits" representing a level of abstraction equivalent to that of "spheres"?) In terms of an interpretation of the counterexample subexperiment, the results may reflect that stimuli from various cells in the design may require very different rejection strategies.

Although the appropriate interpretation of Holyoak and

Glass's findings is unclear, their research may demonstrate an inadequacy of the feature comparison model (Smith, et. al., 1974b). It is not obvious how the model could be amended to handle systematic effects of variables other than subject-predicate similarity on negative RT. Smith, et. al. (1974b) suggest that the second processing stage, in which only defining features of the subject and predicate concepts are compared, may have to be altered from an exhaustive to a self-terminating comparison process. The process would terminate as soon as a defining feature of the subject-concept was encountered which was not also a defining feature of the predicate concept. However, the defining features of the subject- and predicate-concepts should be the same regardless of production frequency or counterexample frequency in the Holyoak and Glass study because the predicate is from the same category as the subject-concept (contradiction situation) or the predicate is an instance of the subject-category (counterexample situation). Thus, there is no basis for a self-terminating comparison process like that proposed by Smith, et. al. (1974b). The difficulty is Smith, et. al.'s claim that only attributes, and not subset/superset information is being evaluated in the sentence-verification paradigm.

While other theories of semantic memory must come to grips with Holyoak and Glass's findings, the theory Holyoak and Glass (1975) offer to account for their results is not consistent with the finding that subject-predicate similarity slows negative RT. They accept the finding that anomalous propositions are rejected very quickly as straightforward and recognize that their theory lacks a well-formulated mechanism to deal with this result, although some suggestions are offered. However, Holyoak and Glass do not view anomalous propositions as representing an extreme point on a continuum of subject-predicate similarity. Rather, they see anomalous propositions as distinct from "meaningful" propositions. Their position (Holyoak and Glass, 1975) is that the similarity/negative RT effect can be reduced to a case of either anomalous propositions being compared with meaningful propositions, or a confounding of similarity with counterexample frequency or production frequency (contradiction situation).

To summarize to this point, Holyoak and Glass have demonstrated that variables other than similarity affect subjects' RTs to reject false propositions. This finding raises problems for a feature comparison model of semantic memory and suggests that the processing of false propositions is a complicated affair. On the other hand, Holyoak and Glass's view of how false propositions are rejected fails to integrate the similarity/negative RT effect satisfactorily. Further, questions have been raised with respect to the appropriate interpretation of Holyoak and Glass's (1975) findings concerning their contradiction and counterexample hypotheses.

<u>Collins and Quillian: A Network Model</u>

As we have seen, there is evidence that subjects base negative decisions on information concerning properties shared by the subject- and predicate-concept of the stimulus proposition (e.g., Smith, <u>et. al.</u>, 1974b). There is also evidence that subjects reject some false propositions on the basis of information retrieved about the subset/superset relationships the subject- and predicate-concepts are involved in (e.g., Holyoak and Glass, 1975). The feature comparison model predicts the former, but not the latter effect; while Holyoak and Glass predict the latter effect, but cannot account for the semantic relatedness effect. Collins and Quillian present a model which suggests a way in which both of these effects might be accomodated. At its present state of development, the model is not sufficiently well-formulated to unambiguously predict the conditions of occurrence of either effect; however, it does offer an heuristic framework within which the two results might be considered further. Thus, those features of the model which provide the framework will be emphasized in the following discussion.

Three aspects of Collins and Quillian's (1969; 1970a; 1972a; 1972b; Collins and Loftus, 1975) theory are of interest. First, their conceptualization of the memory structure is important (Collins and Loftus, 1975). Concepts are represented as nodes, and relations as pathways between the nodes in the structure. Thus, memory is viewed as an interconnected network of information. The information in the network includes subset/ superset information as well as information about the attributes of concepts. Given that a particular concept is accessed in memory, information about that concept is obtained be tracing out along the paths which intersect at that concept-node and retrieving the associates which are encountered. This search process is undirected: The major processing assumption is that when a concept is accessed in memory, a parallel spread of activation is automatically triggered which results in the activation of neighboring concept-nodes. The speed with which these concepts are processed is determined by the strength of the pathway connecting them to the original concept, which, in turn, is presumably a function of factors such as the similarity and co-occurrence frequency of the two concepts. Finally, associates differ in their importance to the definition of a particular concept; thus pathways connecting concepts are tagged according to the "criteriality" of the various associates. Criteriality does not affect the speed of travel on the pathway, but serves as a weight indicating the importance of a particular piece of information during the evaluation of a proposition. Typically, subset/superset information is highly criterial information in Collins and Quillian's model, as it is in Holyoak and Glass's theory.

The second important aspect of Collins and Quillian's (1969; Collins and Loftus, 1975) concerns their conceptualization of how information is retrieved about a concept. Some of the major processing notions have been noted; what is necessary is an elaboration of what is hypothesized to occur when a test proposition is presented for evaluation. As soon as the subject and predicate concepts are encoded, a parallel spread of activation is initiated from each concept. Since this process is occurring within a network structure, there will be interconnecting pathways between the two concepts. When a common associate or intersection is found, the pathway interconnecting the subject and predicate is traced and evaluated against the test proposition. This brings us to the third important contention of Collins and Quillian (1972b; Collins and Loftus, 1975). When information relevant to the evaluation of a proposition is located in memory, it contributes to a positive or a negative decision according to the criteriality of the information. Thus, a false proposition is rejected on the basis of contradictory information found in memory, as in the Holyoak and Glass model. That is, negative decisions are based upon the positive retrieval of disconfirming information; rather than upon the failure to find a "match" between information in memory and the stimulus proposition, or between the subject- and predicate-concepts' feature lists.

According to the above model, the information in memory relevant to a particular test proposition might be viewed as an ordered list of propositions. Thus, the speed with which a false proposition is rejected is a function of at least two factors: (1) the length of the list (number of intersecting pathways between the subject- and predicate-concept); and (2) the position the disconfirming information occupies in the list.[4] In terms of

---

[4]A third factor suggested by the theory is the "distance" between successive items on the list, to continue the metaphor. Even if the disconfirming proposition to stimulus A and the disconfirming proposition to stimulus B occupy the same ordinal position on their respective lists, proposition A and proposition

predictions, the longer the "list", the slower negative RT will be in general. However, list-length will be a factor only to the extent that the position in the list of the contradictory information is affected. This position is consistent with that of Holyoak and Glass in that it: (1) emphasizes the importance of subset/superset information; and (2) declares the primary determinant of negative RT to be the accessibility of contradictory information. On the other hand, the "intersecting search hypothesis" generates the same, general prediction -- although by a different mechanism -- as the feature comparison model with respect to the effects of subject-predicate relatedness on negative RT. Both Collins and Quillian and Smith, et. al., predict increasing negative RT with increasing semantic relatedness. A point of distinction between the two positions is that Collins and Quillian claim that common associates of the subject and predicate are responsible for the effect and place the locus of effect in a search stage of processing; whereas Smith, et. al., attribute the effect to the number of shared features (similarity) of the two concepts being compared, placing emphasis on a comparison stage.

B may not be rejected with the same speed because the respective disconfirming propositions would occupy different positions on a ratio scale, or list. Consideration of this factor of inter-item distance on the list may well lead to theoretical predictions opposite to some generated by a consideration of ordinal position and list-length alone. For example, as subject-predicate similarity decreases, list-length decreases and, with it, negative RT. However, inter-item distance probably increases as similarity (list-length) decreases, leading to the prediction that negative RT should increase. Without a serious attempt to examine the relative importance of the "list-length" versus the "inter-item distance" variables, unambiguous predictions cannot be derived from the Collins and Quillian model. However, these problems will be set aside at present in order to make clear the heuristic framework provided by the model for integrating the semantic relatedness effect and Holyoak and Glass's findings.

All of the research reviewed to this point is relevant to the model being considered. The evidence supports the contention that contradictory information is used to reject false propositions (Anderson and Reder, 1974; Glass, et. al., 1974; Holyoak and Glass, Millward, et. al., 1975). Less directly, the intersecting search hypothesis is supported by evidence for the similarity/negative RT effect on the assumption that as subject-predicate similarity increases, the number of intersecting pathways linking the subject- and predicate-concepts in memory (list-length) will increase; thus more items will have to be evaluated, on the average, before a disconfirming piece of information is found. To avoid being redundant with previous sections, only one experiment of particular relevance to the present model will be considered here.

The study of interest is the counterexample subexperiment of Holyoak and Glass's (1975) Experiment 2. It is the only experiment in the literature that simultaneously manipulates the two factors -- accessibility of contradictory information and subject-predicate relatedness -- which the Collins and Quillian model hypothesizes to be the major determinants of negative RT. Recall that the predicted effect of counterexample frequency was found. The model also predicts a main effect of subject-predicate relatedness. The findings were inconclusive with respect to this prediction. There was no effect of relatedness; however, rated similarity of the subject and predicate concepts was shown to account for a significant portion of the variability after variability attributable to counterexample frequency was removed. This finding is consistent with the network model's contention that relatedness is important only to the extent that it affects the accessibility of disconfirming information.

In conclusion, the model derived from Collins and Quillian's work (1969; 1970a; 1970b; 1972a; 1972b; Collins and Loftus, 1975) is probably the most successful of the three theories considered in this review. Its success seems attributable primarily to two assumptions it makes which are not jointly common to either the Holyoak and Glass, or the Smith, _et. al._, theories: (1) attribute _and_ subset/superset information is stored with a concept in memory and is used in the process of evaluating a proposition; and (2) the search in memory for information about the relationship between the subject and predicate concepts of a test proposition proceeds from _both_ the subject _and_ the predicate concept, not just from the subject concept. Although Collins and Quillian's theory does a reasonable job of integrating the research findings produced thus far, it is not unambiguously supported. Further, it is not clear how much of its success is due to particular assumptions it adopts as opposed to a lack of specification at crucial points in the model. Thus, Collins and Quillian's position is perhaps most fruitfully considered in terms of its points of contrast with Holyoak and Glass's model and the feature comparison model of Smith, Shoben, and Rips. It is within this context that the research to be reported was developed.

THE RESEARCH

The two experiments to be reported may be interpreted from two perspectives. At one level, the experiments are an investigation of Holyoak and Glass's (1975) contradiction and counterexample hypotheses, respectively. At a more general level, the experiments represent a point of intersection between the three major theories of semantic memory which have been discussed.

First, it has been noted that Holyoak and Glass have not provided unambiguous support for their contradiction and counterexample hypotheses. It is not clear that their subjects followed the hypothesized "strategies" concerning how particular types of false statements are rejected. The present experiments attempt to clarify the situation by adopting somewhat different operational definitions of the relevant dependent variables. Experiment 2, which tests the counterexample hypothesis, also extends Holyoak and Glass's investigation by examining not one but two dependent variables which concern the distribution in memory of counterexamples to a proposition.

Despite some difficulties in interpreting exactly what kinds of strategies their subjects may be using, Holyoak and Glass demonstrate that false responses are based on the positive retrieval of information about the subset/superset relationships of the subject- and/or predicate-concepts of the stimulus propositions -- at least in some situations. At the same time, they raise some questions about the validity of the semantic relatedness/negative RT effect. However, their measure of semantic relatedness is unusual; thus it is difficult to evaluate the significance of their finding. The present experiments re-examine the semantic relatedness effect -- in the context of an investigation of the contradiction and counterexample hypotheses -- using a traditional measure of semantic relatedness.

The findings from the two experiments should also provide evidence relevant to a number of issues generated by a consideration of the three models of semantic memory. A comparison of the feature comparison model with Holyoak and Glass's, and with Collins and Quillian's positions raises two questions: What types of information are stored in memory?; How is that information organized? Contrasting Holyoak and Glass's model with that of Collins and Quillian raises the question of whether the retrieval process involves a search from both the subject- and predicate-concept in memory, or from the subject-concept only. Finally, a consideration of the three models taken together points up the issue of whether the retrieval process is directed such that only items bearing particular relationships to the subject- (and predicate) concept are searched; or whether the search is undirected. Of course, definitive answers cannot be given to any of these important questions. However, the following experiments should contribute to the existing literature bearing on these issues.

Experiment 1

Experiment 1 tests Holyoak and Glass's contradiction hypothesis. According to this hypothesis, when the subject- and predicate-nouns of a proposition represent instances from the same semantic category, the proposition will be rejected as soon as the subject (in the experiment) retrieves the information from memory that the two concepts have the same immediate superordinate. Holyoak and Glass hypothesize that retrieval of that information involves a search in memory from the subject-concept during which the superordinate of that concept is generated and the predicate concept is then found to be an instance of the same category. However, a question was raised as to whether this was an adequate description of how subjects processed all the propositions they received in Holyoak and Glass's (1975, Experiment 2) experiment; or whether two different bases for rejection were used corresponding to the two levels of false production frequency (Holyoak and Glass's measure of the accessibility of contradictory information). For instance, high production frequency propositions (e.g., "All men are women") may have been rejected by finding the information directly-stored in memory that the subject and predicate concepts are mutually-exclusive ("Men are not women"). On the other hand, low production frequency propositions (e.g., "All dogs are rabbits") may well have been rejected by the hypothesized process of generating the subject-concept's superordinate, finding that the predicate was an instance of that category, then making the appropriate inference. The present experiment attempts to avoid this problem by using different techniques of stimulus collection.

In addition to manipulating the accessibility of contradictory information, the rated similarity of the subject- and predicate-concepts is varied in Experiment 1. Although Holyoak and Glass considered the influence of this variable, they did so in a _post hoc_ manner and consequently their results were inconclusive with respect to the relative influences of the two variables examined -- production frequency and similarity.

Thus, the design of Experiment 1 is essentially a two by two orthogonal design, with subject-predicate similarity and the accessibility of contradictory information ("predicate dominance") the two independent variables. The task is a sentence-verification task in which the stimulus sentences are universal affirmative propositions ("All A are B") and reaction time is the dependent variable of primary interest. With respect to the three theories of semantic memory being considered, three different predictions are generated: (1) Holyoak and Glass predict a main effect of predicate dominance for two reasons. First, they hypothesize that the search process proceeds from the subject-concept only and is unaffected by the particular predicate-concept. Secondly, they expect that the only basis for rejecting a false proposition in this experiment should be the contradiction "strategy" since no anomalous items are included in the stimulus materials. Because only subset/superset information is hypothesized to be used to reject false propositions, and because the search process is presumed to proceed from the subject-concept only, similarity is predicted to have no effect. (2) Smith, Shoben, and Rips (1974b) predict a main effect of similarity. Predicate dominance is predicted to have no effect because it represents a manipulation of the accessibility of subset/superset information, information which the feature model claims is not used in a sentence-verification task. (3) Collins and Quillian predict main effects of both variables. Predicate dominance should affect RT for two reasons. One, the Collins and Quillian model allows for a rejection strategy based on subject-predicate similarity which is analogous to that of the feature model. The second reason is due to their conceptualization of the search process. Increasing subject-predicate similarity essentially increases the number of common associates of the two concepts, thus increasing the amount of information which will probably have to be evaluated before a rejection is possible.

Method

Design: The design is a two by two by twenty-six, within-subjects design. Two levels of subject-predicate similarity, two levels of predicate dominance, and twenty-six different subject-nouns define the partitioning of the stimulus materials. It is an unusual feature of the design of this experiment that there is no "items" variable nested within the four cells defined by the similarity and predicate dominance variables. Instead, the twenty-six subject-nouns serve as the subjects of four different propositions, one corresponding to each of the four cells. An "items" variable is best defined as the interaction of the three within-subjects variables -- similarity, predicate dominance, and subjects of the propositions.

Materials: There were three stages in the collection of the critical stimulus materials. First, groups of five nouns were selected from as many different semantic categories in the Battig and Montague (1969) norms as possible. Within each category, a noun of intermediate production frequency was selected to serve as a subject noun in the test propositions. Next, two pairs of nouns were selected from the same category with the constraints that (1) the items within a pair were of comparable production frequency and intuitively at different levels of similarity to the subject-noun and (2) that the pairs of nouns represented two extreme levels of production frequency. These four nouns would be paired with the subject-noun to create four propositions corresponding to the four, similarity by predicate dominance cells. For example, the subject-noun "lead" was paired with "steel" (high production frequency and high similarity); "gold" (high, low); "nickel" (low, high); and "bronze" (low, low). Forty-seven sets of four subject-predicate pairs were generated in this manner.

Next, the subject-predicate pairs were presented to nine subjects (graduate and undergraduate volunteers) to rate in semantic similarity. Each subject received a booklet consisting of pages with five or six sets of nouns on each page. Each group of five nouns chosen together in the first stage of stimulus collection were presented together on a page in the booklet. Predicate nouns were randomly ordered; noun sets were randomly assigned to a page; and pages were randomly presented to individual subjects. Subjects were required to rate how similar the subject-noun presented on the left-hand side of the page was to each of the four predicate nouns presented opposite to it on the right. A seven-point rating scale was used in which "1" was to be interpreted as "extremely similar" and "7" was "extremely dissimilar". On the basis of subject ratings, twenty-six sets of four noun pairs each were selected for use in the sentence-verification task. The resulting 104 stimulus pairs are presented in Appendix A; while their values on the independent variables and word-frequency (Kucera and Francis, 1967) are summarized in Appendix B. Note that the stimuli in each of the four cells of the design are approximately equated for frequency of English language usage and word length.

The final stage in the collection of stimulus materials was the generation of "true" items to serve as fillers for the above "false" propositions. For each of the twenty-six subject-nouns, an average of three (no less than two nor more than four) predicate terms were generated which would make a "true" statement. This was to insure that subjects would not be able to respond on the basis of the subject-noun alone after a few blocks of trials. In addition, fifty-six "true" items were constructed in which the predicate noun occupied the intuitively equivalent level in its respective semantic hierarchy as the subject-nouns in the critical false propositions. Further, twenty-four "false" filler items were constructed in which the predicate terms represented an abstract, relatively high level in their respective semantic hierarchies. The reasoning behind this control was that the predicate terms of the critical false propositions and their corresponding true propositions represent distinct populations with respect to the level they occupy in a semantic hierarchy, and possibly on an abstractness-concreteness dimension. Representative predicates for false items are "cherries", "lamps", and "plates"; for true items, they are "vegetables", "fruits", and "clothing". The additional true and false fillers were intended to abolish any systematic differences between true and false propositions that might encourage subjects to adopt a processing strategy irrelevant to the concerns of the present study.

Twenty stimulus propositions were randomly assigned to each block in the sentence-verification task with the following constraints: (1) two false propositions from each of the four cells of the experimental design appeared in each block; (2) two additional false fillers; (3) five or six true fillers with critical subject-nouns; and (4) five or four additional true fillers were represented in each block; and (5) a subject-noun appeared only once in a particular block. A total of twelve blocks of twenty trials each were generated in this manner and an additional two practice blocks were constructed which maintained the characteristics of the test blocks.

With respect to the two independent variables of interest in the study, the objective measure of similarity was, of course, the subject rating data. The predicate dominance variable is the objective measure of the accessibility of contradictory information about a critical proposition. The mean strength of the relationship between the subject-noun and its superordinate was equated across all four cells of the design by crossing the subject-noun with the similarity and predicate dominance variables. The strength of the relationship between the predicate noun and the same superordinate was varied by selecting two pairs of predicate nouns at two levels of production frequency (Battig and Montague, 1969) in their respective semantic category for each subject-noun.

Procedure: Subjects were instructed that they would be shown many simple sentences of the form "All A are B", and the task was to decide whether the statement was "true" or "false" as quickly as possible. If the statement was "true", subjects were to pull a response trigger with their right hand; the left-hand trigger corresponded to a "false" response.

Each subject was shown fourteen blocks of twenty propositions for evaluation. The first two blocks were practice blocks for which no data was recorded. The remaining twelve blocks were presented in a different random order for each subject and the sequencing of trials within each block was random as well. The sequence of events on each trial was as follows. First, the words "ALL ARE" appeared, "ALL" above "ARE", on a screen in front of the subject. After an interval of 400 milliseconds, a warning tone of 150 milliseconds in duration was presented, followed by a second silent interval of 400 milliseconds. Next, the subject-noun and predicate were presented, one above the other. The complete proposition looked as follows:

```
ALL
      S U B J E C T
ARE
      P R E D I C A T E
```

The proposition remained in view until a response was made, which erased the screen. If the response was correct, a two second interval followed before the start of the next trial. If the response was incorrect, the word "ERROR" replaced the proposition on the screen and the two-second interval did not begin until subjects indicated they were ready by pulling either response trigger. RT was measured from the presentation of the subject and predicate words. The presentation of trials and the sequencing of blocks and trials was controlled by a PDP-8I computer. Subjects were run individually and an experimental session lasted approximately forty-five minutes.

Subjects: Twenty undergraduate volunteers from an introductory psychology course participated in the experiment for course credit.

Results

Only data from critical false proposition trials were considered in analyses. The set of propositions for which the subject-noun was "ladles" was not included in analyses because a number of subjects reported that they had misread that subject-noun as "ladies". Thus, there were 100 data points per subject.

Error and RT data were each submitted to a two by two by twenty-five, within-subjects analysis of variance. Similarity and predicate dominance were considered fixed-effects variables (Clark, 1973). However, since the definition of subject-nouns as a random-effects factor means that tests of treatment effects are conservative (Clark, 1973), the results of analyses in which treatment effects were tested against the subject-by-treatments term will be reported when they differ from the results of quasi-F computations. Quasi-F ratios were computed according to the formula suggested by Myers (1972). In addition to the results for the critical, false propositions reported below, a summary of the results for the false and true filler items is presented in Appendix C. Results of error analyses are reported first.

Errors: The overall error rate for the experiment was 8.55 percent. A summary of the error data is presented in Table 1 on the following page. As can be seen, errors are positively correlated with reaction time. The analysis of variance demonstrated a number of significant effects. The subjects and subject-nouns effects were significant ($F(19,456)=14.50$, $p<.001$; $F(24,456)=6.08$, $p<.001$, respectively). In addition, the interactions of similarity with subject-nouns ($F(24,456)=5.80$), predicate dominance with subject-nouns ($F(24,456)=6.28$), and similarity by predicate dominance by subject-nouns ($F(24,456)=5.20$) were all significant at the .001 level. These interactions and the main effect of subject-nouns taken together indicate significant variability attributable to the stimulus materials.

The theoretically interesting observations are the main effects of similarity and predicate dominance. The effect of similarity was significant ($F'(1,23)=17.87$, $p<.01$). There were more errors when subject-predicate similarity was high (13.2%) than when it was low (2.9%). The effect of predicate dominance did not reach significance according to quasi-F computations ($F'(1,23)=1.21$, $p>.1$), but was significant when tested against the subjects-by-predicate dominance term ($F(1,19)=6.95$, $p<.05$). There were more errors with low predicate dominance (10.0%) than with high (7.1%). The interaction of similarity with predicate dominance was not significant ($F'<1$), nor were any other effects.

Table 1: Mean reaction times and errors (in parentheses) for Experiment 1.

		Similarity	
		High	Low
Predicate Dominance	High	1475.2 (13.0%)	1260.1 (1.2%)
	Low	1600.0 (15.4%)	1338.9 (4.6%)

Reaction time: For purposes of the analysis of variance, errors were replaced with the subjects' mean RT for the respective similarity-by-predicate dominance cell of the design. Mean RTs are presented in Table 1.

The results of the reaction time analysis are consistent with the results of the error analysis. Again, the main effects of subjects ($F(19,456)=22.52$, $p<.001$) and subject-nouns ($F(24,456)=3.19$, $p<.001$) are significant. The similarity-by-subject nouns ($F(24,456)=3.68$), predicate dominance-by-subject nouns ($F(24,456)=3.30$), and similarity-by-predicate dominance-by-subject nouns ($F(24,285)=2.56$) interactions all are significant at the .001 level.

The main effects of both similarity ($F'(1,27)=20.21$, $p<.001$) and predicate dominance ($F'(1,37)=4.70$, $p<.05$) are significant. Subjects are 102 milliseconds faster to reject high predicate dominance propositions, and are 238 milliseconds faster to reject propositions of low subject-predicate similarity. The only remaining significant effect is the interaction of similarity with subjects ($F(19,456)=1.77$, $p<.05$). The significant interaction is attributable primarily to variability in the size of the similarity effect between subjects rather than to differences in its direction. Nineteen of twenty subjects showed slower reaction times when similarity was high than when it was low.

Discussion

At this point, the results of Experiment 1 will be interpreted only with respect to the contradiction hypothesis of Holyoak and Glass. The implications of the results for the three models of semantic memory reviewed and the issues they raise will be postponed until the second experiment is considered.

Holyoak and Glass's contradiction hypothesis predicts that when the subject and predicate of a proposition are instances of the same semantic category, subjects' recognition of this fact is the basis for their rejection of the proposition. The present study varies the accessibility of this information by varying the strength of the relationship between the predicate instance and its immediate superordinate (predicate dominance), while holding the subject-noun constant. The finding that RT and errors increase as predicate dominance decreases thus supports the contradiction hypothesis.

The finding that similarity had a large effect on both RT and error rate refutes Holyoak and Glass's suggestion that previous demonstrations of slower negative RT with increasing subject-predicate similarity (e.g., Smith, et. al., 1974b; Glass, et. al., 1974) are due to: (1) a confounding of similarity with variables determining the accessibility of superset/subset information common to the subject- and predicate-concepts of a proposition; and/or (2) comparisons of RTs to reject meaningful "false" propositions (high similarity) with anomalous propositions (low similarity), which presumably represent a distinct population of stimuli. Neither criticism can be applied to the present experiment. However, while the similarity effect in the present experiment does not appear attributable to a confounding with other potentially significant factors, the locus of its effect is not clear. Two possibilities are suggested by the feature comparison model and Collins and Quillian's theory, respectively. First, the effect of similarity may indicate that subjects do use a processing strategy similar to that hypothesized by Smith, et. al. (1974b). That is, judgments of subject-predicate similarity are being used as a basis for response in the experimental task. An observation which is inconsistent with this hypothesis is that the RTs for errors on critical false items are quite long (2023 milliseconds) and thus could not be based on a fast computation of subject-predicate similarity as the feature comparison model suggests. The second possibility is that similarity has its effect not on the subjects' processing strategies, but at a more fundamental level of processing. Specifically, given an hypothesized search process which locates any information common to the subject- and predicate-concept during the evaluation of a proposition (e.g., Collins and Quillian, 1969), high similarity will slow negative RT because much irrelevant information will have to be considered before some relevant information is found which will result in the proposition's rejection. Unfortunately, the present experiments are not designed to distinguish conclusively between these alternatives, although there will be further discussion of this issue in the general discussion section.

Experiment 2

Experiment 2 tests Holyoak and Glass's counterexample hypothesis. This hypothesis states that when subjects are presented with a universal proposition in which the subject term is the superordinate of the predicate term (e.g., "All animals are dogs"), they will reject the assertion on the basis of finding in memory an instance of the subject-category which is not the predicate-concept (e.g., "cat"). Holyoak and Glass hypothesize that the primary determinant of RT in this situation is accessibility of the most accessible counterexample, which is measured in the present experiment by production frequency (Battig and Montague, 1969). Although Holyoak and Glass have provided evidence for the counterexample hypothesis, questions were raised concerning whether there were qualitative differences within their stimulus sample corresponding to the various cells of their experimental design. For instance, one cell (high production frequency, high counterexample frequency) included the items: "All birds are robins" and "All fruits are oranges"; while another cell (low, low) included stimuli which partitioned the same subject-category in unusual ways -- ways which subjects may not have stored directly in memory: "All birds are flyers" and "All fruits are spheres". The potential problem is that these two types of false propositions may require different rejection strategies. For example, the first two statements may be rejected by the strategy proposed by Holyoak and Glass. On the other hand, the latter two statements seem to require an extra processing step: subjects do not have stored alternatives to the above two predicate partitions at comparable levels of abstraction -- they probably do not explicitly categorize a subclass of nonspheres or nonflyers, for instance. Consequently, they probably must retrieve instances of the subject-categories and then determine whether each instance may be classified as a member of the predicate-category as well. Alternatively, rather than retrieving exemplars which may or may not turn out to be counterexamples of the proposition, subjects may use a more "analytic" rejection strategy for these abstract propositions. For example, they might reason that "citrus" denotes a class of fruits, therefore there must be other types of fruits. The present experiment attempts to insure that subjects use a consistent rejection strategy by making "counter-example frequency" a between-categories variable, thus allowing similar predicate-category partitions of all the semantic categories used as subject-nouns.

The experiment is also concerned with a second variable which may influence time to retrieve a counterexample to a false proposition. While the primary definition of counterexample frequency is that proposed by Holyoak and Glass -- the frequency of occurrence of the most frequently produced instance of the subject-category (Battig and Montague, 1969); the number of highly accessible counterexamples to a statement may affect RT. Thus a second definition of counterexample frequency adopted is the number of instances of a subject-category with a production frequency of fifty percent or higher (Battig and Montague, 1969).

Finally, the analog of the similarity variable of Experiment 1 is included in the present experiment. "Typicality" (of the predicate as an instance of the subject-category) is included as a variable for the same reasons as in Experiment 1.

The predictions of the three theories of semantic memory are the same as in Experiment 1: (1) Holyoak and Glass predict a main effect of counterexample frequency only; (2) the feature comparison model predicts a main effect of typicality only; and (3) the network model of Collins and Quillian predicts main effects of both variables.

Method

Design: The design of the present experiment is the same for each definition of counterexample frequency. In each case, the design is essentially a two-by-two, within subjects design. Typicality and counterexample frequency cross and are considered fixed effects. In addition, twenty-four propositions are nested within each cell defined by the crossing of typicality and counterexample frequency. This "items" variable is considered a random effects variable, as is the subjects variable.

Materials: Stimulus materials were selected from the Battig and Montague (1969) production frequency norms. The first step in the collection procedures involved identifying all the categories in the norms which could be labelled with a familiar, one-word title. (However, the categories "cities", "states" and "colleges" were excluded from consideration because it was felt that there would be too much variability in the normative structures of these categories between different subject populations.) Thirty categories were selected by this criterion to serve as the subject-nouns in the critical stimulus propositions. They were divided into two groups of fifteen categories each corresponding to the two levels of counterexample frequency (obviously, which categories fell into which level depended upon which definition of counterexample frequency was being examined at the time).

The next step was to select predicate-instances to be paired with the subject-categories. As many pairs of instances were selected as could be found for each of the thirty subject-categories which satisfied the following criteria: (1) the two instances be of approximately equivalent, and preferably intermediate levels of production frequency; (2) the two instances be intuitively of two different levels of "typicality" as examples of the category from which they were selected; (3) no instance was to be among the four most-frequently given instances of its category (only one predicate was produced by more than fifty percent of Battig and Montague's subjects -- "potato" was given as an instance of the category "vegetables" by fifty-one percent of their subjects).

After pairs of instances were selected, stimuli were prepared for presentation to subjects who were to rate instances for typicality. Categories and instances were grouped together and the order of instances was randomized within each category group. Category groups were randomly assigned to pages, six or seven groups to a page, and pages were randomized for each separate booklet. The same nine subjects who did the similarity ratings in Experiment 1 did the typicality ratings in Experiment 2. Subjects were asked to rate how typical each instance was as a member of the category it was paired with. A seven point scale was used in which "1" meant "extremely typical" and "7" meant "extremely atypical". On the basis of these data, one or two pairs of predicate-instances were selected to be paired with each subject-category selected in the first stage of stimulus collection. Thus, the two groups of stimuli already classified according to counterexample frequency were further subdivided into two groups on the basis of typicality ratings, completing the four cells of the basic experimental design: high counterexample frequency, high typicality (e.g., "All furniture are divans"); high, low (furniture - stereos); low, high (ships - yachts); and low, low (ships - canoes). A total of ninety-six category-instance pairs were generated, twenty-four per cell. From these items, ninety-six propositions were created of the form: "All category are instance". Critical stimuli used in Experiment 2 and summaries of their characteristics are presented in Appendices F through I.

After the critical stimuli were constructed, "filler" items were generated. Since the same subject-noun was used to construct two or four critical false propositions, some true items had to be generated which used the same subject-nouns. This was a difficult task since the subject-nouns represented semantic categories at relatively high levels of abstraction. For example, it is difficult to generate four different predicates which make the incomplete proposition "All professions are _____" true. Only half (forty-three) as many true fillers could be generated as there were critical false items. Since no comparisons were planned between "true" and "false" responses, the two-to-one ratio of false to true items was accepted.

In addition to the above filler items, sixteen more false and eight more true filler propositions were constructed. These additional items were included for the same reasons as in Experiment 1: (1) to prevent subjects from responding on the basis of predicate terms alone, which were primarily abstract (high level superordinates) for the true items and concrete concepts (common objects) for the critical false stimuli; and (2) to prevent the same predicate term from appearing repeatedly in true propositions only (e.g., "objects" was the predicate for three true propositions).

After the stimuli had been generated, eight test blocks of twenty-one trials each were constructed. Each block contained: (1) twelve critical false propositions, three from each of the four cells of the design (the primary definition of counterexample frequency was adopted in defining the cells); (2) two false fillers; and (3) seven true fillers. Stimulus propositions were randomly assigned to blocks with these constraints plus the additional limitation that a particular subject of a proposition appear no more than once in a block. Finally, an additional two blocks of twenty-one trials each were constructed in accordance with the characteristics of the test blocks to serve as practice blocks in the sentence-verification task.

Procedure: The procedure was the same as that for Experiment 1 with the exceptions that there were only eight rather than twelve test blocks; there were twenty-one instead of twenty trials per block; and there was no warning tone presented during trials.

Subjects: Twenty-one subjects from an introductory psychology course participated for course credit.

Results

The results will be reported in two parts corresponding to the two definitions of counterexample frequency. A summary of the results for the true and false filler propositions is presented in Appendix J.

Definition 1 (single counterexample): This section reports the results when counterexample frequency is defined as the production frequency of the most frequently given instance of the subject-category. These results are summarized in Table 2.

Table 2: Mean reaction times and errors (in parentheses) for the primary definition of counterexample frequency in Experiment 2.

		Typicality	
		High	Low
Counterexample Frequency	High	1290.4 (2.7%)	1214.7 (0.8%)
	Low	1294.9 (3.6%)	1265.9 (2.0%)

Errors: The overall error rate was only 2.23 percent, but the pattern of errors was consistent between subjects. The error data was analyzed with counterexample frequency and typicality considered fixed effects, and subjects and items nested within counterexample frequency and typicality considered random effects. The analysis of variance showed significant effects of subjects ($F(20,1840)=2.33$, $p<.01$), items ($F(92,1840)=1.46$, $p<.05$), and typicality ($F'(1,29)=5.74$, $p<.025$). There were more errors when typicality was high (3.6%) than when it was low (1.2%). The effect of counterexample frequency was not significant ($F'(1,29)=1.92$, $p>.1$). None of the interactions were significant.

Reaction times: The reaction time data was submitted to a two (counterexample frequency) by two (typicality) by twenty-four (items) by twenty-one (subjects) analysis of variance. Counterexample frequency, typicality, and items nested within counterexample frequency and typicality were within-subjects variables.

Counterexample frequency and typicality were considered fixed effects; items and subjects were considered random effects variables. Errors were replaced with the subject's mean RT for the appropriate counterexample frequency by typicality cell.

The results of the RT analysis were consistent with those of the error analysis. The main effects of subjects ($F(20,1795)=41.51$, $p<.001$) and items ($F(92,1795)=1.96$, $p<.001$) were significant. With respect to the variables of interest, two sets of F-ratio calculations were performed for each variable. First, quasi-F (Myers, 1972) ratios were calculated for each variable. If the result was significant at the .05 level or beyond, the effect was considered significant. However, if the effect failed to reach significance by quasi-F computations, a second criterion for acceptance was applied. Forster and Dickinson (1976) have demonstrated that under some conditions, quasi-F ratios represent conservative approximations to standard alpha levels. One such condition cited by Forster and Dickinson (1976) occurs when the subjects-by-treatments effect and/or the main effect of "items" (two, potential error terms against which an effect might be tested) are small. In this situation, they suggest that the most accurate test of whether an effect is significant ($p<.05$) or not is to test the effect of interest against the subjects-by-treatments effect and against the effect of items. If both F-ratios exceed the .05 level, the effect is considered significant. This second criterion will be referred to as the joint criterion (Forster and Dickinson, 1976).

By quasi-F computations, the main effect of typicality is marginally significant ($F'(1,42)=3.77$, $p<.07$). Since the subject-by-typicality interaction effect does not deviate significantly from zero ($F(20,1795)=1.10$, $p>.1$), the joint criterion for acceptance was applied. The effect of typicality is significant when tested against both the subjects-by-typicality effect ($F(1,20)=7.08$, $p<.01$) and the effect of items ($F(1,92)=3.96$, $p<.05$). Subjects were fifty-two milliseconds slower when typicality was high than when it was low.

The effect of counterexample frequency does not attain significance by either criteria: neither when tested by the quasi-F procedure ($F'(1,4)=1.46$, $p>.1$); nor when tested against the subjects-by-counterexample frequency effect ($F(1,20)=3.56$, $p<.08$) or items effect ($F(1,92)=1.22$, $p>.1$). No other effects approached significance.

Definition 2 (many counterexamples): This section gives the results when counterexample frequency is defined as the number of instances of the subject-category with a production frequency value of fifty percent or higher in the Battig and Montague norms (1969). For this definition of counterexample frequency, some of the propositions switched levels of counterexample frequency from the first definition of counterexample frequency. The items within levels of typicality stayed the same, of course, The results are summarized in Table 3 on the following page.

Errors: The error data were submitted to an analysis of variance in which subjects and items were considered random effects and typicality and counterexample frequency were fixed effects. All treatment effects were considered within-subjects

Table 3: Mean RT and errors (in parentheses) for the second definition of counterexample frequency in Experiment 2.

		Typicality: High	Typicality: Low
Counterexample Frequency	High	1259.7 (2.2%)	1213.3 (0.8%)
	Low	1324.4 (4.0%)	1267.1 (2.0%)

variables.

Both main effects of items and subjects were signficant ($F(92,1840)=1.43$, $p<.05$; $F(20,1840)=2.33$, $p<.01$, respectively). The effect of typicality was significant ($F'(1,28)=5.88$, $p<.025$); while the effect of counterexample frequency was marginally significant by quasi-F computations ($F'(1,29)=3.46$, $p<.08$). The conjoint criterion test is applicable here (the subjects-by-counterexample frequency interaction was not significant ($F(20, 1840)=1.09$, $p>.1$). When tested against the subjects-by-counterexample frequency term, the effect was significant ($F(1,20)=4.84$, $p<.05$). However, when tested against the items term, counterexample frequency is only marginally significant ($F(1,92)=3.69$, $p<.07$). There tends to be more errors when counterexample frequency is low (3.0%) than when it is high (1.5%).

Reaction times: Errors were replaced with the subject's mean RTs for the appropriate cell of the design for the purpose of the analysis of the reaction time data. The analysis of variance model was the same as that used for all prior analyses. The results are generally consistent with those for the error analysis. Subjects ($F(20,1795)=41.88$, $p<.001$) and items ($F(92,1795)=1.93$, $p<.001$) were both significant effects.

The effect of typicality was marginally significant by quasi-F computations ($F'(1,41)=4.01$, $p<.06$). Again, the subjects-by-typicality effect is small ($F(20,1795)=1.07$), $p>.1$), thus the joint criterion for significance was applied. The typicality effect is significant when tested against the subjects-by-typicality effect ($F(1,20)=7.49$, $p<.05$) and when tested against the items term ($F(1,92)=4.17$, $p<.05$).

A similar situation applies for the effect of counterexample frequency. The effect of counterexample frequency is marginally significant by quasi-F computations ($F'(1,37)=3.85$, $p<.08$). The subjects-by-counterexample frequency effect is significantly different from zero ($F(20,1795)=1.78$, $p<.05$), but is relatively small.[5] When counterexample frequency is tested against the subjects-by-counterexample frequency effect, it is significant ($F(1,20)=5.86$, $p<.025$). Counterexample frequency is also significant when tested against the items term ($F(1,92)=5.40$, $p<.025$). Subjects were fifty-nine milliseconds faster to reject high counterexample frequency propositions than they are to reject low counterexample frequency propositions. No other effects were significant.

A summary of the results of Experiment 2 is in order. First, high typicality slows RT and results in more errors. Second, when counterexample frequency is defined as the production frequency of the most frequently given instance of the subject-

[5] The variability attributable to the subjects-by-counterexample frequency effect is small: θ_{SxCF} = 2,753 (Myers, 1972). By way of comparison, the combined contribution of error and the subjects-by-items components amounts to 169,772.

category, the effect of counterexample frequency is not significant (although errors and RT tend to decrease as counterexample frequency increases). However, when counterexample frequency is defined in terms of the number of highly-accessible counterexamples to a proposition, the main effect of counterexample frequency is significant in the RT analysis: RT is faster and errors tend to be fewer when counterexample frequency is high than when it is low. Finally, a significant interaction of subjects and counterexample frequency (definition 2) qualifies interpretation of the main effect of counterexample frequency: one third of the subjects were slower when counterexample frequency was high than when it was low.

Discussion

At this point, the discussion will be restricted to an interpretation of the results with respect to Holyoak and Glass's counterexample hypothesis.

Holyoak and Glass contend that universal propositions in which the subject term is a semantic category to which the predicate belongs (e.g., "All animals are cats") are rejected when a counterexample to the statement is found in memory. This hypothesis is supported by the finding that subjects are faster to reject propositions for which there are many, highly-accessible counterexamples, and they tend to make fewer errors in doing so. The nature of the support for the counterexample hypothesis presented here is different from that reported by Holyoak and Glass. Holyoak and Glass found an effect of counterexample frequency defined in terms of the production frequency of the single, most accessible instance of the subject-category. The present experiment analyzed the data according to this definition of counterexample frequency with inconclusive results. It is clear that the number of frequently produced instances of the subject-category, rather than the frequency of the most frequently produced instance, was the better predictor of RT in the present experiment. What is not clear is whether this result is indicative of the nature of the processes underlying performance in the task; or whether the former normative measure is simply a more stable predictor of the speed with which an individual subject will retrieve a single counterexample to a false proposition.

The finding that typicality slows the rejection of false propositions is consistent with the findings of Experiment 1 and with the literature in general (e.g., Smith, et. al., 1974b; Glass, et. al., 1974). The interpretation of this result is the same as that given for the similarity effect found in Experiment 1; namely, that the effect is a legitimate one. The question of the locus of the effect was raised in the discussion section of Experiment 1. The present experiment allows some speculation on the two suggested alternatives. However, since the fundamental question of the nature of the memory search process is involved, this discussion will be presented in the final section of the paper.

Finally, the two-way interaction of counterexample frequency with subjects deserves some note. This result suggests that subjects may have used different strategies of processing in performing the experimental task. This would not be surprising given that two-thirds of the propositions were "false" and in almost all of the false items the predicate term was an instance of the category denoted by the subject term. This fact naturally produced a strong bias to respond "false" (subjects took an average of 1516 milliseconds and had a 16.1 percent error rate on true propositions), which probably acted to "level" the counterexample frequency and typicality effects. It would be interesting to investigate the context sensitivity of the interaction effects by using an equal number of true and false propositions and including additional false items which must be rejected on grounds other than the counterexample strategy.

Both experiments have been described and the implications of their results have been considered with respect to two rejection strategies proposed by Holyoak and Glass. It is now time to consider the general theoretical implications of the findings of the two experiments.

GENERAL DISCUSSION

A consistent package of results emerges from Experiments 1 and 2. Each experiment supports the hypothesis it is designed to test. Experiment 1 provides evidence for the contradiction hypothesis, while Experiment 2 supports the counterexample hypothesis. In doing so, they demonstrate that subjects utilize information about the subset/superset relationships existing between the subject and predicate terms in the process of evaluating a proposition. Both experiments also demonstrate effects of subject-predicate similarity on negative RT: as subject-predicate similarity increases, RT and errors increase as well. Overall, this pattern of results supports Collins and Quillian's network model of semantic memory while demonstrating deficiencies in Holyoak and Glass's conceptualization and in the feature comparison model of Smith, Shoben and Rips. The interesting questions concern the nature of the deficiencies: What makes the Collins and Quillian model successful? Although the present study is not designed to provide a conclusive answer to this question, it allows some speculation.

Memory Structure

The general issue of the structure of semantic memory is addressed by the findings of the study. This issue may be set in the context of a comparison of Collins and Quillian's theory and the feature comparison model.

Contrary to the feature comparison model's contentions, the present experiment indicates that information about a concept's subset/superset relationships is stored in memory and is used in a sentence-verification task. Further, this information must be organized such that rapid access is provided to it. Both the contradiction and the counterexample strategy are based on the use of information about indirect relationships between the subject and predicate of a proposition. The contradiction strategy uses information about the subject and predicate's relationship to a third concept, their common superordinate. The counterexample strategy is based on the use of information related directly to the subject-concept only. A feature model of semantic memory structure is well-adapted to explain decisions based upon the evaluation of direct relationships between explicitly presented concepts, but it cannot easily provide an account of decisions based on indirect relationships between two concepts. The major difficulty would seem to be its requirement that concepts be processed in terms of their elementary concepts at the level which labels the concept, e.g., "Winston Churchill" and "Great Britain". Rather, all relationships between concepts must be inferred from "match-mismatch" relationships between the features of which they are composed. Thus the feature model's conceptualization of the structure of semantic memory places great restrictions on the flexibility of any processing system operating within it.

The success of Collins and Quillian's model thus seems attributable -- at least in part -- to the fact that it does not make a structural distinction between "features" and "concepts"; thus it allows much greater processing flexibility than does the feature comparison model. A concept is described not only in terms of its features, but also in terms of its relationships to other concepts. "Features" are simply concepts with particular types of relationships to other concepts.

Memory Process

The study also bears on issues of semantic memory process.[6] Specifically, the results are relevant to questions concerning retrieval and decision processes in semantic memory. We can be fairly certain that the effects demonstrated in the two experiments were located in these later stages of processing as opposed to an encoding stage of processing for several reasons. First, word length and frequency of English language usage was controlled across levels of the independent variables in each experiment. Second, although Meyer (1973a) has demonstrated semantic-relatedness effects at encoding, the present experiments show an effect of similarity opposite to that demonstrated by Meyer: negative RT was inhibited by subject-predicate similarity, not facilitated. Finally, the effects of similarity are much larger than those reported by Meyer, at least in Experiment 1 (238 milliseconds).

While the locus of effects in the two experiments is thus attributable to later stages of processing, it is difficult to isolate the locus more narrowly. Specifically, the effects could be at a memory retrieval stage, or at a decision stage of processing, or both.

---

[6]The model of processing implicitly accepted in this discussion is one which hypothesizes a series of relatively well-distinguished stages. Those stages and the order in which they handle information are: encoding; memory search, including some sort of mechanism to evaluate retrieved information; decision; and response execution.

Subject-predicate similarity (or typicality) affected reaction time in both experiments. Two alternative explanations of this effect have been offered: subject-predicate similarity has its effect on decision certainty (e.g., Smith, et. al., 1974b); or similarity slows negative RT by resulting in the retrieval of much information which is irrelevant to the evaluation of the proposition (e.g., Collins and Quillian, 1969).[7] Two observations argue against the hypothesis that similarity was the basis for decisions in the present study. First, according to the feature comparison model, all errors result from first-stage processing -- errors occur when the degree of subject-predicate similarity (high or low) contradicts the usual subject-predicate relationship for a proposition with a particular truth value. Thus, error RTs should be short, but in both experiments they were quite long. Second, if a processing strategy is based upon rapid judgment of subject-predicate similarity, similarity must be confounded with truth value if the strategy is to be at all successful. Although no subject ratings of subject-predicate similarity were obtained for true propositions, it is unlikely that similarity and truth value were confounded in either experiment since the range of similarity for false propositions was extreme. Thus, it seems unlikely that information about subject-predicate similarity

---

[7]Some models of the decision process are consistent with Collins and Quillian's hypothesis; specifically, those decision models which state that irrelevant information is not filtered before a decision stage of processing is entered. For present purposes, I have adopted a working model which breaks down processing into separate evaluation and decision stages where the evaluation stage decides whether the information is relevant or irrelevant and the decision stage simply decides whether the relevant information is sufficient for a positive or negative

formed the basis for a decision in the present experiments; rather, it represented readily accessible, but largely irrelevant information.

The preceding analysis suggests that much irrelevant information is retrieved during the evaluation of a proposition. This observation, in turn, suggests that the search process governing retrieval of information from memory is not restricted within narrow limits in the sense that at least two types of information about subject- and predicate-concepts are being retrieved in the experimental task. Subjects process information concerning both the similarity (irrelevant "property" information) and subset/superset (relevant) relationships of the subject- and predicate-concepts. Other researchers have also suggested that subjects have little control over the directions their memory searches take. For example, subjects do not seem capable of restricting the search process to newly-acquired information about a concept in an episodic memory task; rather, the search process is evidently influenced by pre-experimental associations as well as experimentally-introduced associations (Anderson, 1975; Perlmutter, Harsip, and Myers, 1976). Also, when asked to produce instances of a category under time pressure, subjects sometimes mistakenly produce related noninstances even though they are quickly aware of their errors (Collins and Loftus, 1975). These findings indicate an

---

response. Beyond this consideration however, the distinction to be made between Smith, Shoben, and Rips and Collins and Quillian is that the former model hypothesizes that similarity information forms the basis for semantic decisions; while Collins and Quillian consider similarity information to be irrelevant -- in most cases -- to such decisions.

undirected search process (or search process with very little control) which seems inefficient in that it results in the retrieval of much irrelevant information. In the present study, a more efficient, directed search process would have considered only subset/superset relationships between the subject- and predicate-concepts. However, while such a strategy is logically possible in most categorization experiments, it would not be feasible in most linguistic (conservational) situations. This is because a directed search process depends upon the very information to give it direction that a search process is designed to locate so that a statement may be comprehended and evaluated; that is, it assumes the very knowledge it is designed to find.

Turning to the remaining effect of interest in Experiments 1 and 2, respectively, the most straightforward interpretation of the predicate dominance and counterexample frequency effects is that they occur at the retrieval stage. This is because both variables represent a manipulation of the accessibility of information which disconfirms a proposition. However, there is some indication in Experiment 2 that the decision stage may be indirectly responsible for the counterexample frequency effect also. Although the retrieval of a single counterexample logically should be sufficient for the rejection of a proposition, it was found that the number of highly accessible counterexamples to a proposition (definition 2) was a better predictor of RT than the production frequency of the single most frequent counterexample (definition 1). This finding may indicate nothing more than that definition 2 was a better measure of the production frequency of the single most frequent counterexample for a random, individual subject. However, if the results are not artifactual, they may indicate that subjects are cautious in their decision-making, often delaying their response to a proposition until a couple of counterexamples have been retrieved. On the other hand, perhaps one counterexample is sufficient for a decision and the findings of Experiment 2 can be explained entirely in terms of the nature of the search process.

Definition 1 of counterexample frequency implicitly assumes that the retrieval of information about a proposition involves a search from the subject-concept only and is unaffected by the particular predicate-concept with which it is paired (Holyoak and Glass, 1975). Neither the feature comparison model (Smith, et. al., 1974b) nor the network model (Collins and Quillian, 1969) accepts this assumption, proposing instead that the search proceeds from both the subject- and predicate-concept. If this is so, then the first instance of the subject-category retrieved may not be the single most frequent instance of the category, but the instance most strongly associated to the subject-category and predicate-instance combined. This single counterexample may or may not be the most frequent instance of the category (definition 1), but more likely than not would be among the top few instances of the category (definition 2).

Loftus (1973) has conducted a categorization experiment which lends some support to the hypothesis that the search process is bidirectional, proceeding simultaneously from the subject- and predicate-concept in memory. She found that instance-dominance (the strength of the association from a category to an instance) and category-dominance (the strength of the instance-to-category association) each account for significant portions of the positive RT variance. If the unidirectional search hypothesis was correct, only category-dominance should have affected RT. In the present study, the finding that subject-predicate relatedness affected RT even though the accessibility of subset/superset (disconfirming) information (predicate dominance and counterexample frequency) was held constant supports the bidirectional search hypothesis.

A bidirectional search mechanism would seem to necessitate more processing capacity than a unidirectional search mechanism. However, the advantages of the bidirectional search outweigh this apparent disadvantage. Given that many semantic interpretations are available for a particular lexical item, a bidirectional search is well-adapted to the task of identifying the intended sense of a word because it is designed to quickly determine what semantic connections exist between lexical items in a sentence, for example. By way of contrast, a unidirectional search does not take into account the fact that it is trying to establish a connection between a subject-concept and a particular predicate. Rather, a search originating from a particular concept in memory will locate associations to that concept in the same order regardless of the predicate- paired with the subject-concept. A second, related advantage of a bidirectional search concerns its ability to use context to disambiguate words. Specifically, a bidirectional search provides for the use of following information to disambiguate preceding lexical items. On the other hand, a unidirectional search requires that comprehension proceed only in a forward direction during the processing of linguistic information. In sum, a bidirectional search is clearly a more flexible search mechanism for linguistic processing than a unidirectional search mechanism. Further, there is some empirical support for the bidirectional search hypothesis in the work of Loftus (1973) and in the finding of the present study that subject-predicate relatedness affected RT even though the accessibility of disconfirming information (measured in a production task) was held constant across levels of relatedness.

To conclude, the present study has provided evidence for two ways in which false propositions are rejected in a sentence-verification task (Holyoak and Glass, 1975). Also, it substantiates the similarity/negative RT effect reported by other investigators (e.g., Smith, et. al., 1974b; Glass, et. al., 1974). Further, some circumstantial evidence has been collected concerning some fundamental questions about the nature of semantic memory structure and process. However, the experiments raise more questions than they answer. Some of the important, interrelated issues to be investigated are: How conscious are the "strategies" of rejection investigated in this study?; Can their use be manipulated by altering the context within which they are called?; Can other rejection strategies be identified?; If, as suggested, the search process is undirected and proceeds from both the subject- and predicate-concept simultaneously, then what is the nature of the mechanism which evaluates retrieved information and decides upon a response?

APPENDICES

Appendix A

Critical Stimuli for Experiment 1

Subject	High Dominance, High Similarity	High Dominance, Low Similarity	Low Dominance, High Similarity	Low Dominance, Low Similarity
Lead	Steel	Gold	Nickel	Bronze
Jaguars	Tigers	Cows	Panthers	Monkeys
Blizzards	Hurricanes	Tornadoes	Gales	Twisters
Benches	Sofas	Lamps	Divans	Pictures
Surfing	Swimming	Tennis	Diving	Hunting
Admirals	Generals	Privates	Marshals	Cadets
Bankers	Lawyers	Doctors	Judges	Artists
Skunks	Cats	Elephants	Beavers	Turtles
Suicides	Murders	Rapes	Homicides	Treason
Polo	Tennis	Football	Pool	Boxing
Prunes	Grapes	Peaches	Figs	Mangos
Vests	Shirts	Socks	Suits	Belts
Trolleys	Busses	Trucks	Subways	Sleds
Banjos	Violins	Trumpets	Fiddles	Bugles
Piccolos	Flutes	Pianos	Recorders	Bongos
Dolphins	Sharks	Trout	Marlins	Shrimps
Ladles	Spoons	Forks	Cups	Plates
Axes	Hammers	Nails	Hatchets	Ladders
Raisins	Cherries	Bananas	Berries	Melons
Termites	Ants	Bees	Ticks	Hornets
Turnips	Potatoes	Corn	Beets	Parsley
Girdles	Pants	Blouses	Shorts	Boots
Arrows	Knives	Rifles	Lances	Missiles
Buzzards	Eagles	Robins	Falcons	Jays
Finches	Sparrows	Eagles	Swallows	Chickens
Parsnips	Carrot	Beans	Turnips	Parsley

Appendix B

Summary of Stimulus Characteristics for Experiment 1

		Dominance High	Dominance Low	
Similarity	High	288	25	157
	Low	274	24	149
		281	25	

Mean Battig-Montague values (1969) for the predicate nouns in Experiment 1. Each number represents the mean number of subjects producing the predicate noun as an instance of its respective category in the Battig and Montague (1969) study. These figures represent the measure of predicate dominance in Experiment 1.

		Dominance High	Dominance Low	
Similarity	High	2.94	2.93	2.93
	Low	5.39	5.50	5.45
		4.16	4.22	

Mean subject-predicate similarity ratings for Experiment 1.

		Dominance High	Dominance Low	
Similarity	High	42.1	34.5	38.5
	Low	32.1	31.0	31.6
		37.1	32.8	

Mean frequency of English language usage (Kucera and Francis, 1967) of the predicate nouns in Experiment 1.

Appendix C

Summary of Results for Filler Items in Experiment 1

Reaction Time	Number of Errors
1352 milliseconds	39 (8.0%)

Summary of results for the false, filler propositions in Experiment 1.

Reaction Time	Number of Errors
1396 milliseconds	195 (7.5%)

Summary of results for the true propositions in Experiment 1.

Appendix D

Analysis of Variance of Error Data for Experiment 1

Source of Variance	df	MS	F	p
SIMILARITY (fixed)	1	6.38	17.87*	.001
DOMINANCE (fixed)	1	.42	1.21*	
PROPOSITION (random)	24	.39	6.08	.001
SUBJECTS (random)	19	.87	14.50	.001
SIM x DOM	1	.01	.04	
SIM x PROP	24	.36	5.80	.001
SIM x S	19	.06	.89	
DOM x PROP	24	.34	6.28	.001
DOM x S	19	.06	1.11	
PROP x S	456	.06		
SIM x DOM x PROP	24	.32	5.20	.001
SIM x DOM x S	19	.05	.75	
SIM x PROP x S	456	.06		
DOM x PROP x S	456	.05		
SIM x DOM x PROP x S	456	.06		
	1999			

*These F-ratios are based on quasi-F computations (Myers, 1972).

Appendix E

Analysis of Variance of Reaction Time Data for Experiment 1

Source of Variance	df	MS	F	p
SIMILARITY (fixed)	1	21,316,789	20.21*	.001
DOMINANCE (fixed)	1	4,835,763	4.70*	.05
PROPOSITION (random)	24	703,671	3.19	.001
SUBJECTS (random)	19	4,970,224	22.52	.001
SIM x DOM	1	12,475	.02	
SIM x PROP	24	873,212	3,68	.001
SIM x S	19	418,679	1.77	.05
DOM x PROP	24	784,681	3.30	.001
DOM x S	19	244,179	1.03	
PROP x S	456	220,735		
SIM x DOM x PROP	24	773,917	2.56	.001
SIM x DOM x S	19	156,969	.52	
SIM x PROP x S	456	237,047		
DOM x PROP x S	456	237,668		
SIM x DOM x PROP x S	285**	301,839		
	1828			

*These F-ratios are based on quasi-F computations (Myers, 1972).
**The degrees of freedom were calculated as 456 minus the number of errors (replaced reaction times), 171.

Appendix F

Critical Stimuli for Experiment 2: First Definition of Counterexample Frequency

Subject, High Dominance	Subject, Low Dominance	Predicate, High Typicality	Predicate, Low Typicality
Furniture		Divans	Stereos
		Rockers	Mirrors
Colors		Maroon	Indigo
		Navy	Beige
Gems		Topaz	Onyx
		Opals	Jade
Relatives		Parents	Grandsons
Tools		Files	Pencils
		Chisels	Rulers
Fruits		Berries	Prunes
		Nectarines	Pumpkins
Animals		Wolves	Giraffes
		Panthers	Camels
Vehicles		Vans	Sleds
Weapons		Pistols	Rope
		Daggers	Poisons
Cloth		Denim	Burlap
		Flannel	Mohair
Dwellings		Duplexes	Tepees
		Cabins	Igloos
Sports		Softball	Fencing
Trees		Cedars	Palms
Crimes		Homicides	Speeding
Alcohol		Rye	Brandy
	Utensils	Ladles	Ovens
		Tongs	Sponges
	Birds	Larks	Penguins
		Doves	Chickens
	Professions	Nursing	Plumbing
		Banking	Farming
	Metals	Nickel	Mercury
	Clothing	Suits	Scarves
		Slacks	Gloves
	Music	Concert	Dixieland
	Insects	Locusts	Worms
	Money	Bills	Bonds
		Coins	Lira
	Instruments	Bugles	Bells
		Cellos	Harps
	Fuels	Charcoal	Uranium
	Vegetables	Potatoes	Lettuce
	Elements	Cobalt	Acids
	Toys	Marbles	Bicycles
		Rattles	Bikes
	Fish	Flounders	Shrimp
		Haddock	Lobsters
	Ships	Yachts	Canoes
		Clippers	Barges

Appendix G

Summary of Stimulus Characteristics for Experiment 2: First Definition of Counterexample Frequency

		Dominance High	Dominance Low	
Typicality	High	36	47	41
	Low	41	43	42
		38	45	

Mean Battig-Montague (1969) values for the predicate nouns in Experiment 2.

		Dominance High	Dominance Low	
Typicality	High	1.93	1.81	1.87
	Low	4.14	4.46	4.30
		3.03	3.13	

Mean ratings of the typicality of the predicate instances with respect to the category nouns with which they are paired.

		Dominance High	Dominance Low	
Typicality	High	16.1	21.0	18.6
	Low	7.7	11.9	9.8
		11.9	16.5	

Mean frequency of English language usage (Kucera and Francis, 1967) of the predicate nouns in Experiment 2.

	Dominance High	Dominance Low
Word Frequency	73.6	76.0
Definition 1, Counterexample Frequency	413.6	318.1
Definition 2, Counterexample Frequency	7.1	6.3

Summary of the characteristics of the subject (category) nouns used in Experiment 2.

Appendix H

Critical Stimuli for Experiment 2:
Second Definition of Counterexample Frequency*

Category	Number of Instances*	Category	Number of Instances
Colors	11	Utensils	6
Relatives	11	Tools	6
Clothing	11	Fuels	6
Fruits	10	Insects	6
Instruments	10	Money	6
Vegetables	9	Trees	6
Animals	8	Gems	5
Furniture	8	Music	5
Alcohol	8	Dwellings	4
Birds	8	Crimes	4
Vehicles	8	Ships	4
Fish	8	Weapons	3
Metals	7	Professions	3
Cloth	7	Elements	3
Sports	7	Toys	3

*According to the second definition of counterexample frequency, categories (propositions) are partitioned on the basis of the number of instances produced by fifty percent or more of the subjects in the Battig and Montague (1969) category norms. Since the instances paired with the categories remain the same as for the first definition of counterexample frequency, they are not repeated here.

Appendix I

Summary of Stimulus Characteristics for Experiment 2: Second Definition of Counterexample Frequency

		Dominance High	Dominance Low	
Typicality	High	42	41	41
	Low	44	41	42
		43	41	

Mean Battig-Montague (1969) values for the predicate nouns in Experiment 2.

		Dominance High	Dominance Low	
Typicality	High	1.83	1.90	1.87
	Low	4.05	4.57	4.31
		2.94	3.24	

Mean ratings of the typicality of the predicate instances with respect to the category nouns with which they are paired.

		Dominance High	Dominance Low	
Typicality	High	15.0	22.1	18.6
	Low	7.3	13.0	10.1
		11.1	17.5	

Mean frequency of English language usage (Kucera and Francis, 1967) of the predicate nouns in Experiment 2.

	Dominance High	Dominance Low
Word Frequency	66.3	83.8
Definition 1, Counterexample Frequency	379.9	351.8
Definition 2, Counterexample Frequency	8.7	4.7

Summary of the characteristics of the subject (category) nouns used in Experiment 2.

Appendix J

Summary of Results for Filler Items in Experiment 2

Reaction Time	Number of Errors
1516 milliseconds	190 (16.1%)

Summary of results for true propositions in Experiment 2.

Appendix K

Analysis of Variance of Error Data for Experiment 2:
Definition 1 of Counterexample Frequency

Source of Variance	df	MS	F	p
Counterexample Frequency (fixed)	1	.06002	1.92*	
Typicality (fixed)	1	.14335	5.74*	.05
Items/FT (random)	92	.03086	1.46	.05
Subjects (random)	20	.04926	2.33	.01
FT	1	.00050	.02*	
FS	20	.02148	1.02	
TS	20	.01523	.72	
FTS	20	.01820	.86	
SI/FT	1840	.02113		
	2015			

*These F-ratios are based on quasi-F computations (Myers, 1972).

Appendix L

Analysis of Variance of Reaction Time Data for Experiment 2:
Definition 1 of Counterexample Frequency

Source of Variance	df	MS	F	p
Counterexample Frequency (fixed)	1	412,143	1.46*	
Typicality (fixed)	1	1,336,065	3.77*	.07
Items/FT (random)	92	337,525	1.96	.001
Subjects (random)	20	7,141,155	41.51	.001
FT	1	270,286	.71*	
FS	20	115,909	.67	
TS	20	188,776	1.10	
FTS	20	217,499	1.26	
SI/FT	1795**	172,025		
	1970			

*These F-ratios are based on quasi-F computations (Myers, 1972).
**The degrees of freedom were calculated as 1840 minus the number of errors (replaced RTs), 45.

Appendix M

Analysis of Variance of Error Data for Experiment 2:
Definition 2 of Counterexample Frequency

Source of Variance	df	MS	F	p
Counterexample Frequency (fixed)	1	.11161	3.46*	.07
Typicality (fixed)	1	.14335	5.88*	.05
Items/FT (random)	92	.03027	1.43	.05
Subjects (random)	20	.04926	2.33	.01
FT	1	.00446	.16*	
FS	20	.02307	1.09	
TS	20	.01523	.72	
FTS	20	.01801	.85	
SI/FT	1840	.02111		
	2015			

*These F-ratios are based on quasi-F computations (Myers, 1972).

Appendix N

Analysis of Variance of Reaction Time Data for Experiment 2:
Definition 2 of Counterexample Frequency

Source of Variance	df	MS	F	p
Counterexample Frequency (fixed)	1	1,769,205	3.85*	.08
Typicality (fixed)	1	1,365,417	4.01*	.06
Items/FT (random)	92	327,640	1.93	.005
Subjects (random)	20	7,109,367	41.88	.001
FT	1	15,038	.04*	
FS	20	301,937	1.78	.05
TS	20	182,240	1.07	
FTS	20	213,965	1.26	
SI/FT	1795**	169,772		
	1970			

*These F-ratios are based on quasi-F computations (Myers, 1972).
**The degrees of freedom were calculated as 1840 minus the number of errors (replaced RTs), 45.

REFERENCES

Anderson, J.R. Item-specific and relation-specific interference in sentence memory. Journal of Experimental Psychology, 1975, 104, 249-261.

Anderson, J.R., and Bower, G.H. Human Associative Memory. Washington, D.C.: V.H. Winston, 1973.

Anderson, J.R., and Reder, L.M. Negative judgments in and about semantic memory. Journal of Verbal Learning and Verbal Behavior, 1974, 13, 664-681.

Anderson, R.C., and Ortony, A. On putting apples into bottles -- a problem of polysemy. Cognitive Psychology, 1975, 7, 167-181.

Battig, W.F., and Montague, W.E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut Category norms. Journal of Experimental Psychology Monograph, 1969, 80, (3, Pt. 2).

Carpenter, P.A., and Just, M.A. Sentence comprehension: A psycholinguistic processing model of verification. Psychological Review, 1975, 62, 45-73.

Clark, H.H. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 335-359.

Clark, H.H., Cohen, J., Smith, J.E.K., and Keppel, G. Discussion of Wike and Church's comments. Journal of Verbal Learning and Verbal Behavior, 1976, 15, 257-266.

Collins, A.M., and Loftus, E.F. A spreading activation theory of semantic processing. Psychological Review, 1975, 82, 407-428.

Collins, A.M., and Quillian, M.R. Retrieval time from semantic memory. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 240-247.

Collins, A.M., and Quillian, M.R. Does category size affect categorization time? Journal of Verbal Learning and Verbal Behavior, 1970a, 9, 432-438.

Collins, A.M., and Quillian, M.R. Facilitating retrieval from semantic memory: The effect of repeating part of an inference. In A.F. Sanders (Ed.), Attention and Performance III. Amsterdam: North-Holland Publishing Company, 1970b, 304-314.

Collins, A.M., and Quillian, M.R. How to make a language user. In E. Tulving and W. Donaldson (Eds.), Organization of Memory. New York: Academic Press, 1972a, 310-349.

Collins, A.M., and Quillian, M.R. Experiments on semantic memory and language comprehension. In L.W. Gregg (Ed.), Cognition in Learning and Memory. New York: John Wiley and Sons, Inc., 1972b, 117-137.

Collins, A.M., Warnock, E.H., Aiello, N., and Miller, M.L. Reasoning from incomplete knowledge. Bolt, Beranek and Newman Report No. 3019, March, 1975.

Conrad, C. Cognitive economy in semantic memory. Journal of Experimental Psychology, 1972, 92, 149-154.

Fiksel, J.R., and Bower, G.H. Question-answering by a semantic network of parallel automata. Journal of Mathematical Psychology, 1976, 13, 1-45.

Forster, K.I., and Dickinson, R.G. More on the language-as-fixed-effect fallacy: Monte Carlo estimates of error rates for F_1, F_2, F', and min F'. Journal of Verbal Learning and Verbal Behavior, 1976, 15, 135-142.

Foss, D.J., and Harwood, D.A. Memory for sentences: Implications for Human Associative Memory. Journal of Verbal Learning and Verbal Behavior, 1975, 14, 1-16.

Freedman, J.L., and Loftus, E.F. Retrieval of words from long-term memory. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 107-115.

Freedman, J.L., and Loftus, E.F. Retrieval of words from well-learned sets: The effect of category size. Journal of Experimental Psychology, 1974, 102, 1085-1091.

Frijda, N.H. Simulation of human long-term memory. Psychological Bulletin, 1972, 77, 1-30.

Glass, A.L., and Holyoak, K. The effect of "some" and "all" on reaction time for semantic decisions. Memory and Cognition, 1974, 2, 336-340.

Glass, A.L., Holyoak, K., and O'Dell, C. Production frequency and the verification of quantified statements. Journal of Verbal Learning and Verbal Behavior, 1974, 13, 237-254.

Gough, P.B. Grammatical transformations and speed of understanding. Journal of Verbal Learning and Verbal Behavior, 1965, 5, 107-111.

Grober, E., and Loftus, E.F. Semantic memory: Searching for attributes versus searching for names. Memory and Cognition, 1974, 2, 413-416.

Hayes-Roth, B., and Hayes-Roth, F. Plasticity in memorial networks. Journal of Verbal Learning and Verbal Behavior, 1975, 14, 506-522.

Henley, N.M. A psychological study of the semantics of animal terms. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 176-184.

Hollan, J.D. Features and semantic memory: Set-theoretic or network model? Psychological Review, 1975, 82, 154-155.

Holyoak, K., and Glass, A.L. The role of contradictions and counterexamples in the rejection of false sentences. Journal of Verbal Learning and Verbal Behavior, 1975, 14, 215-239.

Juola, J.F., and Atkinson, R.C. Memory scanning for words versus categories. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 522-527.

Kucera, H., and Francis, W.N. Computational analysis of present-day English. Providence, R.I.: Brown University Press, 1967.

Landauer, T.K., and Meyer, D.E. Category size and semantic memory retrieval. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 539-549.

Loftus, E.F. Activation of semantic memory. American Journal of Psychology, 1973, 86, 331-337.

Loftus, E.F. Category dominance, instance dominance, and categorization time. Journal of Experimental Psychology, 1973, 97, 70-74.

Loftus, E.F., and Cole, W. Retrieving attribute and name information from semantic memory. Journal of Experimental Psychology, 1974, 102, 1116-1122.

Loftus, E.F., and Freedman, J.L. Effect of category-name frequency on the speed of naming an instance of the category. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 343-347.

Loftus, E.F., Freedman, J.L., and Loftus, G.R. Retrieval of words from subordinate and superordinate categories in semantic hierarchies. Psychonomic Science, 1970, 21, 235-236.

Loftus, E.F., Senders, J.W., and Turkletaub, S. The retrieval of phonetically similar and dissimilar category members. American Journal of Psychology, 1974, 87, 57-63.

Loftus, E.F., and Scheff, R.W. Categorization norms for 50 representative instances. Journal of Experimental Psychology, 1971, 91, 355-364.

Loftus, E.F., and Suppes, P. Structural variables that determine the speed of retrieving words from long-term memory. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 770-777.

Loftus, G.R., and Loftus, E.F. The influence of one memory retrieval on a subsequent memory retrieval. Memory and Cognition, 1974, 2, 467-471.

McFarland, C.E., Kellas, G., Klueger, K., and Juola, J.F. Category similarity, instance dominance, and categorization time. Journal of Verbal Learning and Verbal Behavior, 1974, 13, 698-708.

Meyer, D.E. Components of semantic-information processing in memory. Paper presented at the annual meeting of the Psychonomic Society, St. Louis, November, 1969.

Meyer, D.E. On the representation and retrieval of stored semantic information. Cognitive Psychology, 1970, 1, 242-300.

Meyer, D.E. Correlated operations in searching stored semantic categories. Journal of Experimental Psychology, 1973a, 99, 124-133.

Meyer, D.E. Verifying affirmative and negative propositions: Effects of negation on memory retrieval. In S. Kornblum (Ed.), Attention and Performance IV. New York: Academic Press, 1973b, 379-394.

Meyer, D.E. Long-term memory retrieval during the comprehension of affirmative and negative sentences. In Wilkes and Kennedy (Eds.), Studies in Long-Term Memory. London: John Wiley and Sons, 1975.

Meyer, D.E., and Ellis, G.B. Parallel processes in word-recognition. Paper presented at the annual meeting of the Psychonomic Society, San Antonio, Texas, November, 1970.

Meyer, D.E., and Ruddy, M.G. Lexical-memory retrieval based on graphemic and phonemic representations of printed words. Paper presented at the annual meeting of the Psychonomic Society, St. Louis, November, 1973.

Meyer, D.E., and Ruddy, M.G. Bilingual word-recognition: Organization and retrieval of alternative lexical codes. Paper presented at the meeting of the Eastern Psychological Association, Philadelphia, April, 1974.

Meyer, D.E., and Schvaneveldt, R.W. Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. Journal of Experimental Psychology, 1971, 90, 227-234.

Meyer, D.E., Schvaneveldt, R.W., and Ruddy, M.G. Activation of lexical memory. Paper presented at the annual meeting of the Psychonomic Society, St. Louis, November, 1972.

Meyer, D.E., Schvaneveldt, R.W., and Ruddy, M.G. Loci of contextual effects on visual word-recognition. In P. Rabbitt (Ed.), Attention and Performance V. New York: Academic Press, 1974a.

Meyer, D.E., Schvaneveldt, R.W., and Ruddy, M.G. Functions of graphemic and phonemic codes in visual word-recognition. Memory and Cognition, 1974b, 2, 309-321.

Millward, R.B., Rice, G., and Corbett, A. Category production measures and verification times. In A. Kennedy and A. Wilkes (Eds.), Studies in Long-Term Memory. New York: John Wiley and Sons, 1975.

Myers, J.L. Fundamentals of Experimental Design. Boston: Allyn and Bacon, Inc., 1972.

Neisser, U., and Beller, H.K. Searching through word lists. British Journal of Psychology, 1965, 56, 349-358.

Norman, D.A., and Rumelhart, D.E. Explorations in Cognition. San Francisco: W.H. Freeman and Company, 1975.

Perlmutter, J., Harsip, J., and Myers, J.L. The role of semantic knowledge in retrieval from episodic long-term memories: Implications for a model of retrieval. Memory and Cognition, 1976, 4, 361-368.

Perlmutter, J., and Myers, J.L. Spreading activation: A demonstration of inhibitory effects on retrieval. Report 74-2, Cognitive Processes Laboratory, UMass-Amherst, September, 1974, 67-78.

Perlmutter, J., Sorce, P., Harsip, J., and Myers, J.L. Further consideration of a two-trace strength model of retrieval: The effect of interpolated activity in recall. Report 74-2, Cognitive Processes Laboratory, UMass-Amherst, September, 1974, 1-29.

Perlmutter, J., Sorce, P., and Myers, J.L. Retrieval processes in recall. Cognitive Psychology, 1976, 8, 32-63.

Rips, L.J. Quantification and semantic memory. Cognitive Psychology, 1975, 7, 307-340.

Rips, L.J., Shoben, E.J., and 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., and Shoben, E.J. Set-theoretic and network models reconsidered: A comment on Hollan's "Features and Semantic Memory". Psychological Review, 1975, 82, 156-157.

Rosch, E. On the internal structure of perceptual and semantic categories. In T.E. Moore (Ed.), Cognitive Development and the Acquisition of Language. New York: Academic Press, 1973.

Rosch, E. Cognitive reference points. Cognitive Psychology, 1975, 7, 532-547.

Rosch, E., and Mervis, C.B. Family resemblances: Studies in the internal structure of categories. Cognitive Psychology, 1975, 7, 573-605.

Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M., and Boyes-Braem, P. Basic objects in natural categories. Cognitive Psychology, 1976, 8, 382-439.

Rubinstein, H., Garfield, L., and Millikan, J.A. Homographic entries in the internal lexicon. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 487-494.

Rubinstein, H., Lewis, S.S., and Rubinstein, M.A. Evidence for phonemic recoding in visual word recognition. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 645-657.

Rumelhart, D.W., Lindsay, P.H., and Norman, D.A. A process model of long-term memory. In E. Tulving and W. Donaldson (Eds.), Organization of Memory. New York: Academic Press, 1972.

Sanford, A.J., and Seymour, P.H.K. Semantic distance effects in naming superordinates. Memory and Cognition, 1974, 2, 714-720.

Schaeffer, B., and Wallace, R. Semantic similarity and the comparison of word meanings. Journal of Experimental Psychology, 1969, 82, 343-346.

Schaeffer, B., and Wallace, R. The comparison of word meanings. Journal of Experimental Psychology, 1970, 86, 144-152.

Schvaneveldt, R.W., and Meyer, D.E. Retrieval and comparison processes in semantic memory. In S. Kornblum (Ed.), Attention and Performance IV. New York: Academic Press, 1973.

Schvaneveldt, R.W., Meyer, D.E., and Becker, C.A. Lexical ambiguity, semantic context, and visual word recognition. Manuscript in preparation, 1975.

Smith, E.E. Effects of familiarity on stimulus recognition and categorization. Journal of Experimental Psychology, 1967, 74, 324-332.

Smith, E.E., Haviland, S.E., Buckley, P.B., and Sack, M. Retrieval of artificial facts from long-term memory. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 583-593.

Smith, E.E., Rips, L.J., and Shoben, E.J. Semantic memory and psychological semantics. In G.H. Bower (Ed.), The Psychology of Learning and Memory, Vol. 8. New York: Academic Press, 1974a.

Smith, E.E., Shoben, E.J., and Rips, L.J. Structure and process in semantic memory: A feature model for semantic decisions. Psychological Review, 1974b, 81, 214-241.

Sternberg, S. The discovery of processing stages: Extensions of Donders' method. In W.G. Koster (Ed.), Attention and Performance II. Amsterdam: North-Holland Publishing Company, 1969, 276-315.

Tulving, E. Episodic and semantic memory. In E. Tulving and W. Donaldson (Eds.), Organization of Memory. New York: Academic Press, 1972.

Walker, J.H. Real-world variability, reasonableness judgments, and memory representations for concepts. Journal of Verbal Learning and Verbal Behavior, 1975, 14, 241-252.

Warren, R.E. Stimulus encoding and memory. Journal of Experimental Psychology, 1972, 94, 90-100.

Wike, E.L., and Church, J.D. Comments on Clark's "The language-as-fixed-effect fallacy". Journal of Verbal Learning and Verbal Behavior, 1976, 15, 249-256.

Wilkins, A.J. Conjoint frequency, category size, and categorization time. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 382-385.