

## Negative Judgments In and About Semantic Memory<sup>1</sup>

JOHN R. ANDERSON AND LYNNE M. REDER

*University of Michigan*

This research is concerned with determining how subjects falsify statements like *A collie is a cat*. A multiple regression analysis was performed which used 22 variables to try to predict the negative judgment times. The predictive variables are time to generate the superordinate of the instance (for example, *dog*), time to falsify that the superordinate is the predicate (for example, *A dog is a cat*), and time to encode the instance. This finding and others indicate that a prominent negation strategy is one in which the subject generates the superset of an instance and falsifies that the superset is the predicate. Auxiliary regression analyses are also reported for other reaction time measures gathered in the experiments. It is argued that large-scale regression experiments are critical to the inferential logic of a semantic memory experiment.

In the past five years or so, the topic of semantic memory has become a very active area of experimental research. This paper presents a relatively different approach in trying to understand subjects' performance in semantic memory tasks. Several of the major theoretical analyses in this field are discussed first. We raise some interpretive problems with earlier studies, and with these considerations in mind, offer our own method for investigating the semantic memory paradigm and analyzing its data.

The study of Collins and Quillian (1969) was instrumental in stirring interest in semantic memory. They proposed that our categorical knowledge was stored in a strict hierarchical fashion. They claimed, for instance, that we have directly stored *Collies are dogs* and *Dogs are mammals* but not *Collies are mammals*. To verify this last fact, a subject would have to retrieve the first two and with these infer the truth of the third. This proposal has been called the cognitive economy hypothesis because it reduces memory requirements to a logical minimum. The

hypothesis predicts subjects should be faster to verify facts like *Collies are dogs*, which can be directly retrieved, than facts like *Collies are mammals*, which require inferences. This prediction was verified by Collins and Quillian (1969).

The problem with this outcome is that it is subject to multiple interpretations. For instance, Anderson and Bower (1973, Chap. 12) argue that subjects do store information like *Collies are mammals* even when that information is redundant with other information in memory. However, information like *Collies are dogs* tends to be more available than *Collies are mammals*. This is because one fact is more frequently thought of than another. For instance, *dog* occurs more frequently as a superset to *collie* in a constrained association task (Loftus & Scheff, 1971). Glass, Holyoak, and O'Dell (1974) have found such generation frequencies to be good predictors in semantic memory tasks.

Another explanation (for example, Meyer, 1970; Landauer & Meyer, 1972) is that the Collins and Quillian results are an artifact of the sizes of the categories *mammal* versus *dog*. Because *dog* has fewer instances it can be more quickly searched to see if *collie* is a member. Meyer and Ellis (1970) found category size effects in judgments about whether

<sup>1</sup> This research was supported by Grants MH-24360-1 from the National Institute of Mental Health and GB-40298 from the National Science Foundation. We would like to thank Rebecca Paulson for her help in many phases of this research.

a nonword like *mafer* is a member of a category. In such a situation subjects could not be searching for *mafer* in a conceptual hierarchy.

Another explanation has to do with the similarity (measured in terms of feature overlap) between *collie* and *dog* versus the similarity between *collie* and *mammal*. Concepts that tend to be more distant in a conceptual hierarchy have fewer features in common. They tend to be rated as less similar by subjects (Rips, Shoben, & Smith, 1973). Schaeffer and Wallace (1970) propose subjects prime a positive response (lower their threshold for saying true) when two concepts are similar. Smith, Shoben, and Rips (in press) propose that a second stage of a two stage model is either invoked or not on the basis of a first-stage similarity judgment. They find ratings of similarity are better predictors of judgment times than distance in a conceptual hierarchy.

There have been two developments in response to this multiplicity of theories. The first has been the recognition of the diagnostic value of negative judgments. For instance, how do subjects judge a probe like *A St. Bernard is a cat* or *A typhoon is a wheat*? The general result is that subjects are slower the more similar the two concepts or the closer they are in conceptual hierarchy (Meyer, 1970; Collins & Quillian, 1972; Smith *et al.*, in press). Thus subjects are slower in falsifying the first negative example above. This result is completely counter to what Collins and Quillian had originally predicted. The result does correspond nicely to the predictions of the similarity model since similar negatives would tend to bias a positive response. Unelaborated, the associative availability and category size models do not make any predictions with respect to similarity. However, the conditions in Meyer's (1970) model for obviating stage two act functionally as a test for semantic similarity; Holyoak and Glass (unpublished) also provide an explanation for the similarity effect within an associative frequency model.

#### *Interpretive Problems*

The other recent development in semantic memory has been a recognition of the interpretative problems posed by the nature of a semantic memory experiment (for example, Anderson & Bower, 1973; Clark, 1973; Landauer & Meyer, 1972). The basic problem is that the experimenter cannot randomly construct the material that he will present to his subject; he must select his stimuli from what the language makes available. The typical strategy is to partition materials into conditions on the basis of some property they possess such as similarity or category size. If differences are found between conditions, there is no way to discover whether this was due to the prescribed variable or some unknown variable that covaried with it. We have noted how hierarchical distance, associative availability, category size, and similarity tend to covary making the original Collins and Quillian result uninterpretable. There are many other variables that might be underlying the observed effects.

The basic point is that the underlying logic of a semantic memory experiment is not that of a true experiment with controlled manipulation of an independent variable; rather, it is correlational. That is, there is some variable, such as similarity, which naturally varies over stimulus materials. An experimenter obtains reaction time judgments to these materials and determines how the naturally occurring variable correlates with the experimentally obtained reaction times. Despite the frequent bias in experimental psychology against correlational research, this approach is capable of being used for developing and testing scientific theories. It has been used to great effect in other fields, for example, economics. However, if correlational research is to be useful it must be recognized as such and analyzed in a sophisticated manner. One necessary step to achieving this sophistication is to abandon past procedures which looked at only one or two variables; instead one must perform multiple regression analyses

that consider simultaneously the effects of many partially covarying variables to determine which are the effective predictors of reaction time. We do not mean to imply that by simply examining dozens of variables we will come to an understanding of semantic memory. Careful theoretical analysis and model testing is required. However, a preliminary to these more focused enterprises is simply the task of finding out what variables are the effective predictors in a semantic memory experiment.

Loftus and Suppes (1972) provide an example of the use of multiple regression to analyze a semantic memory task. In their experiment, subjects were required to generate an instance of a specified category that began with a specified letter. They used twelve dependent variables to predict generation times. These variables included measures of frequency of category, frequency of the correct instance, and dominance of the instance in Battig-Montague norms. Battig-Montague dominance was the best predictor. This supports the Anderson and Bower and the Glass *et al.* models which relate reaction time to production frequency. Frequency measures were also good predictors. However, Loftus and Suppes did not use the measure of similarity which Smith *et al.* claim should be more effective.

Another development in semantic memory is the recent statistical critique provided by Clark (1973). He pointed out that typically in semantic memory experiments words are grouped into a number of conditions depending on variables such as category size. The statistical test usually employed attempts to assess whether the differences between conditions are reliable across subjects, and not whether they are reliable over the words in the conditions. Clark calls this sort of analysis the language-as-fixed-effects fallacy. He recommends a procedure whereby the statistical test simultaneously generalizes over materials and subjects.

As Clark, points out, however, this procedure is only appropriate when the ma-

terials are randomly sampled. Verbal learning experiments exemplify those paradigms in which random selection of stimuli for conditions is possible: Study time, number of presentations, or any number of other factors can be used with any sets of words. In a semantic memory experiment, however, one never really samples materials at random from the total population available. Rather one chooses materials that constitute the experimental conditions in a somewhat haphazard way. The sampling is necessarily haphazard because the total population of stimuli with a certain trait (for example, high cooccurrence frequency) is not well-defined.

The procedure advocated by Clark uses variability within the sample material to estimate how well the obtained results will generalize to other samples of material. Mosteller and Tukey (1968) point out that there is a potential danger in any statistical procedure that uses variability within the sample materials to generalize the results to other samples. Variability within the sample may be much less than within the total population. As a consequence it has frequently been found in correlational research that significant (measured by within-sample variance) effects were not replicated when different samples were used. There are no fully adequate solutions to this dilemma. However, a move in the direction of safety is to try to use large samples of wide heterogeneity or to sample exhaustively. This helps remove the problem of generalizing from one sample to others. This experiment uses a large sample of materials.

The statistics advocated by Clark are still an improvement over existing procedures. We too, use a statistical procedure that makes use of variability among the materials in the sample. However, we wish to point out that, because of biased sampling, one should be suspicious of the conclusions of these tests and that one should attempt to deal with the problems of generalization by using large and heterogeneous samples.

### *Theoretical Attitudes*

Our principal concern in this paper is with the times taken to judge that an instance like *collie* is not a member of a category like *cat*. As noted earlier, these data on negative judgments seem to have a particularly diagnostic role in identifying the correct theory. Our attitude in this research is somewhat atheoretical. That is, we are simply trying to determine which of the possibly relevant variables predict negative time. Our preexperimental notions about what is involved in the task, however, caused us to collect some dependent measures that we might not have otherwise. These theoretical biases should be stated: We assume that if a subject sees *A collie is a cat*, he searches the set of propositions he knows about collies and cats. If he finds a subset of propositions which contradicts the original statement he will respond false. There are many possible sets of propositions that one could retrieve that logically imply the falsity of the statement, for example:

1. Collies are not cats.\*
2. Lassie is a collie.  
Lassie is not a cat.\*
3. A collie is a dog.  
I like all dogs.  
I do not like cats.\*
4. Collies are big.  
Cats are small.  
Big is the opposite of small.\*
5. A collie is a dog.  
A dog is not a cat.\*

Two factors would seem to control which of the many conceivable means a subject uses for falsifying a particular statement. The first is whether the requisite set of propositions is stored in memory. For instance, the first possibility *Collies are not cats* would be a quick route for falsification, but it is unlikely that many of us have that proposition stored in memory. The second factor is simply the relative availability of the various sets of propositions in memory.

We suspect that subjects are likely to falsify statements involving specific categories like *collie* by retrieving some more general superset of the category like *dog* and then falsifying *A dog is a cat*. That is, we suspect methods like (3) and (5) are particularly likely. Therefore, among our measures are the times to generate supersets to low level instances like *collie* and the time to falsify that the superset is another category. We expected both these times would predict the time to falsify that the low level instance is a member of the other category.

### RESEARCH STRATEGY

In a number of experiments to be described below, we obtained 23 measures to predict various reaction time decisions in our semantic memory tasks. The materials or stimuli remained constant across the various experiments; so, their construction should be described first.

Forty category names were selected. This exhausted most of the categories we could think of that passed the criteria that (a) they be one word (that is, not categories like *musical instruments*); (b) they be distinct from all other categories used; (c) they have generally recognized instances; and (d) they have relatively many instances of both high and low dominance. From each category, two instances were selected such that, intuitively, one instance seemed a salient member of the category and the second a nonsalient member of the category. This yielded 80 instance-category pairs. For each of the forty categories a second category was chosen such that it was disjoint from the first. These would serve as the negative categories for the false statements. The negative categories were constructed by repermuted the positive categories so that the range of similarity of the instances to the negative categories varied as much as possible. Thus the basic material were 80 triples consisting of an instance, a positive category, and a negative category. These are given in the Appendix.

Table 1 lists the 23 variables collected in the experiments to be described. The variables are organized into groups to be explained later. More detailed explanations of the meaning of each variable and how the statistics were obtained is given in the experiments below.

as the predicate in a categorical statement.

The procedure used for gathering reaction times was as follows: Approximately one second after the experimenter said "ready" the slide was backprojected onto a screen, starting a timer. The subject's hand rested on a neutral telegraph key between positive and

TABLE 1  
SUMMARY STATISTICS FOR REGRESSION VARIABLES

Variable	Mean	Standard deviation	Error	Proportion attenuation
1. <i>INT</i> —instance negation time	1026	92	42	.89
2. <i>INF</i> —instance and negative category conjoint frequency	1.40	1.19	.49	.91
3. <i>INS</i> —instance and negative category similarity	1.71	.96	.29	.95
4. <i>CNT</i> —category negation time	1021	99	30	.95
5. <i>CNF</i> —category and negative category conjoint frequency	2.21	1.41	.21	.99
6. <i>CNS</i> —category and negative category similarity	2.23	1.02	.25	.97
7. <i>IVT</i> —instance verification time	972	72	42	.81
8. <i>CGT</i> —corrected generation time	398	252	81	.95
9. <i>ICF</i> —instance and category conjoint frequency	4.73	.98	.46	.88
10. <i>ICS</i> —instance and category similarity	5.26	.74	.39	.85
11. <i>CGD</i> —category generation dominance	22.2	6.6	2.4	.93
12. <i>IFR</i> —instance frequency	32	82	—	—
13. <i>IIF</i> —inverse instance frequency	.27	.33	—	—
14. <i>IRT</i> —instance reading time	739	166	73	.90
15. <i>IMT</i> —instance meaning time	518	95	56	.81
16. <i>CFR</i> —category frequency	108	172	—	—
17. <i>CIF</i> —inverse category frequency	.05	.09	—	—
18. <i>CRT</i> —category reading time	641	134	73	.84
19. <i>CMT</i> —category meaning time	496	65	56	.51
20. <i>NFR</i> —negative category frequency	108	172	—	—
21. <i>NIF</i> —inverse negative category frequency	.05	.09	—	—
22. <i>NRT</i> —negative category reading time	641	134	73	.84
23. <i>NMT</i> —negative category meaning time	496	65	56	.51

### Experiment I

This experiment obtained the times for subjects to judge whether one word was a subset of another. Two words were typed on a slide, one above the other. The subject was to treat the top word as *A* and the bottom word as *B* in the phrase *An A is a B*, or in other words, the top word as the subject and the bottom word

negative keys. The subject moved his finger to hit one of the keys and indicate his decision. This stopped the timer. The experimenter recorded the reaction time and informed the subject if his decision was incorrect.

The subject was given 10 practice trials and then proceeded to make judgments of 320 pairs. There were 80 pairs of each of four types. There were positive instances which were

created by pairing each of the 80 instances in the Appendix with its positive category. There were negative instance statements created by pairing each instance with its negative category. There were negative category statements which were constructed by pairing the positive category with its negative category. Since there were only 40 distinct pairs of positive categories and negative categories, each negative category statement was repeated twice in the 320 trials. Finally, there were positive category statements which were taken by pairing the positive category with some superset. There were two copies of each of these 40 positive category pairs.

The mean verification time for each negative instance across subjects was the first variable, *INT* for *instance negation time*, in our regression analysis. This variable is the principal measure to be predicted. Mean time to negate categories (for example, *A dog is a snake*) was variable 4 in the regression analysis, *CNT* for *category negation time*. Mean time to verify positive instances (for example *A collie is a dog*) was variable 7, *IVT* for *instance verification time*. The positive category times were not used in the regression analysis because a number of categories shared the same superset, (for example, *activity*) and hence the superset had to be used an uncontrolled number of times. These trials were only employed to provide a balance of materials. The 320 pairs were placed in random order in four Carousel trays of 80 slides. Order of these slide trays was counterbalanced over subjects.

Twenty-four subjects were recruited from the general population of students and local residents available in the Yale University community during the summer. Each of the 24 possible permutations of the four slide trays was used with a different subject. They were paid \$2.00 for participation in an experiment that lasted about an hour.

### *Experiment II*

This experiment was performed to obtain an estimate of how long it took a subject to

generate a superset to each of the 80 instances. An instance was backprojected onto a screen before the subject, starting a timer. He was to respond vocally with a superset. His response triggered a voice key which stopped a clock. His response and reaction time were recorded. Subjects went through three lists of these 80 instances. The order of lists was counterbalanced across subjects. The order of instances within a set was random.

The voice key apparatus seemed to introduce a lot of systematic error into the reaction times. For instance, the category *state* did not tend to trigger the voice key until the emission of the stop *t*. In general, the apparatus seemed insensitive to fricatives like *s*. Also there probably were differences in the speed with which a subject could initiate the articulation of different words. This motivated us to derive a measure of generation time that would be independent of articulation. Subjects showed a considerable speed-up over the three passes through the 80 instances. Mean reaction time for the first set was 1407 msec; for the second set, 1095 msec; and for the third set, 1009 msec. We speculated that by the third pass the subject's response was usually an automatic recall of a response he had learned in the first two passes. If so, a useful measure would be first generation time minus third generation time. The third generation time would subtract out the stimulus encoding and response generation times in the first generation time. This would just leave the time to mentally generate the category. This difference measure gave us variable 8 in the regression analysis, *CGT* for *corrected generation time*.

Table 2 presents the matrix of correlations among reaction times for first, second, third, and corrected generation times. For each instance we used its mean reaction time averaged over subjects. The correlations are computed across these 80 instances. Note that all correlations are positive. In particular the correlation between third generation time and corrected generation time is positive. This

indicates that whatever corrected generation time measures, some of it still remains in the third generation time. Presumably, subjects occasionally still generate in this third generation condition rather than just retrieve a learned response.

All the subjects did not generate the same superordinate to all instances. We counted the frequency of the most frequent superordinate to each instance on the first pass through the material. There were 30 subjects in this experiment and the response frequency of an instance's most salient category ranged from 6 to 30. This was the measure of the existence of a dominant categorization response to each instance. This variable provided variable 11 in the regression analysis, *CGD* for *category generation dominance*.

judge for each item whether it was a word or nonword. In the second part of the experiment the subject was shown the 120 words and was asked to indicate by a button press when he had thought of the word's meaning. The first task has been studied rather extensively (for example, Meyer & Schvaneveldt, 1971; Meyer & Ellis, 1970; Rubenstein, Garfield, & Millikan, 1970). There appear to be effects of word frequency, polysemy, and associative priming. The only study we know of that used the second task was Moore (1915). *A priori*, we had no way to decide which task would give a better estimate of time to encode a word.

The experiment was run in five groups on the IBM 1800 at the Human Performance Center at the University of Michigan. Subjects

TABLE 2  
CORRELATIONS AMONG GENERATION TIME MEASURES

	First generation time	Second generation time	Third generation time
Second generation time	.75		
Third generation time	.40	.61	
Corrected generation time	.94	.57	.31

Thirty subjects were recruited from the general population of students and local residents at Yale in the summer. Five subjects were tested with each of the six possible orderings of three lists of 80 instances. They were paid \$1.50 for participation in an experiment that lasted about half an hour.

### *Experiment III*

Time to encode the words may be an important factor contributing to differences among semantic memory stimuli. Two measures were obtained in this experiment of the time to encode the words that comprised the triples in the Appendix. In the first part of the experiment subjects saw the 120 words (80 instances plus 40 categories) mixed in with 120 nonwords. The subject's task was to

sat in individual booths and responded to stimuli presented to them on the *CRT* screen. In the word judgment task, a word or nonword appeared on the screen and the subject pressed a positive or negative button to indicate his judgment. Feedback consisted of either the phrase *You are right* or *You are wrong* printed on the screen. One second after this feedback disappeared from the screen the next word was presented. Reaction time was measured from the word's appearance to the subject's button press. A short rest period intervened between the first and second tasks. The procedure was similar for the meaning task. Of course, no feedback was given. The next stimulus appeared 2 sec after the disappearance of the old stimulus from the screen.

This experiment yielded six more measures

for the regression analysis. Variable 14 was the time to make a word judgment about the instance, *IRT* for *instance reading time*. Variable 15 was *IMT*, *instance meaning time*. Variable 18 was time to make a word judgment about the *positive category for a triple*, *CRT*. Variable 19 was the *corresponding category meaning time*, *CMT*. Variable 22 was the *reading time for the negative category*, *NRT*. Variable 23 was the *corresponding meaning time*, *NMT*.

Thirteen subjects were recruited for this experiment from the University of Michigan student population. The experiment lasted about 30 min and they were paid \$1.50 for their services.

#### *Rating measures*

A number of analyses of semantic memory identify frequency as a relevant variable in semantic memory (for example, Anderson & Bower, 1973; Conrad, 1972; Glass, Holyoak, & O'Dell, 1974; Loftus & Suppes, 1972; Wilkins, 1971). In particular, the frequency with which two words co-occur together is thought to be critical in determining the number of connections between the words and the availability of these connections. Therefore we had subjects rate how frequently they thought of each of the three types of pairs derivable from the triples. Judgments of frequency of instance and category co-occurrence constituted variable 9, *ICF*, for *instance and category conjoint frequency*. Variable 2 is the rating of *cooccurrence frequency of the instance and the negative category* with which it appeared in Experiment I, *INF*. Variable 5 is the rating of *cooccurrence frequency of the positive and negative category*. Different groups of subjects made the ratings of different types of pairs. Subjects were asked to rate the pairs on a 0–7 scale where 0 indicated they never thought of the words together and 7 that they always thought of the words together. The word pair *vanilla* and *embrace* was given as an illustration of a pair which was almost never thought of together.

*Peanut butter* and *jelly* were suggested as a pair which was frequently thought of together.

The Smith *et al.* analysis identifies similarity in meaning as the controlling variable in semantic memory tasks. Three more variables were obtained by getting measures of the similarity of the three types of pairs. Variable 10 was *similarity of instance and positive category*, *ICS*. Variable 3 was *similarity of instance and negative category*, *INS*. Variable 6 was *similarity of positive and negative category*, *CNS*. Different groups of subjects gave the different ratings. For *INS* and *CNS*, subjects were simply told to rate the pairs as to similarity of meaning on a 0–7 scale. On that scale 0 meant totally different and 7 meant practically identical. As an example of a very dissimilar pair they were given *vanilla* and *embrace*. As an example of a very similar pair they were given *jelly* and *jam*. We thought it strange to ask subjects to rate the similarity of an instance to its category. So for *ICS* ratings we asked subjects to rate how typical the instance was of the category on a 0–7 scale with 0 indicating very atypical and 7 indicating most typical. Such typicality ratings are considered by Smith *et al.* to be the best measure of what they mean by similarity.

Ten Yale summer subjects gave the *INF* ratings. The other ratings came from Michigan subjects. There was relatively high agreement among raters, so large numbers of subjects were considered unnecessary. The numbers of subjects giving the various ratings were: *ICF*, 8; *NCF*, 9; *ICS*, 14; *INS*, 12; and *CNS*, 13.

#### RESULTS

In addition to the variables described above we used the *word frequency of the instance* (Variable 12, *IFR*), of the *positive category* (Variable 16, *CFR*), and of the *negative category* (Variable 20, *NFR*). Word frequency has been found to correlate negatively with word judgment time (Rubenstein *et al.*, 1970). Thus we are using word



frequency as a measure of encoding time. The measures were taken from the Kučera and Francis (1967) word count. Loftus and Suppes (1972) suggest that the inverse frequency measure is better because it eliminates the extreme range of scores. Since we had an occasional zero frequency word, we took the inverse of one plus the frequency. This gave the following inverse frequency measures: Variable 13,  $IIF = 1/(1 + IFR)$ ; Variable 17,  $CIF = 1/(1 + CFR)$ ; and Variable 21,  $NIF = 1/(1 + NFR)$ .

Summary statistics about the 23 variables are displayed in Table 1. For each variable (except *CGD* and *word frequency measures*) and for each of the 80 triples a mean value was computed over subjects. These means, the *CGD* count, and the frequency counts for each triple served as the data for the regression analysis. The mean of these values, across triples, is given under Mean in Table 1 for each variable. The standard deviation of these values is also given. The standard deviation reflects both systematic differences among the triples and random error in measurement. For those variables whose means were obtained by averaging subject data, it is possible to obtain an estimate of the error in measurement from the subject  $\times$  item interaction. These standard errors are given under Error in Table 1. An error estimate for variable 11, *CGD*, was obtained by the formula  $(p(1-p)n)^{\frac{1}{2}}$  for a binomial variable. It was not possible to calculate an estimate of error for the word frequencies.

In the final column the proportion attenuation due to error is given of a correlation between that variable and another variable. Let us explain the calculation of this quantity: Let  $s_x$  denote the standard deviation of the  $X$  scores and  $s_e$  their standard error of measurement. The proportion attenuation is given by the following formula (see McNemar, 1962):

$$r_{xx} = 1 - \frac{s_e^2}{s_x^2}$$

The square root of  $r_{xx}$  is the quantity we are referring to as proportion attenuation. Suppose  $r_t$  is the true correlation between  $X$  and  $Y$ , then the observed correlation,  $r_{xy}$ , will be somewhat smaller because of errors of measure in  $X$  and  $Y$ . McNemar gives a formula for  $r_{xy}$  in terms of  $r_t$ ,  $r_{xx}$ , and  $r_{yy}$ :

$$r_{xy} = r_t \cdot (r_{xx})^{\frac{1}{2}} \cdot (r_{yy})^{\frac{1}{2}}$$

Thus, we want the proportion of attenuation to be near 1 to have a maximally sensitive regression analysis. The attenuations are all tolerable except for variable 19, *CMT*, and variable 23, *NMT*.

The variables have been organized in Table 1 into seemingly natural groups. Variables 1–3 measure the connection between instance and negative category; variables 4–6 the connection between positive and negative category; variables 7–11 the connection between instance and positive category; variables 12–15 the encodability of the instance; variables 16–19 the encodability of the categories; variables 20–23 the encodability of the negative categories.

#### *Analysis of Instance Verification Time*

A regression analysis was performed using variable 7, *IVT*, as the independent variable and the other 22 variables as dependent variables. The results of a stepwise forward multiple regression analysis are reported in Table 3a.

We report six multiple regression analyses in Table 3. These are central to the conclusions of the paper. Therefore, it is important to understand the nature of the stepwise forward method. This method first selects that variable, among the 22, which is most highly correlated with the dependent variable. The first variable selected will often not predict all explainable variance in the dependent variable. If not, the stepwise forward method will select next the variable that best predicts the unexplained variance. That is, it selects second the variable which has the highest correlation with the dependent

variable, partialling out the first predictor variable. It keeps adding variables to the regression equation as long as they significantly increase the amount of variance predicted. The regression equations at the bottom of the panels in Table 3 give the best fitting linear equations using all the variables selected by the stepwise forward method.

There are serious problems in such multiple regression analysis. We are trying to find which set of variables from the 22 available does the best job of predicting the dependent variable. Obviously, we could use all 22 and account for the maximum variance, but many variables would probably make insignificant contributions to the variance produced.

TABLE 3  
SUMMARY OF STEPWISE REGRESSION ANALYSES

Variable to be predicted	Variables entering regression	Multiple R	Significance
(a) 7. <i>IVT</i>	11. <i>CGD</i>	.386	.0004
	12. <i>IFR</i>	.461	.0152
	$IVT = 1078 - 4.46 CGD - .22 IFR$		
(b) 1. <i>INT</i>	4. <i>CNT</i>	.545	.0001
	14. <i>IRT</i>	.630	.0006
	8. <i>CGT</i>	.669	.0100
	$INT = 526 + .59 CNT + .38 IRT + .29 CGT$		
(c) 4. <i>CNT</i>	5. <i>CNF</i>	.659	.0001
	21. <i>NIF</i>	.713	.0011
	1. <i>INT</i>	.748	.0041
	18. <i>CRT</i>	.770	.0144
	$CNT = 573 + 39.5 CNF + 279 NIF + .25 INT + .14 CRT$		
(d) 14 + 15. <i>IET</i>	13. <i>IIF</i>	.483	.0001
	12. <i>IFR</i>	.530	.0269
	23. <i>NMT</i>	.564	.0447
	$IET = 856 + 287 IIF - .28 IFR + .23 NMT$		
(e) 8. <i>CGT</i>	11. <i>CGD</i>	.585	.0001
	10. <i>ICS</i>	.638	.0048
	2. <i>INF</i>	.669	.0210
	15. <i>IMT</i>	.693	.0317
	16. <i>CFR</i>	.716	.0280
	$CGT = 1009 - 18 CGD - 105 ICS + 64.6 INF + .55 IMT - .28 CFR$		
(f) 1. <i>INT</i> without <i>CNT</i> , <i>IRT</i> , or <i>CGT</i>	5. <i>CNF</i>	.494	.0001
	10. <i>ICS</i>	.574	.0025
	$INT = 1140 + 35.8 CNF - 33.5 ICS$		

The multiple-*R* reported in Table 3 is a measure of the goodness of regression equation using that variable and the preceding variables. The square of this quantity is the proportion of variance explained. The significance of the additional variance explained by adding each variable is also reported in Table 3.

There are, in fact,  $2^{22}$  possible sets of variables to consider. The stepwise forward method is not guaranteed to find the optimal set, but often will. There are particular problems with any approach when two predictor variables are highly correlated with each other and approximately equally correlated with the dependent variable. Then, it is hard to reliably

assess the contribution of each to the regression. We have tried to be sensitive to this danger. However, so that the reader can assess this matter for himself we have published the entire correlation matrix in the Appendix. These are the correlations among the 23 variables computed across the 80 triples also listed in the Appendix.

As can be seen from Table 3a, two variables are significantly related to *IVT*. The most significant is *CGD*, the measure of the dominance of some categorization response to the instance. The other was the measure of instance frequency. Both variables suggest that a prominent strategy for instance verification is for the subject to enter memory from the instance and generate categories attempting to obtain a match to the probed category. According to this strategy, subjects should be faster with instances which have a dominant category (so long as the dominant category is the predicate of the test pair; this was the case for 67 of the 80 items). This strategy also predicts that the speed at which the subject can encode the instance should be more critical to latencies than the speed at which he can encode the category. The variables *IFR*, *IIF*, *IRT*, and *IMT* are all intended as measures of encoding time and all correlate appropriately with *IVT*. However, *IFR* has the strongest correlation with *IVT* when *CGD* is partialled out. Therefore, it was the one that entered into the regression analysis.

#### *Analysis of Instance Negation Times*

The variable of prime concern to this paper is *INT*. Table 3b presents the regression of this variable, taking all other variables as dependent. The three significant predictor variables were *CNT*, *IRT*, and *CGT*. These are all consistent with the strategy we proposed: A subject falsifies an instance statement by generating a category and determining if the retrieved category is disjoint from the probed category. Subjects should take longer to negate an instance the longer it takes them to generate a category (*CGT*) and the longer

it takes them to negate the category (*CNT*). Since the process starts from the instance, time to encode the instance should be more critical than time to encode the predicate, that is, the negative category. This is reflected by the significant regression on *IRT*, the time to judge an instance is a word.

The regression on these three variables is highly significant,  $F(3, 76) = 20.5$ ; the improvement in regression due to the other variables is not,  $F(19, 57) = .97$ . The deviation from prediction can be evaluated using the estimate of error in *INT* (see Table 1). This residual variance turns out to be quite significant,  $F(76, 1722) = 2.73$ ,  $p < .001$ . This strongly suggests that there are other strategies for negation which have not been measured in any of these variables. Note that the coefficients of *CNT* and *CGT* in the regression equation in Table 3b are not unity as would be expected if this were the only strategy employed. Also note from Table 2 that *INT* (1026 msec) is only slightly larger than *CNT* (1021 msec). If subjects always generated the category and then falsified that the category was the predicate, *INT* would be considerably longer than *CNT*.

Almost all strategies for categorical negation can logically be used for instance negation. That is, the subject can retrieve the category from the instance and just apply the category negation strategy. However, there are strategies for instance negation (such as 1, 2, and 4 in the introduction of this paper) which do not involve category negation. One can think of instance negation as being composed of a mixture of some strategies which proceed through the category and some which do not. This would lead to the following formula:

$$INT = \alpha(CGT + CNT) + (1 - \alpha)XT,$$

where  $\alpha$  represents the probability of a strategy through the category and  $XT$  is the mean time for noncategory strategies. If  $XT < CGT + CNT$ , this would yield the observed near equality of *INT* and *CNT*.

Meyer (1970) also observed near equality of what we call *INT* and *CNT*. From this, he concluded that subjects do not falsify categorical statements by going through a superordinate. This conclusion would be valid under the assumption that there is only one falsification strategy, but it is not valid under the more realistic assumption that subjects use multiple falsification strategies. By looking at the regression of *INT* on *CNT* and *CGT*, we have obtained evidence for falsification by category retrieval and this evidence does not depend on the assumption of a unique strategy.

The high correlation between *INT* and *CNT* ( $r = .55$ ) might, at first, seem suspect. That is, these measures came from the same subjects in the same experiments. Moreover, the pairs yielding these measures have the same predicates. Therefore, part of the correlation may reflect communalities in the encoding of the predicates, rather than communalities in the falsification process. Three facts argue against this interpretation: First, measures of negative category encoding time (*NFR*, *NIF*, *NRT*, and *NMT*) were not strongly related to *INT*. This suggests encoding of the negative category was not a critical variable. Second, consider the relatively low correlation ( $r = .18$ ;  $t(78) = 1.64$ ;  $p > .10$ ) between *IVT* and *CNT*. These two variables also came from the same subjects in the same experiment. They involve encoding of a common category, but should have no overlap in the verification process. Apparently, these sources of communality are not enough to introduce a significant correlation between the two variables.

The third fact is probably the most decisive. For each instance negation time, we found the average reaction time of the two instance verifications whose predicate or category is the same as the false predicate. That is, to predict the time to falsify *tent-metal* we took the mean of the times to verify *copper-metal* and *zinc-metal*. This new variable only correlated .25 with *INT*. Partialing out *CNT*,

this correlation is reduced to  $-.004$ . In contrast, the correlation between *INT* and *CNT* with this new variable partialled out is still .50. Thus, the correlation between *INT* and *CNT* is not just because they shared a common predicate.

In addition to *CNT*, two other reaction time measures were significant predictors of *INT* in the regression analysis, namely *IRT* (*instance reading time*) and *CGT* (*corrected generation time*). Further regression analyses were performed to determine what factors predicted each of the above three variables. The factors which correlate most strongly with these three variables are the nonreaction time measures, particularly the rating variables. The analyses of *CNT*, *IRT*, and *CGT* will be discussed below, followed by a regression of *INT* in which the reaction time variables are removed in order to discover which rating variables predict *INT*.

#### *Regression of Category Negation Time*

Table 3c reports the stepwise regression for *CNT*. The best predictor was *CNF*, subject ratings of how often they thought of the two categories together. *CNF* turns out to be highly correlated ( $r = .72$ ) with *CNS*, subject ratings of the similarity of the two categories: *CNS* is almost as well correlated ( $r = .61$ ) with *CNT* as is *CNF* ( $r = .66$ ). Clearly, we can make no strong discrimination between the two on the basis of this data. The Smith *et al.* theory expects negation time to be positively related to similarity since similarity between subject and predicate biases a positive response. It may not be so clear why one would expect a positive relation between *CNF* and negation time. Cooccurrence ratings could index the existence of irrelevant facts connecting the two items and these irrelevant facts would interfere with the retrieval of the requisite negative information. This is one of the hypotheses offered by Collins and Quillian (1972) to account for negative judgments. There were a few pairs in our experiment

that separated frequency and similarity. Examples of high conjoint frequency, low-similarity pairs are *weapon-money* and *music-college*. Examples of high-similarity, low-frequency pairs were *dog-snake* and *tool-jewel*. Unfortunately our sample did not have enough discriminative pairs; nevertheless, these measures do not seem inherently confounded.

Table 3c indicates that both measures of encoding of the subject (*CRT*) and of the predicate (*NIF*) are significantly related to *CNT*. It appears that experimental subjects show less asymmetry in their processing of negative category statements than they do for either positive or negative instance statements. When subjects deal with these instance statements, only measures of subject (instance) encoding are significantly related to reaction time. For the pairs that gave us our *CNT* measures, both subjects and predicate come from the same pool of 40 categories and therefore are at an equal level in a conceptual hierarchy. The fourth significant variable in Table 3c is *INT*. This is not unexpected given the strong relation between *INT* and *CNT* in Table 3b. The regression of *CNT* with *INT* removed still selects the other three variables and no others.

#### *Regression of Encoding Time*

The best predictor of *IRT*, *word judgment time*, is *IMT*, *word meaning time* ( $r = .55$ ). Presumably these are measures of approximately the same thing and so it does not seem too meaningful to let *IMT* enter the regression of *IRT*, thereby soaking up all the explainable variance. Therefore, a regression of the sum *IRT* + *IMT* was performed and is reported in Table 3d. This sum is referred to as *instance encoding time*, *IET*. Both measures of word frequency, *IFR* and *IIF*, are correlated with encoding time in a way that indicates high-frequency words are more rapidly encoded. This supports our use of frequency as a correlate of encoding ease. The entry of the third variable *NMT* into the regression in

Table 3d is almost certainly a spurious, chance correlation. It only produces a marginally significant improvement in prediction.

#### *Regression of Corrected Generation Time*

Table 3e reports the regression of *CGT*. The two most significant variables are the measure of dominance of a category, *CGD*, and the measure of instance prototypicality, *ICS*. The Holyoak and Glass and the Anderson and Bower theories would expect the importance of a production frequency measure like *CGD*. The Smith *et al.* theory predicts the importance of *ICS*. These two variables are basically orthogonal for our materials. Their intercorrelation is .18. Therefore, it might seem that both similarity and frequency theories are confirmed. This is consonant with our notion that the subject variously employs a number of strategies in the verification task.

Three other variables were significant predictors in the regression of *CGT*: *a measure of instance encoding time (IMT)* and *category frequency (CFR)*. Loftus and Suppes also found response frequency a significant variable. Their task was just the opposite of ours: Their subjects generated instances to categories. Presumably frequency is a measure of response availability. The final variable in this regression is *INF*, *subject ratings of instance and negative category cooccurrence frequency*. This may reflect the existence of extraneous paths from the instance which interfere with retrieval of the category.

#### *Rating Variables as Predictors of INT*

Table 3f reports the regression of *INT* when the reaction time measures that proved to be the best predictors are removed from the regression. The two significant measures are *CNF*, *subject ratings of positive and negative category cooccurrence frequency*, and *ICS*, *subject ratings of instance and category similarity*. The first had proved a significant predictor of *CNT* and the second a significant predictor of *CGT*. These two reaction time

measures were significant predictors of *INT* in Table 3b. Note that *INT* comes from pairs that do not mention the positive category, but both rating variables do. Neither measure of the direct connection between instance and negative category, *INF* and *INS*, proved significant. This is strong evidence for our proposed falsification strategy that subjects go through the instance's category to reject the predicate presented with the instance.

#### DISCUSSION

The regression analyses presented above indicated that a predominant strategy for falsifying *As are Cs* is the following: (a) Retrieve a superset *B* of *A*; (b) attempt to verify *Bs are Cs*. A prediction derived from this strategy was confirmed, namely, that the time to generate *B* to *A* and the time to falsify *Bs are Cs* are both correlated with the time to falsify *As are Cs*.

In the introduction, five different sets of propositions were listed as possible strategies for falsifying one particular statement. It should be noted that each set contains one asterisked proposition which is a negative proposition. A negative proposition is logically necessary to reject a statement. Therefore, an important factor in determining what strategy a subject will use is where negative information is stored. It is untenable to suppose that a subject stores with every two disjoint categories that the one is not the other. As stated earlier, it is unlikely that a subject has directly stored that *Collies are not cats*; it does seem more reasonable, however, that he has stored *Dogs are not cats*.

We do not mean to imply that negative information is never stored directly with the instance, nor that it is always directly stored at the immediate superordinate. To negate the statement *Collies are tractors*, one might generate the sequence *Collies are dogs*, *Dogs are animals*, *Tractors are machines*, and *Animals are not machines*. In this case,

retrieving *dog*, the direct superset of *collie*, is not enough. A superset of *dog* must also be retrieved. Still the time to retrieve *dog* to *collie* would be part of the negation process. One might object to the proposed strategy of recursively retrieving supersets until one is found which denies the predicate. This would seem to predict that less similar items are falsified slower than more similar items, which is just the opposite of the usual result. For instance, in the above example, more retrievals were required to reject *A collie is a tractor* than *A collie is a cat*. Two points need to be made to this objection. First, we do not mean to imply that to falsify *A collie is a tractor* a subject must retrieve *animal*. He may simply retrieve *dog* and falsify *A dog is a tractor*. The second, related point is that subjects may examine many facts simultaneously, looking to see if any can be used for confirmation or disconfirmation. This might explain why it is easier to disconfirm *A dog is a tractor* than *A dog is a cat*. The more dissimilar the instance to the predicate, the faster a disconfirming fact can be found. Dogs and cats share many properties (e.g. legs, eyes, fur), while tractors and dogs have many differences (for example, tractors rust, need no food.)

Our *a priori* reason for suspecting supersets are retrieved from the instance to falsify an instance statement is that it seems more likely that negative information would be stored with large categories. That negative information is likely to be found at large categories is probably just a reflection of which negative propositions one is likely to experience and hence learn.

The notion that negative propositions are stored in memory is considered by some to be unappealing and unintuitive. Why this reaction should occur is not clear to us since negative premises must exist in order to logically deduce a negative conclusion. This does not mean that people necessarily have stored in their heads strings of words which include *no* or *not*; rather

the information structure which we call a negative proposition simply must have negative logical force.

There are clear similarities between the original Collins and Quillian proposal that subjects search for intersections in conceptual hierarchies and the proposal that subjects retrieve supersets to verify statements. However, there are a number of critical differences. First, we do not suppose that information is only stored at the logically highest level. Second, we do not suppose that subjects will reject a particular pair such as *collie* and *cat* just because they intersect at a common superset like *animal*. Rather, information must be retrieved to indicate that one is not a subset of the other. Third, this is not proposed as the only falsification strategy; we just think that it is a likely one. Examples 1, 2, and 4, given in the introduction to this paper, are cases of logical falsifications that do not depend on retrieval of the instances' supersets.

#### SUMMARY AND CONCLUSIONS

The primary goal of this paper was to help understand the processes involved in rejecting false categorical statements. Regression analyses were presented using numerous variables to predict the time to reject a categorical statement. Reaction times, rating variables,

and frequency norms were used in the prediction of instance negation time. The variables found to be critical are the time to generate the superordinate, the time to reject a categorical statement which has the same predicate but uses the superordinate rather than the instance, and the time to encode the instance. The prominence of the particular negation strategy found in our experiments is probably due to the logical nature of the task (that is, that subjects should find logically disconfirming evidence) and to the fact that negative propositions tend to be stored with large categories.

Another goal of this paper was to call attention to problems in data analyses in most semantic memory tasks. It was pointed out that variables used in semantic memory studies tend to be correlated and hence confounded. Because one cannot randomly assign materials to conditions to test whether certain "variables" affect reaction times, one cannot be certain that the hypothesized variable has caused the effect rather than some other known or unknown variable related to it.

The present analyses do not tell us anything very critical about the nature of memory (for example, to help us discriminate between network and set-theoretical models). They just identify one prominent falsification strategy.

## APPENDIX

*Experimental Materials*

Positive categories	Instances	Negative categories
dwelling	tent, house	metal
metal	copper, zinc	spice
vehicle	car, motorcycle	weapon
toy	yo-yo, doll	country
beverage	milk, eggnog	fish
weapon	spear, gun	money
flower	camelia, rose	insect
vegetable	turnip, carrot	fruit
sport	polo, football	crime
ship	cruiser, yacht	tool
science	chemistry, geology	toy
fish	trout, herring	disease
fruit	grape, apple	pastry
dance	polka, waltz	vehicle
furniture	chair, sofa	cloth
crime	burglary, arson	science
tool	hammer, chisel	jewel
jewel	diamond, sapphire	tree
city	Reno, London	artwork
money	shilling, dollar	music
pastry	cake, donut	clothing
artwork	etching, painting	dwelling
insect	termite, bee	bird
fuel	petroleum, propane	ship
tree	oak, cedar	flower
college	Harvard, Purdue	profession
country	Cuba, France	relative
bedding	quilt, sheets	animal
cloth	cotton, tweed	furniture
spice	thyme, pepper	beverage
music	jazz, concerto	college
state	California, Idaho	city
bird	robin, pelican	vegetable
disease	cancer, measles	fuel
dog	collie, spaniel	snake
profession	lawyer, fireman	sport
animal	goat, cat	state
relative	aunt, niece	dance
clothing	shirt, vest,	bedding
snake	python, cobra	dog



CORRELATION MATRIX FOR THE 23 VARIABLES

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	
2	.29																						
3	.35	.33																					
4	.55	.34	.49																				
5	.49	.49	.54	.66																			
6	.40	.32	.86	.61	.72																		
7	.07	-.04	-.08	.19	.15	.03																	
8	.21	.13	-.16	-.06	-.11	-.19	.31																
9	-.24	.38	.02	-.06	.01	-.06	-.14	-.21															
10	-.20	.33	.18	.03	.18	.17	-.27	-.35	.65														
11	-.04	-.06	.15	-.08	.09	.17	-.39	-.59	.19	.18													
12	-.11	.30	-.02	-.02	.00	-.04	-.22	.03	.19	.22	-.07												
13	.15	-.31	.01	-.08	-.13	-.02	.06	.12	-.33	-.34	.01	-.30											
14	.34	-.14	.02	.04	.18	.12	.17	.05	-.33	-.25	.06	-.32	.50										
15	.03	-.21	-.11	-.24	-.14	-.12	.22	.20	-.22	-.15	-.03	-.30	.30	.55									
16	.04	.28	.09	.18	.40	.25	-.04	-.25	-.08	.17	.20	.01	-.16	.14	.03								
17	-.15	-.16	-.17	-.23	-.33	-.26	.05	.18	.01	-.03	-.38	-.05	.26	-.04	.04	-.29							
18	.08	.14	-.08	.16	-.06	-.09	.08	.25	.06	-.13	-.22	.20	-.03	-.06	-.12	-.23	.22						
19	-.08	-.16	-.26	-.11	-.28	-.24	.05	.19	.02	-.12	-.14	.01	-.10	-.11	.03	-.20	-.07	.31					
20	-.19	-.02	-.29	-.14	-.11	-.33	-.18	-.12	-.03	-.09	.21	.01	-.09	-.10	-.09	.29	-.04	-.05	-.10				
21	.08	-.08	.13	.26	-.02	.16	-.01	-.12	.04	-.01	.11	-.10	-.10	-.13	-.14	-.17	-.05	-.04	.15	-.29			
22	.10	.01	.02	.22	.02	.08	.07	-.04	.07	.08	.01	.04	.00	.04	.00	.03	-.05	.08	-.07	-.23	.22		
23	.04	.04	-.24	-.15	-.19	-.17	-.02	.04	-.08	-.02	-.01	.13	.00	.17	.09	.05	-.22	-.13	-.07	-.20	-.07	.31	

## REFERENCES

- ANDERSON, J. R., & BOWER, G. H. *Human associative memory*. Washington: Winston, 1973.
- 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.
- COLLINS, A. M., & QUILLIAN, M. R. Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 240-247.
- COLLINS, A. M., & QUILLIAN, M. R. How to make a language user. In E. Tulving & W. Donaldson (Eds.), *Organization and memory*. New York: Academic Press, 1972.
- CONRAD, C. Cognitive economy in semantic memory. *Journal of Experimental Psychology*, 1972, **92**, 149-154.
- GLASS, A. L., HOLYOAK, K. J., & O'DELL, C. Production frequency and the verification of quantified statements. *Journal of Verbal Learning and Verbal Behavior*, 1974, **13**, 237-254.
- HOLYOAK, K. J. & GLASS, A. L. Alternative conceptions of semantic memory. Unpublished manuscript, Stanford University, 1974.
- KUČERA, H., & FRANCIS, W. N. *Computational analysis of present-day English*. Providence, R. I.: Brown University Press, 1967.
- LANDAUER, T. K., & MEYER, D. E. Category size and semantic-memory retrieval. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 539-549.
- LOFTUS, E. F., & SCHEFF, R. W. Categorization norms for fifty representative instances. *Journal of Experimental Psychology Monograph*, 1971, **91**, 355-364.
- LOFTUS, E. F., & 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.
- MCNEMAR, Q. *Psychological statistics*. New York: Wiley, 1962.
- MEYER, D. On the representation and retrieval of stored semantic information. *Cognitive Psychology*, 1970, **1**, 242-300.
- MEYER, D. E., & ELLIS, G. B. Parallel processes in word recognition. Paper presented at the meeting of the Psychonomic Society, San Antonio November, 1970.
- MEYER, D. E., & SCHVANEVELDT, R. W. Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 1971, **90**, 227-234.
- MOORE, T. V. The temporal relations of meaning and imagery. *Psychological Review*. 1915, **22**, 177-225.
- MOSTELLER, F., & TUKEY, J. W. Data analysis, including statistics. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology*. Redding, Mass.: Addison-Wesley, 1968.
- RIPS, L. J., SHOBE, E. J., & SMITH, E. E. Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 1-20.
- RUBENSTEIN, J., GARFIELD, L., & MILLIKAN, J. A. Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 487-494.
- SCHAEFFER, B., & WALLACE, R. The comparison of word meanings. *Journal of Experimental Psychology*, 1970, **86**, 144-152.
- SMITH, E. E., SHOBE, E., & RIPS, L. Comparison processes in semantic memory. In G. H. Bower (Ed.), *Psychology of learning and motivation*, Vol. 8, in press.
- WILKINS, A. T. Conjoint frequency, category size, and categorization time. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 382-385.

(Received June 7, 1974)

