Journal of Experimental Psychology: Learning, Memory, and Cognition

(*formerly* JEP: Human Learning and Memory)

VOL. 8, NO. 1 JANUARY 1982

Semantic Distance Effects in Categorization Tasks

Roger W. Schvaneveldt
New Mexico State University

Francis T. Durso
University of Oklahoma

Basabi R. Mukherji
State University of New York College at Old Westbury

Six different categorization tasks were studied in an investigation of the effects of semantic relatedness on categorization performance. A different pair of categories was used in each task. The category pairs were *vowel–consonant* (referring to the initial letters of words), *word–nonword, good–bad, plant–animal, natural–artificial* (or natural vs. manufactured), and *noun–verb*. For the first three category pairs, relatedness facilitated both "same" and "different" decisions when subjects were asked to judge whether both of two items belonged to the same category. For the last three category pairs, relatedness facilitated "same" decisions but did not reliably affect "different" decisions. The findings are generally inconsistent with decision bias theories and with simple versions of spreading activation theory. The intersection model, an elaboration of spreading activation theory, is proposed to account for the effects of semantic relatedness in this and other studies.

In the last decade, cognitive psychologists have begun to analyze the processes and structures underlying people's knowledge. By measuring the time required to decide whether a statement is true (e.g., Collins & Quillian, 1969; Meyer, 1970), whether two words are exemplars of the same category (e.g., Schaeffer & Wallace, 1969, 1970), or whether a string of letters constitutes an English word (e.g., Meyer & Schvaneveldt, 1971), researchers have begun to make claims about the nature of semantic and lexical memory.

Portions of this research were presented at the 1979 meetings of the Psychonomic Society (Schvaneveldt, Durso, & Mukherji, Note 1). We are indebted to Frank Jagusiak, Mary Ann Hayes, Donna Silverglade, Gary Miller, and Warren Dodge for their assistance with various aspects of the research. Also, Don Dearholt, Steven Keele, and John Kiger provided several helpful comments on an earlier draft of the manuscript.

Requests for reprints should be sent to Roger W. Schvaneveldt, Department of Psychology, New Mexico State University, Las Cruces, New Mexico 88003.

Semantic Distance Effects

Much of the research in semantic memory has focused on the influence of semantic relatedness on sentence verification, category judgments, and lexical decisions. Various terms, such as *semantic similarity, semantic relatedness,* and *semantic distance*, have been used to refer to the degree to which concepts are related in meaning. The distance metaphor comes from an analogy to a multidimensional space where concepts are located according to their values on various dimensions of meaning. Presumably, concepts near one another in multidimensional space are more closely related to one another than are concepts that are farther apart in the space.

Many researchers have reported that decisions based on pairs of semantically similar concepts can be made more rapidly than decisions based on pairs of semantically dissimilar concepts (e.g., Collins & Quillian, 1969; Meyer & Schvaneveldt, 1971; Schvaneveldt & Meyer, 1973). Facilitation has not always been observed, however, when the semantic relatedness between items is increased.

Schaeffer and Wallace (1970) first reported that when the semantic similarity of the members of a pair of words was increased, the time required to deny that both members of the pair belong to the same category increased. The categories they used were *trees, flowers, birds*, and *mammals*. Subjects were slower in deciding that related words such as MAPLE and DAISY belonged to different categories than they were in making the same decision about relatively unrelated words such as MAPLE and CHICKEN. In general, when the items from different categories were both plants or both animals, the decision that they were from different categories was slower than when one item was a plant and the other an animal.

Given the design of the Schaeffer and Wallace experiment, however, subjects could have responded to some of the stimuli simply on the basis of similarity information, without explicitly considering information specific to the category decision. In particular, whenever an unrelated pair of items was presented, the "different" decision was correct. Thus if subjects were able to rapidly assess the overall semantic similarity of the words in a pair, a "different" decision could have been made simply on the basis of the low degree of similarity. In contrast, semantically related words were sometimes from the same category and sometimes from different categories. Subjects could not respond correctly to pairs of related words simply on the knowledge that the words were related. The point is that responses to pairs of unrelated words from different categories may have been based on information about the lack of a relationship rather than on information about category membership. Because a lack of semantic similarity was inherent in the "different" pairs, it is perhaps not surprising that the similarity information affected the decision criteria. Similarly, Shoben (1976) conducted a study in which unrelated pairs always required a "different" response; he also found that relatedness between the items of the pair slowed decisions.

Further, in sentence verification tasks, semantic relatedness between the subject and predicate increases the time required to deny the veracity of the statement (e.g., Smith, Shoben, & Rips, 1974). For example, it takes longer to deny that "A bat is a bird" than it does to deny that "A rock is a bird." However, Glass, Holyoak, and colleagues (Glass & Holyoak, 1975; Glass, Holyoak, & Kiger, 1979; Glass, Holyoak, & O'Dell, 1974; Holyoak & Glass, 1975) found that semantic relatedness can facilitate rather than inhibit negative decisions in sentence verification tasks. Glass et al. (1974) used meaningful sentences to create both unrelated and related "false" sentences. They found that people classified the related false sentences faster than the unrelated ones. Their related false sentences contained direct contradictions, whereas the unrelated ones did not. This is consistent with the claim that people are able to assess contradictions rapidly and respond accordingly (Friedman & Bourne, 1976).

The present study was undertaken to provide additional information about the role of semantic relatedness in categorization tasks. We used tasks at various levels of abstraction to determine the limits of relatedness effects. By including both related and unrelated pairs of items for both positive and negative decisions, we attempted to reduce the utility of a decision bias based on the overall relatedness of the items in a pair. With this design, we should be able to determine how relatedness affects the speed of retrieving categorical information from memory. Several models lead to definite predictions regarding our experiment. Rather than review these predictions here, we first present our findings and then discuss the models in relation to one another and in relation to our results.

Method

Design

Six different pairs of categories were used in a series of classification tasks. The pairs of categories were *vowel–consonant, word–nonword, good–bad, plant–an-*

imal, natural–artificial, and *noun–verb.* In a given classification task, subjects were required to use one of the pairs of categories to make judgments about the category correspondence of two simultaneously presented stimuli. If both members of the pair came from the same category, the appropriate response was to move the response lever in the direction labeled "same." If one of the stimuli came from one category and the other stimulus from the other category, the appropriate response was to move the response lever in the direction labeled "different."

The members of the pairs of stimuli differed in semantic relatedness (related vs. unrelated) and in whether they belonged to the same category ("same" vs. "different"). The orthogonal combination of six tasks, two levels of relatedness, and two response categories yielded 24 conditions. The order of the tasks was counterbalanced with a Latin square. The order of stimuli, relatedness, and response categories was independently randomized within each task for each subject. Within each task there were four lists of stimuli that counterbalanced the occurrence of particular stimuli over the four conditions resulting from the combination of relatedness and response category. Thus each stimulus was presented an equal number of times in each of these four conditions. However, each subject saw each stimulus only once.

Materials

The stimulus materials were different for each of the six category-decision tasks. For each task, the stimuli were constructed from four sets of items. Each set consisted of two quadruples of items. Within a quadruple, two of the items belonged to one of the categories and two belonged to the other category. For example, in the natural–artificial task, two of the items were HAND and FOOT (both natural) and two other items were GLOVE and SHOE (both artificial). The members of these pairs were highly related and required a "same" response. Further, each of these items was also highly related to one of the items in the other category (e.g., HAND–GLOVE and FOOT–SHOE). Pairing these items together yielded the semantically related pairs from different categories. The semantically unrelated pairs were constructed by re-pairing an item from one quadruple with an item from another quadruple to form either unrelated-same pairs (e.g., HAND–RIVER, both natural) or unrelated-different pairs (e.g., HAND–CANOE, one natural and one artificial). In this way each set of two quadruples yielded 16 pairs, 4 in each of the four combinations of relatedness and response category. By re-pairing items in this manner, each item occurred in each of the four conditions. When this re-pairing was done for a set, a total of 64 pairs were generated for each task. Examples of the item quadruples are shown in Table 1 and of the pairs generated from two quadruples in Table 2.

Table 3 shows the average similarity ratings for the various pairings of the items in each task. The similarity ratings were obtained in a norming procedure in which 14 students judged the semantic relatedness of the pairs on a 7-point scale.

From the total pool of pairs, four lists were formed with 16 pairs in each list for each task. Each stimulus item was present in each list, but the item was paired with different words so that all four conditions were represented by the same words across the four lists.

For each of the six tasks, half of the stimuli belonged to one of the two designated categories, and half belonged to the other. For the pairs requiring a "same" response, half were composed of instances of one category and half of instances of the other category. For the pairs requiring a "different" response, the first item of a pair was equally often from one category as the other.

In the word–nonword task, the nonwords were misspellings of the words. The misspellings preserved the phonemic pattern (e.g., DOCTER for DOCTOR). In this way, we were able to construct pairs of nonwords that were derived from related words and whose phonemic patterns led to associated semantic codes (e.g., DOCTER–NERSE). Lapinski (1979) showed that nonwords con-

Table 1
Examples of Two Item Quadruples for Each Task

Task	Quadruple 1	Quadruple 2
Vowel	ENTIRE–ALL	OBOE–UKULELE
Consonant	WHOLE–TOTAL	CLARINET–GUITAR
Word	WORK–JOB	DOG–CAT
Nonword	JAWB–WERK	KAT–DAWG
Good	ANGEL–HEAVEN	DELICIOUS–TASTY
Bad	DEVIL–HELL	SOUR–BITTER
Plant	FLOWER–LEAF	PEANUT–BANANA
Animal	BEE–CATERPILLAR	ELEPHANT–MONKEY
Natural	HAND–FOOT	RIVER–LAKE
Artificial	GLOVE–SHOE	CANOE–BOAT
Noun	LAUNDRY–DETERGENT	BRAVERY–HERO
Verb	RINSE–CLEANSE	PROTECT–DEFEND

Table 2
Example of Item Pairings for the Good–Bad Task

Category	Semantically related	Semantically unrelated
Same	ANGEL(G)–HEAVEN(G)	ANGEL (G)–DELICIOUS(G)
	HELL(B)–DEVIL(B)	HELL(B)–BITTER(B)
	TASTY(G)–DELICIOUS(G)	TASTY(G)–HEAVEN(G)
	SOUR(B)–BITTER(B)	SOUR(B)–DEVIL(B)
Different	ANGEL(G)–DEVIL(B)	ANGEL(G)–BITTER(B)
	HELL(B)–HEAVEN(G)	HELL(B)–DELICIOUS(G)
	TASTY(G)–BITTER(B)	TASTY(G)–DEVIL(B)
	SOUR(B)–DELICIOUS(B)	SOUR(B)–HEAVEN(G)

Note. G = good; B = bad.

structed in this manner produce priming of "semantically related" words. A word–nonword pair was never composed of a word and its nonword derivative. Instead, the related word–nonword pairs consisted of a word and the nonword derived from a related word (e.g., DOCTOR–NERSE).

Each stimulus pair was typed in all uppercase letters in the center of a beige card. One item was typed directly above the other.

Procedure

The subjects were tested individually in sessions of approximately 40 min. duration. Each subject was seated in front of the tachistoscope and instructed in the use of the start button and the response lever. Assignment of switch direction (up or down) to response category ("same" or "different") was counterbalanced. Subjects were instructed to respond quickly and accurately, but it was emphasized that accuracy was more important. This particular form of speed–accuracy instruction was deemed necessary because of the high error rate found with pilot subjects.

The subjects were then given instructions for an initial block of practice trials. The category membership decision required for the practice block was based on the categories large–small. Subjects were instructed to respond "same" if both items in a pair referred to large objects or if both referred to small objects. If one item referred to a large object and the other to a small object, they were to respond "different." The practice block was intended to accustom the subject to the task in general. During this block, subjects were corrected if they made an error, and the trial was rerun.

Following the general practice on large–small decisions, subjects were presented with the six experimental tasks in a counterbalanced order. Each task began with instructions appropriate for the new category designations. A block of 8 practice trials, similar to the experimental trials, preceded the block of 16 experimental trials. Accuracy was again stressed during each block of specific practice. For the experimental trials, no feedback was given and error trials were not rerun.

A typical trial began with a signal from the experimenter informing the subject that the trial could be initiated. The subject pressed the start button, and 500 msec later the stimulus pair appeared in the viewing field for 3 sec, after which a blank field replaced the stimulus. Latency and response were recorded by the experimenter, and the sequence began again.

For each task, subjects were to respond "same" if both items were from the same predesignated category and "different" otherwise. A detailed explanation of what constituted category membership was presented in addition to the verbal labels of the categories. For the good–bad, noun–verb, and word–nonword tasks, the explanations were simply lengthier descriptions of the category labels. For the vowel–consonant task, a pair was defined as belonging to the same category whenever the first letters of the two words were either both vowels or both consonants. Otherwise, they belonged to different categories. For the natural–artificial task, membership was defined by whether "the object occurs naturally in the environment or is something made by man." For the plant–animal task, the categories were expanded to include "plants or parts of plants" or "animals or parts of animals."

Each subject responded to 16 pairs within each of the six tasks for a total of 96 trials excluding practice. Half of the pairs required a "same" response and half a "different" response. In half of the pairs, the items were

Table 3
Semantic Similarity Ratings for the Four Conditions in Each Task

	Related		Unrelated	
Task	Same	Different	Same	Different
Vowel–consonant	4.90	4.98	1.37	1.20
Word–nonword	*	*	*	*
Good–bad	5.55	4.16	2.41	1.82
Plant–animal	3.80	4.12	2.27	1.42
Natural–artificial	4.49	4.72	1.34	1.20
Noun–verb	4.91	4.67	1.64	1.58

Note. The ratings are the averages for 14 judges using a 7-point rating scale. Ratings were not obtained for items in the word–nonword task.

highly related, and in the other half they were unrelated. For each task, the initial ordering of the stimulus cards was made with a random number table, and the cards were shuffled between subjects so that each subject received a different order of the 16 pairs for each task.

Apparatus

The stimuli were presented with a three-field Scientific Prototype tachistoscope adapted to measure latencies. A lever movement stopped a msec clock and turned on a light indicating the direction of the lever movement. Subjects started each trial by pressing a button on the tachistoscope control panel.

Subjects

Twenty-four student volunteers from introductory psychology courses at the State University of New York at Stony Brook participated in the experiment. They received extra course credit for participating.

Results

Only the latencies for correct responses were included in the computation of average response times. Mean latency and mean number of errors were computed for each subject's data in each condition. These data were submitted to separate 6 × 2 × 2 (Task × Response × Relatedness) analyses of variance.

The average error rates for each condition in each task are shown in Table 4. In the error analysis, there were reliable effects of task, $F(5,\ 115) = 8.49$, $MS_e = .511$, $p < .001$, and relatedness, $F(1,\ 23) = 4.73$, $MS_e = .449$, $p < .05$, and an interaction between response and relatedness, $F(1,\ 23) = 5.24$, $MS_e = .612$, $p < .05$. The task effect reflects the fact that the tasks varied in overall error rates. There were fewer errors overall for related pairs than for unrelated pairs (11.0% vs. 14.2%). However, the difference was large for "same" responses (6.9%) and essentially zero for "different" responses (−0.7%). Overall, there was a correlation of .596 between mean error rates and mean reaction time across all conditions and all tasks.

The analysis of latencies revealed reliable effects of task, $F(5,\ 115) = 69.10$, $MS_e = 124{,}786$, $p < .001$; response, $F(1,\ 23) = 24.58$, $MS_e = 63{,}515$, $p < .001$; and relatedness, $F(1,\ 23) = 31.56$, $MS_e = 51{,}113$, $p < .001$. There was also a significant interaction between response and relatedness, $F(1,\ 23) = 7.56$, $MS_e = 69{,}035$, $p < .025$, and a three-way interaction of Task × Response × Relatedness, $F(5,\ 115) = 3.65$, $MS_e = 57{,}817$, $p < .005$. The three-way interaction is shown in Figure 1, where we have plotted the interaction of response and relatedness separately for each task.

Inspection of Figure 1 suggests that the three-way interaction may result from the tendency of the vowel–consonant, word–nonword, and good–bad tasks to show an advantage for related pairs over unrelated pairs for both "same" and "different" responses, whereas the plant–animal, natural–artificial, and noun–verb tasks show facilitation from relatedness for "same" responses and interference for "different" responses. The amount of facilitation for positive responses also appears to vary with the task. However, no task shows a detrimental effect of relatedness for positive responses.

Individual Task Analysis

Further analysis of the three-way interaction was accomplished by performing sep-

Table 4
Error Rates (Percentages) for the Four Conditions in Each Task

Task	Same: Related	Same: Unrelated	Different: Related	Different: Unrelated	Task average
Vowel–consonant	8.2	10.5	7.3	5.3	7.8
Word–nonword	20.5	27.0	15.8	11.5	18.8
Good–bad	2.0	7.3	4.3	8.3	5.5
Plant–animal	9.3	19.8	11.5	12.5	13.3
Natural–artificial	9.5	15.8	10.5	12.5	12.1
Noun–verb	11.5	22.0	22.0	16.8	18.1
Condition average	10.2	17.1	11.9	11.2	12.6

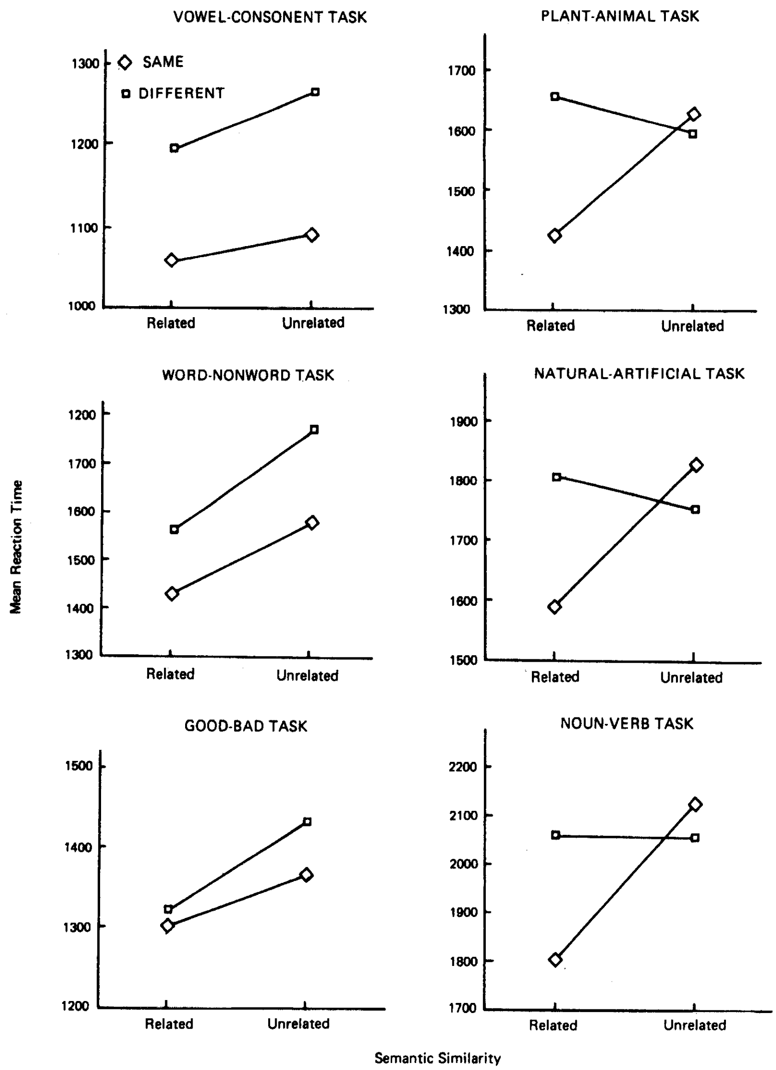

Figure 1. Mean reaction time as a function of relatedness and response, plotted separately for each of the six tasks.

arate 2 × 2 (Relatedness × Response) analyses on each of the six tasks. These analyses generally confirmed the conjectures offered above. Details of these analyses are presented in the following paragraphs. We analyzed error rates as well as reaction times so we could detect any speed–accuracy trade-offs in the various tasks.

Vowel–consonant. There were no significant effects in the error analysis for the vowel–consonant task. The reaction time analysis showed significant effects of response, $F(1, 23) = 21.37$, $MS_e = 27{,}257$, $p < .001$, and relatedness, $F(1, 23) = 6.91$, $MS_e = 9{,}736$, $p < .025$. The interaction was not reliable ($F < 1$). "Same" responses were approximately 150 msec faster than "different" responses. Related pairs were classified about 50 msec faster than unrelated pairs.

Word–nonword. There were significantly more errors made on stimuli requiring "same" responses (23.9%) as compared to "different" responses (13.2%), $F(1, 23) = 9.27$, $MS_e = .449$, $p < .01$. Neither the relatedness effect nor the interaction of response and relatedness was reliable in the analysis of error rates. The reaction time analysis revealed significant effects of response, $F(1, 23) = 16.87$, $MS_e = 38{,}212$, $p < 001$, and relatedness, $F(1, 23) = 12.46$, $MS_e = 61{,}324$, $p < .01$. The interaction was not significant. "Same" responses were faster by about 150 msec. However, the speed advantage for "same" responses should be qualified by a reliable error rate difference favoring "different" responses. Subjects may have been trading speed for accuracy on "same" responses. Responses to related pairs were about 175 msec faster than responses to unrelated pairs.

Good–bad. Only the relatedness variable produced a significant effect in the analysis of error rates, $F(1, 23) = 4.40$, $MS_e = .192$, $p < .05$. Related pairs of words led to 3.2% errors as compared to 7.8% for unrelated pairs. In the reaction time analysis, the only effect to approach significance was relatedness, $F(1, 23) = 3.75$, $MS_e = 47{,}703$, $p = .065$. Related pairs of words were classified about 75 msec faster than unrelated pairs.

Plant–animal. There were no reliable effects in the overall analysis of errors. The reaction time analysis revealed a significant interaction between response and relatedness, $F(1, 23) = 7.48$, $MS_e = 54{,}804$, $p < .025$, and a marginal response effect, $F(1, 23) = 4.01$, $MS_e = 57{,}817$, $p = .057$. Overall, "same" responses averaged about 100 msec faster than "different" responses. Further analysis of the interaction revealed that for "same" responses, related pairs were classified 198 ± 63 msec[1] faster than unrelated pairs ($p < .01$). There were also significantly fewer errors on related pairs, $F(1, 23) = 4.39$, $MS_e = .475$, $p < .05$. "Different" responses showed a nonsignificant difference in the opposite direction for both reaction time (-63 ± 78 msec) and error rate.

Natural–artificial. No effects were reliable in the analysis of errors. The reaction time analysis showed a significant effect of relatedness, $F(1, 23) = 4.72$, $MS_e = 43{,}115$, $p < .05$, and a reliable interaction of response and relatedness, $F(1, 23) = 5.95$, $MS_e = 82{,}448$, $p < .025$. Further analysis of the interaction showed that significantly faster responses were made to related pairs for "same" responses (a difference of 235 ± 74 msec), but for "different" responses, the difference was in the opposite direction and was not significant (-51 ± 70 msec).

Noun–verb. Only the interaction of response and relatedness approached significance in the analysis of errors, $F(1, 23) = 4.18$, $MS_e = .561$, $p = .053$. Relatedness produced a 10.5% effect on error rates for "same" responses and an opposite, −5.2% effect for "different" responses. In the reaction time analysis, there were significant effects of response, $F(1, 23) = 5.87$, $MS_e = 36{,}085$, $p < 0.25$, and relatedness, $F(1, 23) = 8.51$, $MS_e = 69{,}892$, $p < .01$. The interaction was also significant, $F(1, 23) = 5.16$, $MS_e = 122{,}412$, $p < .05$. Detailed analyses of the interaction revealed that related pairs were classified 319 ± 98 msec faster than unrelated pairs for "same" responses ($p < .01$). There were also fewer errors on related pairs, but not significantly so. For "different" responses, there was essentially no difference between related and unrelated pairs.

Regression Analysis

Another way of looking at the data is shown in Figure 2. This figure shows the relatedness effect as a function of the average time to respond to unrelated pairs of items for both "same" and "different" responses in each task. As the diamonds reveal, the magnitude of the relatedness effect for positive responses is a linear increasing function of the absolute amount of time required to respond to unrelated pairs. In other words, the longer it took to decide that an unrelated pair of items belonged to the same category, the greater the facilitation for classifying related pairs belonging to the same category. The slope of the regression line, fit to the data from positive responses, suggests that relatedness reduced latencies by 290 msec for each second of response time. The linear regression accounted for 99% of the variance in the mean latencies.

For the negative responses, the pattern of relatedness effects was quite different. As

[1] In statistics in the form $X \pm Y$, Y represents one standard error of the statistic X. The probability values are based on the t distribution.

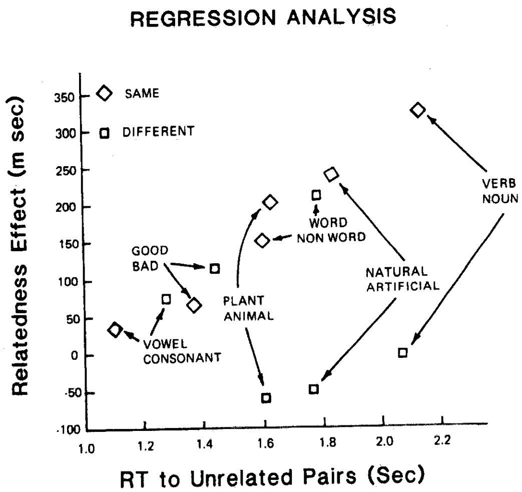

Figure 2. The regression of the relatedness effect on the mean response time (RT) to unrelated pairs.

the squares in Figure 2 indicate, there was no clear relationship between the relatedness effect and the time required to respond to unrelated pairs for the items requiring "different" responses. Instead, the tasks fall into the same two groups we found for the earlier analyses. In particular, the tasks that showed additive effects of response and relatedness appear to follow the regression function for both positive and negative responses. For the three tasks exhibiting interactive effects in the earlier analyses, the regression line is a poor predictor of the magnitude of the relatedness effect with negative decisions.

In summary, the categories we studied produced two different patterns of response times. Relatedness facilitated positive responses for the plant–animal, natural–artificial, and noun–verb categories, but relatedness had no significant effects on negative responses for these categories. There was a tendency for relatedness to inhibit negative responses in these categories, but the effects were small relative to the variability in the data. In contrast, relatedness facilitated both positive and negative decisions for the vowel–consonant, word–nonword, and good–bad categories. The effects of relatedness with the latter categories were well predicted from the amount of time needed to process an unrelated pair of items, whereas only the positive responses from the former set of categories were well described by the regression of relatedness effects on response time to unrelated pairs.

Discussion

Our experiment, in agreement with several other studies, shows that semantic similarity between members of pairs of stimuli affects the time required to decide whether the stimuli belong to the same category. Semantic similarity effects range over a wide variety of categories at various levels of abstraction. Judgments about particular letters in words are affected by the semantic relationships between the words. Also, judgments about semantic classes are influenced by the semantic similarity of the concepts being judged. The semantic effects in the vowel–consonant task are surprising. Decisions about the first letters of words do not require people to deal with semantic information at all. Yet, semantic similarity has a consistent facilitating influence. It may be that people automatically recognize the words before analyzing them into their constituent letters. There is ample evidence that related words can be recognized more rapidly than unrelated words (Meyer & Schvaneveldt, 1971; Schvaneveldt & McDonald, 1981).

The word–nonword task has been shown to be sensitive to semantic factors in several previous experiments (e.g., Meyer & Schvaneveldt, 1971, 1976). The present experiment is consistent with other work (Antos, 1979; Lapinski, 1979; Schvaneveldt & McDonald, 1981; Schvaneveldt & Meyer, 1973; Lapinski & Tweedy, Note 2) in suggesting that the effects of semantic context in the lexical-decision task cannot be explained solely by appeal to decision bias or response bias, since both "same" and "different" responses were facilitated by semantic similarity. Instead, the results suggest that relatedness facilitates accessing relevant information in memory.

For the tasks requiring semantic analysis, we find two distinct patterns of results. Either both positive and negative responses are facilitated by semantic relatedness, or positive responses show facilitation and negative responses show no effect. There was a tendency in two tasks (plant–animal and natural–artificial) for the negative responses to be inhibited by semantic relatedness, but the inhibition was not statistically reliable. The inhibitory effects we observed were considerably smaller than those found by Schaeffer and Wallace (1970). Shoben (1976) also reported large inhibitory effects of semantic relatedness on "different" responses in a similar task. One major difference in the studies was the strong manipulation of relatedness in both "same" and "different" responses in our study. When relatedness is only varied for the "different" responses, as in the Schaeffer and Wallace study, subjects may classify the unrelated "different" stimuli on the basis of their lack of relatedness. When such a strategy is precluded, as it was in our study, the evidence for an inhibitory effect of relatedness on "different" decisions is much weaker.

Despite the lack of evidence for inhibitory effects of relatedness, the evidence for two different patterns of relatedness effects is compelling. Some of the tasks yield reliable interactions of relatedness and response, whereas in other tasks the effects are additive. The basis of the division of the tasks into the two different patterns is not obvious. The distinction does not seem to follow a semantic versus nonsemantic classification since the good–bad task that is semantic yields the same pattern of results as the vowel–consonant and the word–nonword tasks. Also, the noun–verb task that presumably requires syntactic information falls into the same category as the plant–animal and natural–artificial tasks that involve semantic judgments. Even if we argued that the noun–verb task was performed using semantic information, we still have the problem with the good–bad task. If we define task difficulty empirically by the overall time required to perform the task (see Figure 2), we find a tendency for the more difficult tasks to show interactions of relatedness and response, whereas the easier tasks show facilitation from relatedness for both positive and negative responses. This classification scheme is not completely satisfactory, however, since the word–nonword task is more difficult than the plant–animal task by the total-time criterion, and yet the word–nonword task shows the effects found in the easier tasks. Thus the additive versus interactive patterns do not seem to be associated with the depth of processing required by the task for either an a priori or an empirical definition of depth.

Inhibitory Effects of Relatedness

Other recent studies reported inhibitory effects of semantic relatedness in situations requiring several different responses. A. S. Brown (1979) studied performance in a task designed after the tip-of-the-tongue effect reported by R. Brown and McNeill (1966). A. Brown presented subjects with a priming stimulus followed by a definition. The subjects' task was to retrieve and pronounce the target word corresponding to the definition. The priming stimuli were related to the target word in various ways. When the priming stimulus was a word semantically related to the target, the response latencies were longer than when the priming stimulus was neutral (i.e., three asterisks). Thus semantic relatedness inhibited the retrieval of the target word given its definition. In some of the studies A. Brown reported, the semantically related primes also led to poorer performance than unrelated word primes. As we argue in the following section, these inhibitory effects of relatedness and the pattern of results obtained in the current study are not easily explained with either a spreading activation theory or a theory about decision bias. Spreading activation would lead to the expectation that a related prime should *facilitate* the retrieval of the target word (Collins & Quillian, 1969; Schvaneveldt & Meyer, 1973). Since the response is the utterance of the target word itself, it is difficult to see how a decision bias (e.g., Smith et al., 1974) explanation could be applied to these data.

Other studies by Rosch (1975a, 1975b) showed that a category prime can inhibit the decision that two strings of letters are identical. The priming effects occur when the letter strings spell a word that is a member of the category designated by the prime. When the word is a typical member of the category, a decision is facilitated. However, atypical members show inhibition from the category prime. Also, Neill (1979) found that unexpected stimuli that are related to expected stimuli can be more difficult to process than unexpected stimuli that are unrelated to the expected ones.

Semantic Memory Models and Inhibition

Adequate models of semantic memory have been constrained by the finding that relatedness inhibited decisions (e.g., Smith et al., 1974) and are constrained still further by the two patterns of results observed in the current study. Whereas relatedness in all of our tasks facilitated positive decisions, in three of our tasks, relatedness facilitated negative decisions, and in three others, relatedness did not affect negative decisions. This pattern is a problem for traditional spreading activation theories (e.g., Collins & Quillian, 1969) and decision bias models (e.g., Smith et al., 1974). The results also demand further specification from elaborated network models (e.g., Collins & Loftus, 1975). As we shall see, the findings also present problems for simple hybrid models that use both spreading activation and decision bias.

Some of the early theories (e.g., Collins & Quillian, 1969; Meyer & Schvaneveldt, 1971; Schvaneveldt & Meyer, 1973) proposed a process of "spreading activation" to account for the effect of semantic relatedness in tasks requiring retrieval of information from long-term memory. According to the spreading activation theory, concepts are stored in memory with links between semantically related concepts. When one particular concept is activated (by perceptual processes or by internal "thought" processes), the activity spreads to other related concepts via the links. Subsequent retrieval of one of these related concepts is presumably facilitated by the preactivation it has already received. This preactivation (or "priming" as it is sometimes called) provides the basis for explaining the facilitation produced by semantic relatedness. Without further qualification, the spreading activation theory suggests that semantic relatedness should generally facilitate the retrieval of information from memory, and as a result, should facilitate both positive and negative decisions. This is clearly not the case here; even when we controlled for the confound present in the Schaeffer and Wallace (1970) study, relatedness did not facilitate negative responses in all of the tasks. There is no reason to suspect interactions of the kind we observed between response and relatedness according to spreading activation theory.

In another class of models, the decision stage is the locus of relatedness effects. We use the feature comparison model of Smith et al. (1974) as an exemplar of this class, which includes the proposals of Meyer (1970) and Schaeffer and Wallace (1970) as well. In the Smith et al. model, concepts are represented by lists of attributes that vary along a dimension of "definingness." Unlike spreading activation theories, category membership cannot be retrieved in the feature comparison model. Instead, category membership must be computed from the attributes present. This computation takes place in two stages. Stage 1 evaluates the overall similarity between the two concepts in terms of number of shared attributes. If similarity is very high or very low, then a decision can be based on Stage 1. Only the decisions based on Stage 1 processing are influenced by relatedness. With high relatedness, Stage 1 processing results in a positive response. With low relatedness, Stage 1 processing results in a negative response. With intermediate levels of relatedness, Stage 2 is required. It is important to note that the contents of the shared feature lists are not evaluated in Stage 1. Evaluation of the content of the lists rests with Stage 2. In Stage 2 the defining features of the concepts are extracted and compared. In this model, whether Stage 2 is required is the primary determinant of response time. However, Stage 2 processing is not a function of relatedness.

Viewed as a strategy subjects might use in an experiment, the decision process proposed by Smith et al. may ordinarily be efficient, since concepts that belong to the same category tend to be related along several dimensions. Thus relatedness may often provide sufficient information to respond correctly. In studies that confound relatedness and response, this model accounts for the data quite well. For example, in the Schaeffer and Wallace study, subjects could respond negatively on the basis of Stage 1 because all unrelated pairs required a negative response. Similarly, with sentences like "All arrows are intelligent," the subject can rapidly respond "false" using Stage 1, since arrows and intelligence are totally dissimilar.

Despite the utility of a strategy of responding on the basis of relatedness in some situations, relatedness is not a completely reliable criterion of category correspondence. Concepts can be relatively unrelated and belong to the same category or highly related and belong to different categories. In the current study, and in studies that do not use anomalous false sentences (Glass & Holyoak, 1975; Glass et al., 1974), the model does not fare as well. Simple alterations to the model do not increase its explanatory power for these data. For example, one could reasonably argue that because relatedness was not a predictor of the correct response in our study, subjects spread their similarity criteria apart (essentially bypassing Stage 1). If all responses were based on Stage 2, however, we should have observed no relatedness effects, which was clearly not the case.

A hybrid model that uses the constructs of both spreading activation theory and the feature comparison model does somewhat better against the current data, but it does not give a completely satisfactory account. Relatedness could affect two stages. First, it could facilitate access to concepts, regardless of the ultimate response (as in spreading activation theory). Second, it could create a bias to respond positively, thus facilitating positive responses and inhibiting negative ones (as in the feature comparison model). This model would lead to the weaker prediction that relatedness could sometimes facilitate negative responses, depending on the relative contributions of spreading activation and decision bias. A problem with this model is that it predicts that relatedness effects for negative responses should always be smaller than relatedness effects for positive responses in the same task. As can be seen in Figure 1, we obtained effects in the opposite direction for the word–nonword, good–bad, and vowel–consonant tasks.

The model of semantic memory we introduce in the following paragraphs is capable of accounting for inhibitory effects by using spreading activation as a central concept. Although the model is similar to the elaboration of spreading activation theory proposed by Collins and Loftus (1975), it is more specific about certain processes. In particular, it attempts to specify a mechanism that determines the order in which information in memory is evaluated. The resulting order excludes irrelevant information from the search for some pairs but not for others.

Inhibition From Spreading Activation

Perhaps activation may be at the heart of some inhibitory effects of relatedness after all. Our discussion of spreading activation theory has emphasized the positive aspects of a spreading activation process, leading to an expectation of facilitation due to relatedness. A closer look at the potential consequences of spreading activation may suggest a different conclusion, however. We propose the *intersection model*, which specifies the operation of attentional processes in the context of spreading activation theory. With the intersection model, both facilitation and inhibition can be produced by semantic relatedness. The result that occurs depends on the relevance of the relationships between concepts for the task at hand.

Suppose that the presentation of a word leads to a process such as spreading activation, which serves to increase the activity level of concepts in memory that are related to the word. As a result, these concepts are accessed more quickly than they would be otherwise (cf. Schvaneveldt & McDonald, 1981). When two words are presented, spreading activation causes an increase in the activity level of concepts related to each of the words. This is similar to the idea of overlap of feature lists proposed by Smith et al., with the exception that several types of relations could exist in the overlap. In the intersection model, however, the content of the intersection, and not simply the extent of the overlap, is important. Concepts related to both of the words (i.e., the intersection of the sets of concepts related to each of the words) will be more highly activated than those related to only one of the words. Further, when the two words are related to one another, they will mutually increase the activation level of each other, leading to rapid high levels of activation. The intersection model assumes that attention is involuntarily drawn to concepts with unusually high activation levels. The concepts in the intersection of related words will ordinarily be highly activated and will be evaluated first. Whether relatedness will produce facilitation or inhibition will depend on the contents of the intersection and their relevance to the task at hand. Facilitation will occur when the most active concepts are relevant, and inhibition will occur when they are irrelevant. Posner and Snyder (1974, 1975) introduced the "cost–benefit" analysis to explain both the facilitation produced by appropriate primes and the inhibition produced by inappropriate primes. Posner and Snyder proposed that inhibition (or cost) was the result of attention being misdirected by an inappropriate priming stimulus. In the intersection model, we propose that a similar misdirection can occur because of highly activated but irrelevant concepts.

In the intersection model, activation patterns essentially determine the order in which concepts are evaluated in the course of making a category decision. Unlike the feature comparison model, which does not consider the contents of the initial overlap, the intersection model does allow operation on this information. In fact, evaluation of the information at the intersection is an essential difference between decision bias models and the intersection model. In decision bias models, salient but irrelevant information affects the decision indirectly through the similarity index. In the intersection model, decisions need not be based on irrelevant information. Rather, salient but irrelevant information delays the evaluation of relevant information.

The intersection model assumes that category decisions are reached by evaluating information retrieved from memory. If some concepts are activated to a considerable extent beyond other concepts, the highly activated concepts will be evaluated first. Obviously, the information retrieved will depend on what subjects know about the concepts being judged. The stored information might include direct category information (e.g., knowing that a tree is a plant), category information supporting an inference (e.g., knowing that a maple is a tree and that a tree is a plant allows the inference that a maple is a plant), referential information supporting inferences about the target categories (e.g., knowing that cleanse refers to an action allows the inference that cleanse is a verb), or properties supporting inferences about the target categories (e.g., knowing that a glove is manufactured allows the inference that it is artificial). Unlike models in which decisions are based on the overlap of features, the intersection model assumes that category decisions are based either on directly stored category information, if it is available, or on inferences based on whatever information is available. As a result, incorrect decisions are attributed to inappropriate inferences or to information that does not agree with the experimenters' definitions (e.g., all subjects may not agree that RICH is *good*).

Unrelated pairs of words can be viewed as activating two nonoverlapping sets of information. Because there is no intersection for unrelated pairs, the search for relevant category information will be guided by the salient information of each separate semantic set. A "same–different" decision can be reached by determining the category membership of each item, or perhaps by determining the category membership of one item and then using that decision to guide the processing for the other member of the pair (cf. Shoben, 1976). In either case, information is evaluated separately for each of the items.

For related pairs of items, we have different expectations. Applying the intersection model to the tasks we studied requires determining whether the information necessary to make a decision is found among the concepts in the intersection of the concepts related to each of the words in related pairs of items. In the case of the vowel–consonant task and the word–nonword task, the relevant information (letter identities and correct spelling, respectively) is presumably an important part of the concept represented by the presented words. Since activation is presumed to spread *from* these concepts, it seems reasonable to suppose that these concepts would be highly activated and therefore evaluated quickly. Further, when the two words are related to one another, they should produce further mutual activation leading to higher activity levels and more rapid evaluations. Relatedness should have this effect regardless of the correspondence of the letter categories (vowel–consonant) or the lexical categories (word–nonword). In other words, when the task requires accessing information stored directly with the concept, relatedness should consistently facilitate performance, since the lexical entries of related words are by definition related and therefore in the intersection of the activation produced by each of the words.

In contrast, the other tasks (good–bad, plant–animal, natural–artificial, and noun–verb) cannot be performed using lexical information. Presumably, additional information must be retrieved before decisions can be reached in these tasks. For pairs of items in the "same" category, information relevant to their membership in the same category is likely to be found in the intersection of activated concepts. In particular, the items probably share membership in several categories at various levels of abstraction. Even if the target categories are not directly retrievable, the decision may be reduced to a single decision that a less abstract category belongs to one of the target categories. For example, the information that HAND and FOOT are both *body parts* simplifies the decision to the single decision that *body parts* are *natural*. In some cases, the intersection may even contain the category as an individual concept. Thus, for "same" pairs, relatedness should generally facilitate the accessing of relevant information.

When the categories are based on semantic criteria, the members of highly related "different" pairs are by definition related in some way other than sharing membership in the relevant categories. These relationships are likely to delay the evaluation of relevant information. As an example, consider the pair HAND–GLOVE from the natural–artificial task. The relationships that obtain between these two concepts include spatial contiguity, the property of having fingers, and so forth, none of which is relevant to the decision that one is natural and the other artificial. As a result, the early encounter (due to high activation levels) with the commonalities between these two concepts will not facilitate the decision that HAND and GLOVE belong to different categories. Rather, these highly activated concepts may even delay the evaluation of relevant information.

Our experiment did not produce reliable inhibition from relatedness on the "different" pairs. There were some trends in that direction, but the inhibition was small relative to the facilitation found for "same" pairs. The intersection model may help to explain why the inhibition was as small as it was in our study. Relatedness presumably leads to rapid activation of concepts in the intersection. This activation may facilitate the encoding of words as it does in the lexical decision task (Schvaneveldt & McDonald, 1981), and consequently relatedness could lead to an earlier onset of the evaluation process. Thus spreading activation may have two effects: facilitation of encoding and direction of attention to the intersection of related concepts. When the intersection contains irrelevant information, the second effect would be inhibitory, and the first effect would be facilitating. Thus inhibitory effects of relatedness may be at least partially offset by facilitation in accessing related concepts.

Finally, consider the findings from the good–bad task. From the perspective of the intersection model, the additive effects of relatedness and response imply either that evaluative information is directly available from the lexical representations of words or that the intersection of the concepts related to such words as CLEAN and DIRTY or ANGEL and DEVIL contains information relevant to the good–bad categorization. Either of these alternatives is plausible. Glass et al. (1979) suggested that antonyms are processed differently in sentence verification tasks. Antonyms may lead to direct contradictions. In our good–bad task, the antonyms may lead directly to the activation of information about their "oppositeness," thus providing a basis for a "different" response.

There is also some evidence that evaluative information may be more readily accessed than other types of semantic information. For one thing, work with the semantic differential has shown that the evaluative dimension accounts for a considerable amount of the variability in many situations (Osgood, Suci, & Tannebaum, 1957). Also Zajonc (1980) argued that evaluative information may be available prior to other types of semantic information. Thus it may be that the good–bad task makes use of evaluative information that is more readily available than the information required by the other semantic tasks. Earlier we suggested that the distinction between lexical and semantic categories is one means of classifying our tasks. It may be that there is a more psychologically valid distinction between information that can be retrieved via direct connections in memory versus information that is established via inference. The latter distinction would more naturally deal with the availability of evaluative information.

The intersection model is capable of accounting for some cases of inhibitory effects resulting from semantic relatedness. The novel contribution of the model consists of the hypothesis that the most active concepts in memory involuntarily attract attentive processes just as intense stimuli in the environment attract attention (Titchener, 1908). Similar ideas have been proposed in models of episodic memory (Raaijmakers & Shiffrin, 1980). This hypothesis shows promise for increasing our understanding of semantic memory, including such phenomena as the Stroop effect (Stroop, 1935) and the tip-of-the-tongue effect (A. S. Brown, 1979; R. Brown & McNeill, 1966). For example, with the tip-of-the-tongue effect, both failures of retrieval (R. Brown & McNeill) and inhibition of retrieval (A. S. Brown) are attributed to a concentration of activation on an incorrect concept. This activation causes a misdirection of attention to the incorrect concept, perhaps causing still further increases in activation. This may help to explain the common observation that a tip-of-the-tongue experience is often accompanied by repeated recall of a particular incorrect concept.

Regardless of the fate of the intersection model, our study provides some additional findings that further constrain theories of the representation and processing of semantic information. In particular, the processing assumptions of theories require further elaboration. The nature of attentional processes operating in memory retrieval is one possible direction for this development.

Reference Notes

1. Schvaneveldt, R. W., Durso, F. T., & Mukherji, B. R. *Does semantic similarity inhibit negative decisions?* Paper presented at the meeting of the Psychonomic Society, Phoenix, November 1979.
2. Lapinski, R. H., & Tweedy, J. R. *Associate-like nonwords in a lexical-decision task: Paradoxical semantic context effects.* Paper presented at the meeting of the Mathematical Psychology Society, New York University, August 1976.

References

Antos, S. J. Processing facilitation in a lexical decision task. *Journal of Experimental Psychology: Human Perception and Performance*, 1979, *5*, 527–545.

Brown, A. S. Priming effects in semantic memory retrieval processes. *Journal of Experimental Psychology: Human Learning and Memory*, 1979, *5*, 65–77.

Brown, R., & McNeill, D. The "tip of the tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior*, 1966, *5*, 325–337.

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

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

Friedman, A., & Bourne, L. E., Jr. Encoding the levels of information in pictures and words. *Journal of Experimental Psychology: General*, 1976, *105*, 169–190.

Glass, A. L., & Holyoak, K. J. Alternative conceptions of semantic memory. *Cognition*, 1975, *3*, 313–339.

Glass, A. L., Holyoak, K. J., & Kiger, J. I. Role of antonymy relations in semantic judgments. *Journal of Experimental Psychology: Human Learning and Memory*, 1979, *5*, 598–606.

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

Holyoak, K. J., & 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.

Lapinski, R. H. *Sensitivity and bias in the lexical-decision task.* (Doctoral dissertation, State University of New York at Stony Brook, 1979). *Dissertation Abstracts International*, 1979, *40*, 925B. (University Microfilms No. DEL79-18058)

Meyer, D. E. On the representation and retrieval of stored semantic information. *Cognitive Psychology*, 1970, *1*, 242–299.

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.

Meyer, D. E., & Schvaneveldt, R. W. Meaning, memory structure, and mental processes. *Science,* 1976, *192,* 27–33.

Neill, W. T. Switching attention within and between categories: Evidence for intracategory inhibition. *Memory and Cognition,* 1979, *7,* 283-290.

Osgood, C. E., Suci, G. J., & Tannebaum, P. H. *The measurement of meaning.* Urbana: University of Illinois Press, 1957.

Posner, M. I., & Snyder, C. R. R. Facilitation and inhibition in the processing of signals. In P. Rabbitt & S. Dornic (Eds.), *Attention and performance V.* New York: Academic Press, 1974.

Posner, M. I., & Snyder, C. R. R. Attention and cognitive control. In R. Solso (Ed.), *Information processing and cognition: The Loyola Symposium.* Hillsdale, N.J.: Erlbaum, 1975.

Raaijmakers, J. G. W., & Shiffrin, R. M. SAM: A theory of probabilistic search of associative memory. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 14). 1980, New York: Academic Press.

Rosch, E. Cognitive representations of semantic categories. *Journal of Experimental Psychology: General,* 1975, *104,* 192–233. (a)

Rosch, E. The nature of mental codes for color categories. *Journal of Experimental Psychology: Human Perception and Performance,* 1975, *1,* 303–322. (b)

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

Schaeffer, B., & Wallace, R. The comparison of word meanings. *Journal of Experimental Psychology,* 1970, *86,* 144–152.

Schvaneveldt, R. W., & McDonald, J. E. Semantic context and the encoding of words: Evidence for two modes of stimulus analysis. *Journal of Experimental Psychology: Human Perception and Performance,* 1981, *7,* 673–687.

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

Shoben, E. J. The verification of semantic relations in a same–different paradigm: An asymmetry in semantic memory. *Journal of Verbal Learning and Verbal Behavior,* 1976, *15,* 365–379.

Smith, E. E., Shoben, E. J., & Rips, L. J. Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review,* 1974, *81,* 214–241.

Stroop, J. R. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology,* 1935, *18,* 643–662.

Titchener, E. B. *Lectures on the elementary psychology of feeling and attention.* New York: Macmillan, 1908.

Zajonc, R. B. Feeling and thinking: Preferences need no inferences. *American Psychologist,* 1980, *35,* 151–175.

Received January 28, 1981 ■