

EFFECTS OF SEMANTIC JUDGMENTS AND PAIR COMPARISONS
UPON LIST DIFFERENTIATION

by

Mary Mapes O'Connor

Thesis submitted to the Graduate Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

in

Psychology

APPROVED:



J. D. Fritzen, Chairman



L. C. Perlmutter



W. B. Pavlik

May, 1975

Blacksburg, Virginia

LD
5655
V855
1975
027
C.2

ACKNOWLEDGEMENTS

I wish to extend my appreciation to Dr. James D. Fritzen who served as my advisor and to Dr. William B. Pavlik and Dr. Lawrence C. Perlmutter who served on my committee. I have appreciated their time and attention to this thesis and their most helpful guidance and encouragement during its development.

A consistent finding in recent research has been the facilitatory effect of "semantic analysis" upon memory as compared to nonsemantic analysis. Hyde and Jenkins (1969) examined the effects on recall of a variety of orienting tasks. Words were presented singly, and subjects in one group were asked to rate the words on a three-point scale of pleasantness. In other conditions, subjects estimated word length or indicated if an e was present in the word. They found that when subjects were instructed to rate words on pleasantness, they subsequently recalled more words than when they were instructed to estimate word length or detect es in words. This difference was observed for both subjects who were informed of a later recall task and those who were not so informed. Since clustered recall was not found after the e detection task, the authors suggested that nonsemantic task orientation inhibits the assignment of meaning to words and consequently their subjective organization.

Johnston and Jenkins (1971) examined a similar task which forced all subjects to orient to the entire word (rather than to the presence or absence of e or to length) by instructing subjects in the nonsemantic group to suggest a word which rhymed with the word presented. This condition did not require subjects to consider the meaning of the word, according to the authors. The semantic task was to provide an appropriate adjective modifier of the noun presented or vice versa. As in the Hyde and Jenkins (1969) study, more clustering and recall were observed in the semantic condition. Again, no difference was observed in the level of recall or clustering between those subjects instructed

of a later recall task and those not told of the later task. Johnston and Jenkins suggested that this facilitation in recall and clustering was a direct result of semantic analysis and that the nonsemantic task precluded semantic analysis by subjects.

Craik and Lockhart (1972) suggested that such facilitation might be explained within a "levels of analysis" framework. This model suggests that the analysis of any stimulus "...proceeds through a series of sensory stages to levels associated with matching or pattern recognition and finally to semantic-associative stages of stimulus enrichment" (p.675). The completion of such a succession of analyses depends on three factors: the nature of the material (semantic associative potential), available processing capacity (the preexperimental familiarity with the material and the presentation rate), and the task demands (at which level the subject has been instructed to analyze). The later recall or recognition of the material, the authors contended, is a direct function of the depth of analysis of that material (e.g., the deeper the analysis, the better the recall or recognition).

Bower and Karlin (1974) used the ratings of and the later recognition of pictures of faces rather than words. Those pictures rated on honesty or likeableness were better recognized than those judged on gender. As in the Hyde and Jenkins (1969) and the Johnston and Jenkins (1971) studies, this difference was observed in both those subjects who were informed of a later recognition task and those not so informed. The authors were in general agreement with the Craik and Lockhart (1972) view that later memory for an event is dependent on the depth to which

that event is processed. Bower and Karlin, unlike Craik and Lockhart, however, suggested possible mechanisms which might more clearly explain this facilitation of memory at different levels of analysis.

One possibility is that certain judgments require more processing than others. For example, a judgment on gender might require the subjects' noting of hair length, cosmetic makeup, bushy eyebrows, and so on, of the stimulus. Judgments on honesty and likeableness, on the other hand, might require the comparison of the stimulus to an idiosyncratic set of "likeable" or "honest" characteristics such of those people the subject knows to have those traits. This differential and more extensive search through memory for "likeable" or "honest" judgments as compared to gender judgments might then explain the facilitation for later memory following the former two types of analysis.

A similar explanation might be that during the comparison of the stimuli to the subject's memory prototype (instances of or situations involving the criterion) of the experimental criterion, features are extracted from the prototype and stored with the stimulus. As such, more comparisons to the prototypes (as suggested for "likeable" and "honest" judgments) would lead to a larger amount of information stored with the stimulus. Such differential storage might then also explain better recognition of pictures which had previously been judged on "honesty" or "likeableness" than for those pictures judged on gender. In other words, the more prototypes (or features within a prototype) compared to a stimulus during a judgment, the more likely that stimulus will be "enriched" with information about those comparisons.

Gardiner (1974) found that the more semantic attributes necessary for detecting a target within an array of words, the better the later recall of that target. Subjects were required to scan a list of words and check the word which met the criterion (one or two phonemic or semantic attributes). That is, subjects scanning for a two attribute target such as "religious building" recalled target words more readily than those subjects scanning for one attribute targets such as "building" or "religion". However, those conditions in which phonemic attributes were the criteria for target choice revealed no differences in recall as a function of the number of criteria. While Gardiner concurred with the Craik and Lockhart (1972) model of "levels of analysis" as an explanation of his data, the data are also consistent with the Bower and Karlin (1974) position that the semantic analysis of stimuli enriches these stimuli as the analysis requires a more extensive search through memory and possibly more storage with the target than the analysis of phonemic qualities.

If such a complex or extensive memory search facilitates later recall or recognition, then perhaps these features are marked in some manner during the presentation (Anderson and Bower, 1972). To explore the extent to which information is stored in memory about judgments made about words, Fritzen (Note 1) focused upon the later recognition of the type of judgment made rather than upon the later recognition of the words on which the judgments were made. His basic task was to present word-pairs to subjects who were to choose the word from the pair which was more appropriate to the experimenter's criterion. Two

lists of 24 word-pairs were given, and for any one list only one criterion was used (e.g., farming or money). Following the presentation of the two lists, subjects were presented a sheet containing the word-pairs and were asked to rate each pair on a six-point confidence scale according to which list it had appeared on. List differentiation was measured by computing the difference in mean ratings for List 1 and List 2 word-pairs.

In his first experiment, he found that changing the semantic criterion from the first to the second list resulted in better differentiation of list membership relative to a condition in which the same criterion was used in both lists. He observed in his second experiment that the level of list differentiation for those subjects using different criteria for each list did not differ significantly from subjects who were tested for item recognition alone. Furthermore, those subjects who judged words from both lists on the same criterion did not differ from subjects who had no criterion but were told they would receive a memory test later. Fritzen concluded that judgments concerning the relationships of words to a particular semantic criterion resulted in the storage of information which later enhanced list differentiation.

Fritzen examined three possible theoretical accounts for such an observation of facilitation in the remainder of his article. The first account dealt with the possibility that a temporal dating hypothesis (e.g., Yntema and Trask, 1963) might explain the facilitation as a differential temporal tagging of those events (or word-pairs) prior to and following the instructions to switch the criterion. Fritzen

presented four lists of words to his subjects. One group judged all lists on the same criterion, while another group judged Lists 1 and 4 on the same criterion and Lists 2 and 3 on different criteria. The third group judged Lists 2 and 3 on the same criterion and Lists 1 and 4 on different criteria. No support was found for the temporal dating hypothesis as differentiation was highest for those lists for which different criteria were used and as no effect upon differentiation was observed for those lists having the same criterion. The temporal dating hypothesis would have predicted differentiation for those latter lists, particularly those at the extreme list positions.

A second proposed explanation for the facilitation was that the criterion, as an implicit verbal response, became associated with each word. As such, subjects whose criterion (or implicit verbal response) changed on the second list would be better able to later differentiate items than subjects whose criterion remained the same for both lists. Fritzen found no support for this in that subjects whose criterion changed on the second list subsequently differentiated the second list better than those subjects who either had the same criterion on both lists or had the same criterion but whose label for the criterion had changed (e.g., money and later, finance). The performance of the latter two groups did not differ.

A third explanation proposed by Fritzen was that subjects retain information about which word was chosen from the pair and by inference during the LD task, rule out the possibility of that pair having appeared on a particular list (especially in the case when the word-pair appeared

on a particular list and the nonchosen word in the pair better met the other list's criterion). Two groups judged both lists on the same criterion, although one group chose words from one list as more related and from the other list as less related to that criterion. A third group judged the two lists on different criteria. Half of the word-pairs were tested as pairs, while the other half were tested singly. According to the retention of outcome hypothesis, the reversed group which judged both lists on the same criterion should show better differentiation as compared to the standard one criterion group only for those items tested as pairs and not for those words tested singly. As this prediction was supported by the data, Fritzen concluded that the results of the first and second experiments may indeed reflect, at least in part, inferences made by subjects of the likelihood of a pair's list membership from their memory of which word was chosen.

His analysis of items tested singly revealed that switching criterion facilitated differentiation of both those words previously chosen and those not chosen, although facilitation of the latter was somewhat attenuated.

As a followup of this retention of outcome hypothesis, Fritzen re-examined the relationship between recognition and differentiation by presenting one group of subjects two lists of word-pairs with different criteria and later giving recognition and differentiation tests of the words presented singly. Words previously chosen were better differentiated than those words not chosen, while there was no type-of-item effect for recognition. In other words, the recognition and differentiation

rates for the words were not comparable and were differentially affected by the subjects' judgments of words according to some criterion.

The present study was designed to explore two different issues related to the facilitation of differentiation demonstrated by Fritzen (Note 1). The assumption was made that the more extensive the memory search is for a judgment, the more likely information is stored with the word and the better the subsequent list differentiation. The first issue dealt with the effect of semantic versus nonsemantic analysis of word-pairs on later list differentiation. The second issue dealt with the effect of semantic analysis of word-pairs versus single words on list differentiation.

First, Bower and Karlin (1974) suggested that the set of prototypes (instances of the criterion) to which a word is compared directly affects the amount of information stored with the stimulus or the extensiveness of the memory search during the judgment. The more the stimulus is "enriched" as a consequence of this storage or search, the better the later recognition. Also, in Hyde and Jenkins (1969) and Johnston and Jenkins (1971), it was found that recall was differentially affected by semantic and nonsemantic analysis. If then, as Bower and Karlin suggest, a semantic analysis leads to a more extensive memory search, one should expect to find less list differentiation in a task similar to Fritzen's when subjects are judging word-pairs on nonsemantic criteria than when judging the pairs on semantic criteria. The first experiment then was designed to explore this issue by attempting to replicate Fritzen's first experiment using nonsemantic criteria as orienting

tasks.

Second, the Hyde and Jenkins (1969) and the Johnston and Jenkins (1971) studies presented items singly to the subjects to rate on a criterion, while the Fritzen study had subjects choosing the word from each pair which better met the criterion. A recent study by Rips, Shoben, and Smith (1973) suggests that the pair comparison is a more accurate reflection of semantic analysis than is the comparison of a single word to a criterion. They obtained derived semantic distance scores for categories and their instances (e.g., birds and robins) using a multi-dimensional scaling technique. These distances were then used to predict reaction time (RT) in various tasks. They found that these scores served as a better predictor of RT when subjects were to indicate whether two words were both instances of the same category than when subjects were to decide whether one given word was an instance of a given category. To account for this, they suggested that the classification of pairs involved a greater use of semantic information available to the subjects than did the other task. This greater use of semantic information (comparing two instances to a category) might then enhance performance on later memory tasks if, as according to the Bower and Karlin position, a more extensive memory search facilitates later performance on a memory task.

Experiment II then was designed to assess the effects of word-pair comparisons versus single word judgments on later list differentiation. That is, this experiment was designed to compare the effects of comparing one instance to a category (as in the Hyde and Jenkins, 1969,

and the Johnston and Jenkins, 1971, studies) with the effects of comparing two instances to a category (as in Fritzen, Note 1) on later list differentiation. If the latter facilitates semantic memory search, then according to the extensive search position of Bower and Karlin, one should expect superior differentiation of items judged as members of pairs rather than those items judged in isolation.

Consideration of other related issues appears in Appendix A.

Experiment I

Experiment I was designed to assess the effects of semantic and nonsemantic judgments on list differentiation. In the experiment, subjects were presented two lists of word-pairs. For each list a criterion was announced on which subjects were to base their choice of the more appropriate word from each pair. Following this presentation of lists, subjects were asked to rate pairs according to which list they appeared on. The basic questions addressed by this experiment were whether an LD facilitation might be demonstrated by subjects making nonsemantic judgments as compared to an appropriate control and if such a facilitation existed, how it might compare to the LD scores of those subjects making semantic judgments.

Method

Subjects. Sixty-four undergraduate students at Virginia Polytechnic Institute and State University served as subjects for the experiment. All received extra credit in their Introductory Psychology class for their participation. The subjects were assigned to one of four groups in alternating order as they arrived at the experiment.

Procedure. Word-pairs were presented serially on slides at a 5-sec rate. Following the presentation of the first list, the experimenter used 45 sec to change slide trays and then presented a second list of 24 word-pairs. Following the completion of the second list, subjects were instructed to work on a symbol cancellation task for 3 min and 45 sec. Subsequently, the experimenter gave each subject one of two test sheets on which half of the word-pairs from each list were printed. The experimenter then instructed the subject to indicate on a six-point confidence scale (Winograd, 1968) which list each word-pair had appeared on. Subjects were allowed to work on this task as long as they wished. Following the completion of the first test sheet, subjects were given a second test sheet which contained the remaining word-pairs. Subjects were instructed to complete this task in the same manner as the first task.

Prior to the presentation of the first list, subjects were told that they would see various word-pairs projected on the wall and that they were to say aloud the word in the pair which met the experimenter's announced criterion. They were then to say aloud the remaining word in the pair. They were instructed that the decision must be made within 5 seconds of the pair's projection onto the wall and should they be unable to decide on the appropriate word, they should just randomly pick one word to say first. Subjects were given one practice pair, two if necessary, to insure their understanding of the instructions. The criterion used during the practice session was the same criterion used as a basis for judgments on the first list. The words used during the

practice session, however, were neither presented later in the experimental session nor tested following the session.

The subjects were then told that they would make their judgments on a particular criterion (the same as that used during the practice session). For example, subjects might be asked to choose the longer of the two words or the word more related to "Farming and Agriculture". Subjects were not told that a second list would follow.

Design. There were four groups of 16 subjects each: two, an Experimental and a Control, for which criteria for judgments were semantic, and two, an Experimental and a Control, for which criteria for judgments were nonsemantic. For example, one semantic criterion was "word more related to Money and Finances", while one nonsemantic criterion was "word with more vowels".

In both the Semantic and Nonsemantic Experimental groups, subjects judged the second list word-pairs on a criterion different from that on which they judged the first list word-pairs. Following the presentation of the first list, the experimenter announced to these subjects that the criterion would change and that they were to continue to make their judgments in the same manner although to a different criterion. In both Semantic and Nonsemantic Control groups, the criterion remained the same for both lists. Following the completion of the first list, the experimenter announced that a second list would be presented and that the subjects were to continue to make their judgments in the same manner and to the same criterion as List 1. Within both Nonsemantic groups, half of the subjects had criteria from Nonsemantic Subset 1, the other

half from Nonsemantic Subset 2.

Each list served as List 1 for half of each group and as List 2 for the other half. Likewise, in Experimental groups, each semantic ("Farming and Agriculture" and "Money and Finances") or nonsemantic ("First in alphabetical order", "Longer word", "More vowels", and "Word printed in all capital letters") criterion was used equally often on the first and second lists. In Control groups, each semantic or nonsemantic criterion was announced for both lists for half of each group. Each test sheet served as the first LD test tool for half of each group and as the second test for the other half.

Materials. The two lists used in this experiment were the same as those used by Fritzen (Note 1). In that study, each list consisted of 24 word-pairs. Half of the words in each list were taken from the lists given by Winograd (1968). The remaining words were taken from Thorndike-Lorge (1944) such that the distributions of the letters that began the words in the two lists were equivalent. The words represented a wide range of frequencies and were predominately two-syllable nouns. The words were paired randomly except that pairs were avoided which were semantically similar (e.g., student-degree), began with the same letter, or were of equal length.

The distractor task consisted of a sheet on which 10 Greek letters were repeatedly and randomly ordered. A legend at the top of the sheet displayed the letters and their corresponding numbers. Subjects were instructed to cancel out each of the letters on the sheet and mark above each cancellation the appropriate number.

Results

Rating difference scores. Each subject's differentiation between lists was measured by computing the difference between his mean ratings of List 2 and List 1 word-pairs. This is the same measure as that used by Winograd (1968). The larger the LD score, the better the differentiation.

The Semantic Experimental group achieved an LD score of 3.45, while the Semantic Control group scored an overall LD measure of .96. The Nonsemantic Experimental group achieved a 1.10 LD score, while the Nonsemantic Control group scored .59 on the measure. An analysis of variance revealed a significant effect between Experimental and Control conditions, $F(1,60) = 62.63$, $p < .001$, and between Semantic and Nonsemantic conditions, $F(1,60) = 51.85$, $p < .001$. The interaction of these two factors was also significant, $F(1,60) = 27.38$, $p < .001$. Results of a post hoc Newman-Keuls test among the four conditions indicated that the Semantic Experimental group's performance was significantly better than the other three groups' performances, $p < .01$, and that the other three groups did not differ significantly, $p > .05$. Thus, the rating scores appear to indicate that nonsemantic analyses do not lead to LD facilitation.

Correct identifications. As suggested by Winograd (1968), an alternate measure of list differentiation is the simple calculation of the number of word-pairs correctly assigned to the lists. For example, List 1 word-pairs must be rated as either "1", "2", or "3" to be considered correct, while List 2 word-pairs must be rated either "4", "5", or "6" to be considered correct.

The Semantic Experimental group averaged 41.88 correct assignments (out of a possible 48), while the Semantic Control group assigned 29.75 word-pairs accurately. The Nonsemantic Experimental group averaged 30.88 correct identifications, with the Nonsemantic Control group assigning an average of 26.94 pairs correctly. An analysis of variance revealed results similar to those for the difference in ratings: a significant effect between Experimental and Control conditions, $F(1,60) = 54.92$, $p < .001$, and between Semantic and Nonsemantic conditions, $F(1,60) = 40.61$, $p < .001$. The interaction between these two factors was also significant, $F(1,60) = 14.27$, $p < .001$. Results of the post hoc Newman-Keuls test indicated that both the Semantic and Nonsemantic Experimental groups' performances were superior to their respective controls, $p < .01$ and $p < .05$. Thus, unlike the rating difference score, the proportion correct measure indicated a significant effect of nonsemantic analyses.

Further analyses of the Nonsemantic groups revealed mean word-pairs correctly identified as 33.63, 28.50, 28.13, and 25.38 for the Vowel-Capital (V-C) Experimental condition, V-C Control condition, Alphabet-Length (A-L) Experimental condition, and A-L Control condition, respectively. An analysis of variance indicated a significant effect between Experimental and Control conditions, $F(1,28) = 10.13$, $p < .05$, and between V-C and A-L conditions, $F(1,28) = 12.16$, $p < .05$. No significant interaction was found, $F(1,28) = .92$, $p > .05$. It appears that the V-C analysis, whether in the Experimental or Control condition, leads to superior LD. One cannot conclude, however, that the two types of analyses differ in sensitivity to the basic LD task (e.g., switching

criterion across lists) as no interaction was observed.

Analysis of the four tasks for the Nonsemantic Control conditions revealed mean word-pairs correctly identified as 29.25, 27.75, 26.25, and 24.5 for the Capital, Vowel, Alphabet, and Length Control conditions, respectively. An analysis of variance revealed no significant difference between these four conditions, $F(3,12) = 1.28$, $p > .05$.

Serial position curves. To ascertain if the level of LD varied as a function of serial position within lists, an analysis of variance was performed, using the number correct for each successive block of 4 positions (six blocks for each list) within each of the four groups (see Table 1).

For the Semantic Experimental group, there was no effect for serial position, $F(5,75) = 1.57$, $p > .05$, or for List 1 vs. List 2, $F(1,15) = .73$, $p > .05$. In addition, the interaction of these variables was not significant, $F(5,75) = .61$, $p > .05$. For the Nonsemantic Experimental group, similar results were observed. The analysis of variance revealed no significant effect for serial position, $F(5,75) = .36$, $p > .05$, or for List 1 vs. List 2, $F(1,15) = .28$, $p > .05$. The interaction of these variables was also not significant, $F(5,75) = 1.13$, $p > .05$.

For the Semantic Control condition, there was no significant effect for serial position, $F(5,75) = 1.90$, $p > .05$, or for List 1 vs. List 2, $F(1,15) = .86$, $p > .05$. The interaction of these variables was significant, $F(5,75) = 3.32$, $p < .01$. A Duncan-Bonferroni test of the means revealed that the first position for List 1 yielded significantly more correct responses than for the same position in List 2, $p < .01$, and that the

Table 1
 Number of Correct List Assignments
 Across Serial Positions of Two Lists
 as a Function of Switching Criterion

Group	n	Serial Position (in word blocks of 4)					
		1	2	3	4	5	6
Experimental Conditions							
Semantic							
List 1	16	53	58	56	54	56	54
List 2	16	56	55	53	55	61	59
Nonsemantic							
List 1	16	40	40	43	40	44	36
List 2	16	45	47	37	40	39	43
Control Conditions							
Semantic							
List 1	16	47	45	31	33	35	39
List 2	16	33	41	38	40	43	50
Nonsemantic							
List 1	16	47	34	35	35	26	28
List 2	16	31	37	41	37	41	39

sixth position for List 2 yielded significantly more correct responses than for the same position in List 1, $p < .01$.

For the Nonsemantic Control condition, there was no significant effect for serial position, $F(5,75) = .72$, $p > .05$, or for List 1 vs. List 2, $F(1,15) = 1.40$, $p > .05$. The interaction of these two variables was significant, $F(5,75) = 3.68$, $p < .005$. A Duncan-Bonnor test of the means revealed that the first position for List 1 yielded significantly more correct responses than for the same position in List 2, $p < .01$, and that the fifth and sixth positions for List 2 yielded significantly more correct responses than for those same positions in List 1, $p < .01$ and $p < .05$.

Discussion

Semantic vs. nonsemantic effect. On both LD measures, it was observed that switching semantic criteria from the first to second list facilitated later list differentiation. Changing nonsemantic criteria for the two lists, however, resulted in LD facilitation only as measured by the mean number of correct identifications and not as measured by the rating difference scores. It was also observed that the LD facilitation following nonsemantic analysis was significantly lower than LD facilitation observed following semantic analysis.

These results are reminiscent of those of Hyde and Jenkins (1969) and of Johnston and Jenkins (1971) who observed that recall was higher for those subjects who rated stimulus materials on a semantic criterion (e.g., "pleasantness-unpleasantness") than for those performing non-semantic analyses (e.g., detecting es and rhyming words).

Similarly, Bower and Karlin (1974) found higher recognition of faces when subjects had previously judged the pictures on "likeableness" or "honesty" rather than on gender. The authors suggested that this differential level of recognition might have been due to the number of descriptive features stored with the stimulus and that these features were extracted during the analysis of the stimulus to a subject's set of criteria for the experimenter's announced criterion. According to this explanation, semantic analysis leads to more information that a stimulus event has actually occurred.

It is possible that a similar explanation can be extended to LD facilitation, i.e., semantic analysis can lead to more information about when a stimulus event occurs. One might argue that nonsemantic criteria (particularly the criteria in this and previous studies) are rather well-defined and restricted in nature. Semantic criteria, on the other hand, are less restricted and generally more dependent on the subject's idiosyncratic knowledge of those criteria. For example, one's knowledge of money and finances might include last month's bank statement, bills that arrived recently, a job offer, an intended purchase, and so on, while one's knowledge of the vowels, the alphabet, upper case type, and the concept of length might be contained relatively within those very parameters. It is possible then that subjects making semantic judgments may compare the stimulus to a large number of prototypes or criteria, while subjects making nonsemantic judgments need only, and possibly could only, compare the stimulus to one prototype or a relatively small class of prototypes. This "poverty" of prototypes to which the stimulus

can be compared in the nonsemantic condition as contrasted to the potentially large number of prototypes available during semantic analysis might result in differential storage of information with the stimulus words. Consequently, in Experimental conditions where criteria are switched, semantic analysis would result in more information being stored with a word which would help in the assignment of that word to its appropriate list.

According to the above explanation, increased information stored with stimuli could aid LD performance directly or indirectly. It could directly aid LD performance by providing more information about specific list membership. This would be restricted to Experimental conditions in the present study where criteria are switched from one list to the other. It could indirectly aid LD performance by increasing the recognition of the stimuli. Subjects could experience difficulty in assigning words to the correct lists when they do not recognize the words as having occurred at all. This effect could apply to the Control as well as to the Experimental groups in the present study.

Nonsemantic task effects. Further analyses of the Nonsemantic groups revealed that both Experimental and Control subjects using "word with more vowels" (V) and/or "word presented in upper case type" (C) as criteria were better able to later assign word-pairs correctly to lists than were the Experimental and Control subjects using "first alphabetically" (A) and/or "longer word" (L) as criteria. Both Experimental groups scored significantly higher on the LD measure as compared to their respective controls. It appears then that one or both of the

tasks in the V-C conditions might result in better storage of the words in memory than either or both tasks in the A-L conditions.

Bower and Karlin (1974) suggested that memory for a stimulus event is dependent on the number of features stored with that event and that such feature storage is to some degree dependent on the number of prototypes or criteria on which the stimulus is evaluated. This notion may provide a theoretical explanation for the differences between the two Nonsemantic Experimental task groups if one considers the nature of each task.

In the C task condition, subjects were required to choose the word printed in upper case type. This required only one judgment (e.g., if the first letter of the first word the subject examined was a capital letter, the subject said that word first; if this word's first letter were in lower case type, the subject said the other word first). In the L task condition, subjects were again required to make only one judgment (e.g., scanning the two word display and saying the word first which extended farther in "space" than the other). For subjects in the A task condition, it was necessary to compare the first letter of both words simultaneously to the alphabet and say the word which came first alphabetically. The V task condition then appears to be potentially more complex as compared to these other three task conditions. Subjects were required to compare each letter of each word to the vowel set and then compare the total number of vowels for each word before making a decision.

The argument might then be that the stimuli presented during the V task were more enriched than words presented during the other types

of analysis tasks as more prototypes or criteria were necessary for pair judgments and consequently more features were stored with the memory trace of the words. Enhanced list discrimination in the V-C condition should result, and this was observed.

It is not necessary to consider at great length then the C task condition which was paired with the V task for the Experimental group. If more information about judgments about words had been stored with those words on one list and those words are then correctly assigned to that list, then the remaining words on the test sheet might be assigned to the other list by a type of process of elimination. That is to say, if one task results in a superior ability to later assign words presented during that task to the appropriate list, subjects might then infer that all remaining words (regardless of their memory for them) should be assigned to the other list.

If the stimuli are enriched within the vowel task, as suggested by Bower and Karlin (1974), one might expect this enrichment to enhance LD performance by Control subjects using this task as the criterion for their judgments. That is, as stimulus enrichment is facilitated, recognition might be facilitated and consequently the level of LD. Those subjects using this task should subsequently perform better than subjects using tasks evoking less stimulus enrichment (e.g., capitals, alphabet, and length) on LD tests. Analysis of the four nonsemantic tasks within the Control condition failed to indicate any significant differences among these four tasks. Furthermore, the performance on the vowel task was numerically lower than on the capital task. Although

it cannot be ruled out that the vowel task facilitated stimulus enrichment, these data do not offer independent support that such is the case.

While the above account suggests that the V task condition facilitated stimulus enrichment, one might argue that it is the differential study time during tasks which allows for later list differentiation. That is, in the A-L condition, one might speculate that essentially the same amount of time was needed to make either one or two judgments, while in the V-C condition, it is possible that more time (although reaction time was not measured) was needed to make V judgments than to make C judgments. A marked difference in functional study times might then explain the superior LD effects in the Experimental V-C condition. This notion is plausible as Winograd (1968) observed that the more discrepant the relative number of exposures to two lists, the greater the LD. Thus, it is possible that it is the trace that is stronger for V-C subjects rather than that there is more information retained about judgments for this group.

An explanation in terms of study time would lead to specific predictions about performances following each of the nonsemantic tasks within the Control condition. Winograd (1968) demonstrated that two strong lists (e.g., six presentations of each) are better differentiated than two weak lists (e.g., three presentations of each). Thus, if either the V task or the C task leads to relatively strong traces and the other leads to relatively weak traces, one would expect LD performance to be relatively high for one and relatively low for the other.

Although the differential study time explanation cannot be ruled out, the data failed to support this prediction.

It should be noted that the LD effect observed in the A-L condition may be overestimated. As Fritzen (Note 1) notes in his fifth and sixth experiments, subjects may be remembering the word they chose from each pair and making inferences as to the likelihood of the pair's membership from this information about the word chosen. For example, a subject may remember having chosen the word "circus" from a pair which also included the word "sweater". If the subject were in the A-L condition, he might then see that "circus" is shorter than "sweater" and comes first in the alphabet. Given this postpresentation analysis and his memory for having chosen the word "circus", he need only remember which task came first or second to assign that word-pair to the correct list. If such is the case, then during the LD task, the A-L task subjects may test the word remembered as chosen against the other member of the pair on two bases—alphabetical order and length (e.g., "circus"—"sweater" example previously cited).

The V-C task subjects, however, can only test out the word remembered as chosen against the other member of the pair on one base—number of vowels (as on the test sheets, all words were typed in the upper case). For example, a subject may remember having chosen the word "junction" from a pair which also included the word "bread". If the subject were in the V-C group, he might see that "junction" has more vowels than "bread", but he cannot, during the LD task, know that "junction" might have been chosen because it had been presented in upper

case type. He would then be unable to infer from this postpresentation test alone whether the pair had been on List 1 or List 2.

It may be then that the LD measure obtained from subjects within the A-L task condition reflected not only that information about judgments was stored with words but also that subjects in this group had other means (besides this stored information) on which to base list assignments. In other words, it is quite possible that even less information about judgments is stored with words presented during the A-L task than the LD measure suggests.

Serial position curves. It might be argued that the effect of switching criterion between two lists on later LD may not reflect the extent to which information is stored about judgments about words but rather the extent to which stimulus materials are attended to. That is, the LD facilitation observed in Experimental groups as compared to Control groups may be due to the maintenance of attention to the stimulus materials, while the lower LD scores by Control groups might reflect their fatigue or boredom with, and consequently inattentiveness to, the stimulus materials presented during the tasks. Switching criterion then might reestablish interest and/or attentiveness to the stimuli and subsequently facilitate LD scores.

Given this orientation, one would expect the serial position curves for the two lists to differ markedly between Experimental and Control groups. Specifically, one would expect the serial position curves for the two lists of the Experimental groups to be similar. For the Control groups, it would be expected that each subsequent serial posi-

tion would reveal fewer correct identifications and consequently all List 2 positions should be lower than their respective List 1 positions.

When serial position curves were examined for both lists within each group, it was found that the curves for the two lists for either Experimental group (Semantic and Nonsemantic) did not differ, while the two curves differed for both Control groups. Closer inspection of the curves for both Control groups revealed a greater primacy effect for the first list and a greater recency effect for the second list. Clearly, this observation fails to support fatigue or inattentiveness by Control subjects as an explanation of the LD effect.

The apparent primacy and recency effects one observes in both Control groups (if one considers the two lists of 24 word-pairs each as one list of 48 word-pairs) is of interest. Bjork and Whitten (1974) found similar effects (e.g., the primacy effect in the first list and the recency effect in the second list) on recall as the time between presentation and recall increased. They suggested that the effect may be due to the subject's memory of having seen two lists gradually changing over time to a memory of having seen one long list instead. This loss of list integrity might then be similar to this experimental situation of judging words on two lists on the same criterion. It is possible to ascribe a temporal dating hypothesis then to the LD facilitation observed in both Experimental groups. Such an hypothesis would suggest that switching criterion is a psychological event for the subject and that items prior to and following such a switch are "time-tagged" differently (e.g., Yntema and Trask, 1963), thus facilitating

later list differentiation. However, Fritzen (Note 1) has addressed this issue in a more direct test of the hypothesis and found no support for this as an explanation of the LD effects he observed.

Experiment II

Experiment II was designed to assess the effectiveness of semantic judgments which did not involve the comparison of two words. In the experiment, the lists contained single words as well as word-pairs. As in Experiment I, a criterion was announced for each list on which subjects were to base a judgment. In this experiment, however, single words were rated on a three-point scale of relatedness to a criterion, while paired words were compared as in Experiment I. The basic questions addressed by this experiment were whether an LD facilitation might be demonstrated for single words as compared to the appropriate control and if such a facilitation existed, how it might compare to the LD scores of words seen previously in pairs.

Method

Subjects. Sixty-four undergraduate students at Virginia Polytechnic Institute and State University served as subjects for the experiment. All received extra credit in their Introductory Psychology class for their participation. The subjects were assigned to one of two groups in alternating order as they arrived at the experiment.

Procedure. Word-pairs were presented alternately with single words on slides at 6 sec and 3 sec rates, respectively. Following the presentation of the 24 slides for the first list (12 word-pairs and 12 single words), the experimenter used 45 sec to change slide trays and

then presented a second list of 24 slides, alternately word-pairs and single words. Following the completion of the second list, subjects were instructed to work on a symbol cancellation task for 3 min and 45 sec. Subsequently, the experimenter gave each subject one of two test sheets (the first of two test sheets for word-pairs or the one for single words). Subjects were told which type of word (e.g., previously single or paired) appeared on the sheet which they were presented. The experimenter then instructed each subject to indicate on the six-point confidence scale (Winograd, 1968) which list each word had earlier appeared on. Subjects were allowed as much time as they needed for a sheet's completion, but each subsequent test sheet was given to them only following the completion of the prior sheet.

Prior to the presentation of the first list, subjects were told that they would be shown word-pairs and single words alternately projected on the wall. They were instructed that when word-pairs appeared, they would have 6 sec to pick the word more related to the announced criterion. They were told to say this word aloud and then say the remaining word in the pair aloud. When single words appeared, they had 3 sec to say the word aloud and rate the word on a three-point scale (1="highly related", 2="somewhat related", and 3="not related at all") to the same criterion as the pair judgments were based.

The experimenter then gave each subject one practice word-pair as well as one single word, two if necessary, to insure the subject's understanding of the instructions. The criterion used during the practice session was also used as the criterion for the first list. The words used during the practice session, however, did not later appear

during the presentation or during the test session.

Subjects were then told that they would make their judgments on a particular criterion (the same as that used during the practice session). For example, subjects might be asked to choose or rate the word-pair or word according to its relatedness to "Money and Finances". Subjects were not told a second list would follow.

Design. There were two groups of 32 subjects each, one a Control group, the other the Experimental group. Each group was presented two lists of words. Both word-pairs and single words within any one list were chosen or rated on the same criterion. The Experimental group judged each list on a different semantic criterion (e.g., "Farming and Agriculture" and "Money and Finances"), while the Control group judged both lists on the same semantic criterion. Following the presentation of the first list, the experimenter announced that a second list would be presented and that subjects were to continue to make their judgments in a manner similar to the first list to either the same (Control) or another (Experimental) criterion.

Each criterion appeared as the criterion for the first list for half of each group and as the second criterion for the other half of the Experimental group. For the Control group, the criterion used for the first list was again used for the second list.

Subjects were presented 36 of the 48 words from each master word list as used by Fritzen (Note 1). Word presentation was counterbalanced across subjects such that the 48 words from each master list appeared equally often as members of word-pairs and as single words. Each list

served equally often as the first and second list. Each test sheet served as the first test sheet for half of each group and as the second or third test sheet for the other half.

Materials. The words and distractor task were the same as in Experiment I. Test sheets were in the same format as Experiment I, except that all words were tested as single items. The words presented singly appeared on one test sheet, while words from pairs appeared singly on one of two other test sheets. Words previously appearing together in pairs were tested on different sheets.

Results

Rating difference scores. The Experimental group achieved LD scores of 2.60 and 2.19 for single words and words previously in pairs, respectively, while the Control group's LD scores were .67 and .33 for single words and words previously in pairs, respectively. An analysis of variance revealed a significant effect between Experimental and Control conditions, $F(1,62) = 128.86$, $p < .001$, and between single words and words previously in pairs, $F(1,62) = 7.46$, $p < .01$. No significant interaction was observed, $F(1,62) = .32$, $p > .05$. Thus, the LD facilitation task (switching criterion) enhanced the differentiation of both single words and words previously in pairs.

Of those words previously appearing in pairs, words chosen as more related during the presentation obtained an LD score of 1.51 and .26 for Experimental and Control conditions, respectively. Of those words not chosen (i.e., the other member of the pair), LD scores of .71 and .12 were obtained for Experimental and Control conditions, respectively.

An analysis of variance revealed a significant effect between Experimental and Control conditions, $F(1,62) = 159.72$, $p < .001$, and between those words chosen and not chosen, $F(1,62) = 55.06$, $p < .001$. The interaction of these factors was also significant, $F(1,62) = 27.47$, $p < .001$, so a post hoc Newman-Keuls test was used to compare the means. Results indicated that the means for both chosen and not chosen words within the Experimental condition were significantly greater than their respective controls, $p < .01$.

Proportion of correct identifications. The proportion of correct identifications yielded a pattern of results similar to that for the rating difference scores. For the Experimental condition, proportions of correctly assigned words were .77 and .73 for single words and words previously in pairs, respectively. The Control condition's proportions correct were .60 and .55 for single words and words previously in pairs, respectively. An analysis of variance revealed a significant effect between Experimental and Control conditions, $F(1,62) = 75.37$, $p < .001$, and between single words and words previously in pairs, $F(1,62) = 9.64$, $p < .05$. No significant interaction was observed, $F(1,62) = .20$, $p > .05$. Thus, switching criterion from List 1 to List 2 enhances differentiation for both single words and words previously in pairs.

Of those words previously appearing in pairs, words chosen as more related during the presentation were correctly assigned a proportion of .83 and .57 for the Experimental and Control conditions, respectively. Of those words not chosen, the proportion correct was .64 and .52 for the Experimental and Control conditions, respectively.

An analysis of variance revealed a significant effect between the

Experimental and Control conditions, $F(1,62) = 119.78$, $p < .001$, and between those words chosen and not chosen, $F(1,62) = 43.59$, $p < .001$. The interaction of these factors was also significant, $F(1,62) = 16.96$, $p < .001$. A post hoc evaluation of the means with the Newman-Keuls test indicated that the means for both chosen and nonchosen words within the Experimental condition were significantly greater than their respective controls, $p < .01$ and $p < .05$.

Single words: High vs. low rating. Of those single words rated as very related to the criterion, Experimental subjects correctly identified a proportion of .92, while Control subjects correctly identified a proportion of .65. Of those words rated as not related at all, the proportions of correct identifications were .69 and .58 for Experimental and Control subjects, respectively. An analysis of variance revealed significant effects between Experimental and Control conditions, $F(1,62) = 24$, $p < .001$, between high and low ratings, $F(1,62) = 26$, $p < .001$, and their interaction, $F(1,62) = 7$, $p < .05$. A post hoc Newman-Keuls analysis revealed that words in the Experimental condition, whether rated as high or low in relatedness to the criterion, were later correctly identified significantly more than their respective controls, $p < .01$ and $p < .05$.

Word-pairs: Consistency in responses. In order to ascertain consistency in responses for List 1 and List 2, Control subjects were randomly paired with the restriction that both subjects had seen the same words with the same criterion in the same order during presentation. For each subject-pair, the number of choices in common among

12 word-pairs were tabulated separately for List 1 and List 2. The means for the number of common choices were 8.56 and 8.18 for List 1 and List 2, respectively, $t(15) = .84$, $p > .05$.

Discussion

Single vs. paired effect. On both measures of LD, it was observed that switching criterion for the two lists facilitated later list differentiation for both single words and word-pairs. The observed facilitation was somewhat higher for those words presented previously as single words as compared to those words presented previously in pairs.

These results are contrary to those expected from a Bower and Karlin (1974) position if, as Rips, Shoben, and Smith (1973) suggest, a comparison of two instances to a category is a more accurate reflection of semantic memory than is the comparison of a single word to a category. One would have expected pair comparisons to result in better list differentiation.

A possible explanation for the failure to obtain higher list differentiation for the previously paired words is that those words were tested singly. If, as Neisser (1967) suggests, memory is a reconstruction of cognitive operations, then single items have the advantage of being tested in the same context as that in which they were originally presented, while paired items were tested in a context different from that in which they were originally presented. The contextual change could reduce the facility with which subjects reconstruct the original cognitive operations.

The effect of changing context is also related to encoding varia-

bility (Martin, 1968). Tulving and Thomson (1971) have proposed an encoding specificity hypothesis which is essentially an application of Martin's encoding variability hypothesis to situations where the study context differs from the test context. Tulving and Thomson demonstrated that the recognition of a single word was better if that word had been studied alone rather than in the presence of another item. It might then be suggested that words presented previously in pairs in this experiment were "contextually disadvantaged" at the time of the LD test and that such a disadvantage might explain the inferior LD scores for these words.

By testing all words singly, context bias may have been introduced, while postpresentation criterion testing was eliminated. It was argued in the Discussion of Experiment I that LD scores may not only reflect a subject's retention of information about judgments made but also his memory of word-choice. In this second experiment in which LD facilitation for word-pairs was demonstrated, the possibility of this latter factor was precluded. As all words were tested singly, a subject's memory of word-choice only could not lead to later "criterion testing" during the actual LD test. It is then possible that the list differentiation demonstrated in Experiment I was, at least in part, due to the retention of information about judgments made about word-pairs and that the LD demonstration for Experiment II was due to this retention of information about judgments made about words.

Chosen vs. nonchosen effect. Both LD measures revealed that switching criterion facilitated LD for both the chosen and nonchosen

members of word-pairs and that LD was significantly higher for words chosen as more related to the criterion as compared to those words not chosen. In other words, it appears that while information about judgments made about word-pairs is retained with both members of the word-pair (e.g., chosen and not chosen), more information may be available to the chosen member. The observation, however, that nonchosen words also exhibit the LD effect lends further support to the notion that the subject's memory of having chosen a word cannot alone explain the LD facilitation effects observed. This again leads to the conclusion that this observed LD effect is due to information about judgments made about words being stored. This also leads to the possibility that this storage is differentially favorable to those words chosen as related to a criterion. Such differential storage may be due to differential processing.

It is possible to compare this processing of material used within this experiment to a two-stage processing model proposed by Rips, Shoben, and Smith (1973) in which stimuli being compared to a category pass through one or two stages of processing. First stage processing consists of comparing characteristic features of the instance (word presented) to the characteristic features of the category (criterion). The second stage, in which defining characteristics are compared, might be bypassed, or in other words, processing stopped, when either a very high or low degree of common characteristics (extracted during the first stage) trigger a probabilistic "yes"- "no" decision to accept the instance as a member of that category. It is possible then that those words not

chosen as related to the criterion are processed less than those words chosen as related. That is, words assessed by the subject as not related to the criterion would likely be rejected following stage 1 processing, while words not falling above this threshold for a "no" decision would continue to the second stage of processing. As stimulus materials were selected rather randomly, it would be very unlikely that many words, on the basis of characteristic features alone, would have the high degree of commonality with the criterion necessary to trigger a "yes" decision at stage 1. However, for those words with some degree of commonality with the criterion, it would be necessary to continue to the second stage of processing before a decision could be made. This differential processing of these materials may then be a tenable explanation for the differences in LD between words eventually chosen as related to a criterion and those words not chosen.

Single words: High vs. low rating. This measure was taken to assess any subsequent effects on LD as a function of subjective rating of relatedness of a word to a criterion. The analysis revealed that words in the Experimental condition, whether rated as highly related or as not related at all to the criterion, were later differentiated better than those words in each respective Control group. However, those words rated as highly related were assigned to lists correctly more often than those words rated as not related at all. It then appears that on single word judgments, information about judgments about words is retained regardless of the word's relatedness to a criterion. This retention appears, however, to be more extensive for those words sub-

jectively assessed as related to the criterion. This observation is consistent with those observations made about word-pairs (e.g., words related more to a criterion are better differentiated on list membership than are those words judged as not related to the criterion).

It may be that, whether in word-pair or single word presentations, word-choice or rating of relatedness may be made at either of the two stages of processing. Words as the nonchosen member of a pair or given a "3" rating may be quickly processed at the first stage, while words with some relation to the criterion (which are later chosen from the pair or given the "1" rating) cannot be accepted or rejected at this first stage. Better list differentiation may then reflect some product of this more extensive processing.

Word-pairs: Consistency in responses. As previously proposed in the Discussion of Experiment I, the LD facilitation observed may be due to the Control subjects' gradually declining attentiveness to the stimulus materials, while the Experimental subjects' attentiveness was maintained or reestablished by switching criterion for judgments between the first and second lists.

If this is the case, then one might expect that Control subjects would base their judgments on word-pairs presented during the second list less consistently than they might on the first list. That is, while all subjects may be (as instructed) judging List 1 word-pairs on the experimenter's announced criterion, at the onset of the second list, or sometime during the second list, Control subjects might tire of using the criterion and cease to carefully attend to the second list's words.

Should this strategy of random choice be systematically used by the Control subjects, and subsequently affect later LD, one would expect a more random between-subject word-choice pattern on second list word-pairs as compared to their choices on the first list word-pairs. For example, if two subjects within the Control condition, using the same criterion for word-pairs, are randomly choosing words from pairs on List 2, one would expect less agreement between the two subjects on List 2 word-choices as compared to their agreement of choice on List 1.

Analysis of these randomly matched subjects revealed that agreement between the two did not differ significantly between lists and that this agreement was higher than as would be expected by chance responding alone. It appears then that Control subjects were attending to List 2 in a manner similar as that to List 1. This observation then leaves the inattentiveness explanation for LD facilitation again unfounded.

General Discussion

The present study confirmed the results of previous research which observed superior performance on memory tasks following "semantic" analysis as compared to "nonsemantic" analysis. The study failed, however, to confirm the predictions based on the Rips, Shoben, and Smith (1973) notion that comparison of two instances to a category is more reflective of semantic memory search than is the comparison of one instance to a category.

It remains questionable that the second experiment adequately translated the Rips et. al model but rather established a task similar in nature to those of Bower and Karlin (1974), Hyde and Jenkins (1969),

and Johnston and Jenkins (1971). That is, rather than asking subjects whether a word was a member of a category or which word within a pair was a member of a particular class, this second experiment required that subjects rate a word or choose the word more related to a broad class (e.g., not necessarily a member of that class). This difference between judging whether a word or a stimulus is a member of a class or merely related to that class may be the crucial reason for the failure of this study to confirm the predictions arising from the semantic analysis literature in light of the Rips, Shoben, and Smith (1973) notion of semantic analysis.

What we have dubbed as "semantic" may in fact be an orienting task which allows the subject a broader set of prototypes to which stimuli may be judged. As Bower and Karlin (1974) suggested, it may be that memory is facilitated following a search through a large number of prototypes when comparing stimuli to criteria. It then appears that it is not necessary to preclude meaning from the analysis of material to preclude this stimulus enrichment, but rather that the orienting task need only be somewhat limited in the number of prototypes to which the stimuli might be compared. That is, such tasks as length, vowels, capitals, and alphabet, do not preclude the need to understand and perhaps even have a lexical entry. The tasks do, however, limit the prototypes to which the stimuli might be compared (e.g., concept of length, "a,e,i,o,u,and y", "a,b,c,d...z", and "A" vs. "a"), while other tasks such as judging the relatedness of a stimulus to a criterion might include innumerable memories of the stimulus, the criterion, and

possibly a fabrication of their relationship as components of the subject's decision. This "prototypic potential" may then be an adequate explanation of the LD facilitation observed in this study.

For example, Anderson and Bower (1972) suggest that when a subject considers a word to a number of prototypes of a criterion, these prototypes as well as other contextual elements (the experimenter's announced criterion, the subject's mood, his posture, the experimental room, etc.) are recorded as a context bundle which the authors call a "list marker". The chance of a marker becoming associated to a word and of a contextual element becoming part of a bundle (the list marker) are described as probabilistic processes. The most salient feature, however, of this model is that information is stored (or can be stored) with the word by means of this list marker.

For example, if a subject is seeing a list of words which he is to rate on their relatedness to "Farming and Agriculture", he may later remember, when tested on the word "sweater", having thought about sweaters as being made of wool which was gathered on a sheep farm or as being worn on blustery spring days. Likewise, he might think of "sweater" in relation to "Money and Finances" as the sweater just recently purchased or as the sweater seen in an advertisement in the morning paper. When later asked to assign "sweater" to a list, the subject may try to recall what he thought about when he saw "sweater" during the presentation. On the basis of this information, the subject might then assign the word to either the "Money and Finances" or the "Farming and Agriculture" list. Interestingly, a number of subjects in this study confirmed

with the experimenter which criterion had appeared with the first and second lists. For Control groups (where the criterion for both lists were the same), it might be suggested that these subjects were disadvantaged on later list discrimination tasks as they did not have two explicit contexts associated with the two lists to which to refer when assessing a word's list membership. That is, they would have to rely on the strength of a word's trace and more implicit personal contextual elements (posture, mood, fatigue, etc.) as the differentiators, while Experimental group subjects would be able to use this information as well as the additional information related to two distinct criteria for making their decision.

It might be possible to explain the superior facilitation for list discrimination observed following semantic analysis as compared to non-semantic analysis by the probability that a larger number of prototypes (or contextual elements) become tagged with the list marker and subsequently to the word. The observation that the Nonsemantic Experimental group was somewhat better able to assign words correctly to lists than its respective Control group might reflect some tagging of the word to task criterion bundles, but, as argued earlier, the number of prototypes which might be a part of this contextual bundle is relatively small in comparison to those conceivable for semantic criteria. It may be this sparsity of evidence a subject has available to him which accounts for the inferiority of nonsemantic analysis to semantic analysis.

The Anderson and Bower (1972) model has been shown successful in the account of a variety of laboratory phenomena. First, the general

observation that distractor words which are associatively related to words which actually appeared previously are more likely to elicit false-positive recognition judgments than unrelated distractors might be explained by this model.

If a word which is not directly list-marked, by virtue of having not appeared on the list, calls to mind an associated word which appeared on the list and which had been tagged with a list marker, subjects might infer, from this indirect evidence of a list marker, that the distractor word had indeed appeared on the list. In a similar manner, words appearing on the list which were not tagged with a list marker might also be inferred to have appeared on the list by virtue of their elicitation of associated words which had been list marked. This increased probability of retrieving an unmarked list word then offsets the lowering of the "hit-miss" rate resulting from the false recognition of associatively related distractor words. This is consistent with the observation of Kintsch (1968) who reported that overall recognition did not change when distractor words were all highly associated to categories of words appearing during the stimuli presentation.

A subject's ability to "edit" implicit responses from recall might also be related to the Anderson and Bower model. Bower and Winchester (Note 2) reported 90% accuracy in discriminating between the experimenter's stimulus words and their associative responses following a lengthy word association task. Following a task in which subjects had to generate meaningful sentences from a list of unrelated words, Bower and Clark (1969) observed few intrusions of nonlist words in the written

recall. While it has previously been contended that this editing of noncritical material was possible because subjects consciously avoided associating implicit responses with list markers, the Anderson and Bower account may prove preferable. According to this model, the source of each item (critical or implicit) is associated with that item as a part of the contextual bundle. Consequently, the subject would have available these implicit responses to help mediate the recall of critical items and information concerning the source (e.g., the experimenter or himself) which might be used to differentiate between such implicit responses and critical items. Such an explanation would also incorporate the widely published observation that meaningful organization of verbal materials facilitates later performance on memory tasks.

Negative transfer was observed following part to whole learning of lists by Tulving (1966) and Tulving and Osler (1967) who observed that the learning of the whole list took significantly longer following the learning of part of that list than the learning of the list without this prior part-list learning. Such negative transfer might also be accounted for within the framework of the Anderson and Bower model. Bower and Lesgold (1969) reported that the words previously learned were those which took the longer to learn when incorporated within a larger list. According to the Anderson and Bower model, these previously learned words have acquired a list 1 tag and this tag subsequently hinders the tagging of the word with a list 2 marker. Subjects actively edit these words from their recall of list 2 because of the list 1 tag. Consequently, it is imperative for this observation of negative transfer that the subject

is uninformed that the part list is later contained within the larger list. Tulving (Note 3) observed results consistent with this contingency (e.g., that subjects aware that the part was included within the whole list did not reveal negative transfer in the relearning of these words). If, as this model suggests, part-whole negative transfer is a problem in recognition rather than in retrieval, one would also expect this same negative transfer when a subject learns the same list twice if he is under the impression that the lists differ at least in part. Such negative transfer was observed under these circumstances by Schwartz and Humphreys (Note 4).

Another laboratory phenomenon elucidated by the Anderson and Bower model is the estimation of stimulus frequency. Underwood, Zimmerman, and Freund (1971) found that frequency estimates increased linearly as the frequency of words in a list varied from one to six exposures. Such a linear function is consistent with the notion that frequency estimates are made by subjects on the basis of counting list markers associated with an item and that the probability of a marker becoming attached to a word on any one presentation is a constant, α , and independent of previous list markers. Thus, the Anderson and Bower model would predict frequency judgments to be a linear function of the number of exposures to a stimulus.

It now appears that our dichotimization between semantic and nonsemantic may exist only as a convenience rather than as true entities and that the memory of and about stimulus events may depend in large part on the number of ways we think about that stimulus. The Anderson

and Bower (1972) model presents a memory mechanism compatible with this notion that information concerning the context in which a stimulus is considered is stored with that stimulus.

Reference Notes

1. Fritzen, J. D. Unpublished manuscript, 1975.
2. Bower, G. H., & Winchester, P. Unpublished manuscript. (Anderson, J. R., & Bower, G. H. Recognition and retrieval processes in free recall. Psychological Review, 1972, 79, 97-123).
3. Tulving, E. Personal communication, 1971. (Anderson, J. R., & Bower, G. H. Recognition and retrieval processes in free recall. Psychological Review, 1972, 79, 97-123).
4. Schwartz, R. M., & Humphreys, M. S. List differentiation in part-whole free recall. Unpublished manuscript, 1971. (Anderson, J. R., & Bower, G. H. Recognition and retrieval processes in free recall. Psychological Review, 1972, 79, 97-123).

References

- Anderson, J. R., & Bower, G. H. Recognition and retrieval processes in free recall. Psychological Review, 1972, 79, 97-123.
- Bjork, R. A., & Whitten, W. B. Recency-sensitive retrieval processes in long term free recall. Cognitive Psychology, 1974, 6, 173-189.
- Bower, G. H., & Clark, M. C. Narrative stories as mediators for serial learning. Psychonomic Science, 1969, 14, 181-182.
- Bower, G. H., & Karlin, M. B. Depth of processing pictures of faces and recognition memory. Journal of Experimental Psychology, 1974, 103, 751-757.
- Bower, G. H., & Lesgold, A. M. Organization as a determinant of part-to-whole transfer in free recall. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 501-506.
- Craik, F. I. M., & Lockhart, R. S. Levels of processing: A framework for memory research. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 671-684.
- Gardiner, J. M. Levels of processing in word recognition and subsequent free recall. Journal of Experimental Psychology, 1974, 102, 101-105.
- Hyde, T. S. The differential effects of effort and type of orienting task on the recall and organization of highly-associated words. Journal of Experimental Psychology, 1973, 79, 111-113.
- Hyde, T. S., & Jenkins, J. J. The differential effects of incidental tasks on the organization of recall of a list of highly associated words. Journal of Experimental Psychology, 1969, 82, 472-481.
- Hyde, T. S., & Jenkins, J. J. Recall for words as a function of semantic,

- graphic, and syntactic orienting tasks. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 471-480.
- Johnston, C. D., & Jenkins, J. J. Two more incidental tasks that differentially affect associative clustering in recall. Journal of Experimental Psychology, 1971, 89, 92-95.
- Kintsch, W. Recognition and free recall of organized lists. Journal of Experimental Psychology, 1968, 78, 481-487.
- Martin, E. Stimulus meaningfulness and paired-associate transfer: An encoding variability hypothesis. Psychological Review, 1968, 75, 421-441.
- Neisser, U. Cognitive psychology. New York: Appleton-Century-Crofts, 1967.
- Rips, L. J., Shoben, E. J., & Smith, E. E. Semantic distance and the verification of semantic relations. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 1-20.
- Rowe, E. J. Depth of processing in a frequency judgment task. Journal of Verbal Learning and Verbal Behavior, 1974, 13, 638-643.
- Till, R. E., & Jenkins, J. J. The effects of cued orienting tasks on the free recall of words. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 489-498.
- Tulving, E. Subjective organization and effects of repetition in multi-trial free-recall learning. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 193-197.
- Tulving, E., & Osler, S. Transfer effects in whole-part transfer. Canadian Journal of Psychology, 1967, 21, 253-262.

- Tulving, E., & Thomson, D. M. Retrieval processes in recognition memory: Effects of associative context. Journal of Experimental Psychology, 1971, 87, 116-124.
- Underwood, B. J., Zimmerman, J., & Freund, J. S. Retention of frequency information with observations on recognition and recall. Journal of Experimental Psychology, 1971, 87, 149-162.
- Walsh, D. A., & Jenkins, J. J. Effects of orienting tasks on free recall in incidental learning: Difficulty, effort, and process explanations. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 481-488.
- Winograd, E. List differentiation as a function of frequency and retention interval. Journal of Experimental Psychology, 1968, 76 (2, Pt. 2).
- Yntema, D. B., & Trask, F. P. Recall as a search process. Journal of Verbal Learning and Verbal Behavior, 1963, 2, 65-74.

APPENDIX A

Recent followup research related to the original Hyde and Jenkins (1969) tasks has also focused attention on a theoretical orientation which might account for the observed superiority of semantic analysis on later recall as compared to nonsemantic analysis.

A central issue of earlier studies by Hyde and Jenkins (1969) and Johnston and Jenkins (1971) was the subjective organization of meaningfully associated words during recall. This associative clustering was found in groups who had previously rated words on a five-point scale of pleasantness, provided an appropriate modifier of the stimuli, or received no orienting instructions. Clustering was not observed in those groups assigned nonsemantic orienting tasks such as detecting letters, rhyming words, etc. This facilitation of associative clustering and of recall was observed in all semantic orienting task groups, regardless of their prior knowledge of a later memory test.

Such observations that associative clustering appeared consistently higher following semantic analysis led Hyde and Jenkins (1973) to suggest that information somehow "gleaned" from the materials during the presentation was later used during recall to organize highly associated words together. The authors also suggested that this "gleaning" of information was a type of processing available only to particular kinds of orienting tasks.

In the 1973 study, the authors observed that free recall was highest for subjects rating words on a five-point scale of pleasantness, followed, in order, by estimating frequency of word usage on a five-point scale, assigning a word to the appropriate part of speech category, detecting

es and gs, and choosing which of the two sentences the word would be grammatically appropriate in. Although words from lists which had low association values were less recalled than words from lists which had high association values, the level of recall by particular orienting task groups followed the same pattern.

Subjects informed of later recall performed better following semantic orienting tasks than when they had performed nonsemantic orienting tasks with highly associated word lists. No effect of prior knowledge of later testing was observed in the recall of low association word lists. Associative clustering was observed in both intentional and nonintentional conditions of semantic orienting tasks, although not observed following either condition with the nonsemantic orienting tasks.

Thus, it appears that similar responses, checking the appropriate response on a sheet for all orienting tasks, resulted in differential levels of recall and associative clustering. The authors suggest that subjects orienting to the semantic features of stimuli (semantic orienting tasks) process these words in such a manner that information related to these semantic features is available for the later organization of recall in a semantic fashion (as observed in associative clustering), while subjects orienting to nonsemantic aspects of the stimuli process words in a manner which does not make the semantic features available for later organization of recall.

Walsh and Jenkins (1973) suggest, however, that the results of the Hyde and Jenkins (1973) study may be due to the difficulty or effort necessary to perform a particular task rather than to the processing of

semantic features of the stimuli. That is, the authors argue that the tasks which Hyde and Jenkins (1973) contend allow the processing of a stimulus's semantic features might, on the other hand, be less difficult, allowing more time to review and rehearse the material, or require more effort, allowing more interaction with the materials, than those tasks suggested to be nonsemantic in nature.

To test these possibilities, Walsh and Jenkins (1973) compared the effects on free recall of singular orienting tasks and of combinations of orienting tasks. If degree of difficulty leads to lower free recall, one would expect that combination tasks would lower free recall as compared to the level of recall following either task alone. On the other hand, if increased effort facilitated later recall, high recall would be expected following combination tasks as compared to the level of recall following either task performed alone. A process model, as suggested by Hyde and Jenkins (1973), would predict the level of recall following combination tasks to closely resemble the level of recall following the task which, when performed alone, would exhibit the higher level of recall.

Tasks included rating words on a five-point scale of pleasantness, detecting es and gs, and estimating the number of syllables within a word. Later, rating on pleasantness and detecting a particular voice on a tape were used as orienting tasks. In both experimental conditions, results were consistent with the process model. Specifically, processing words as "meaningful units" (rating on pleasantness) led to higher recall than processing during other orienting tasks. Furthermore, when

such a semantic orienting task was used in combination with any of the other tasks, the level of recall was quite similar in magnitude to the level of recall following the semantic task alone.

In light of this finding, the authors voiced support of the original process model as proposed by Hyde and Jenkins (1973) that semantic processing of stimuli makes available information to be later used in memory tasks.

In a similar study, Hyde (1973) also found support for a processing model as an explanation of the superiority of semantic analysis on later memory tests. Hyde predicted that if the later inferiority on memory tasks following nonsemantic analysis was due to those tasks requiring more effort than semantic orienting tasks, then tasks requiring two judgments, whether semantic or nonsemantic, should lead to less recall than tasks requiring one judgment.

Subjects were assigned to one of six groups which rated words on pleasantness, stated whether words were active or passive, performed both of these two tasks, detected es in words, detected es and/or gs in words, or rated the words on pleasantness and detected es and/or gs in words. Control subjects were told of a later recall test and were given no orienting task.

On measures of recall and clustering, subjects using semantic tasks (one or two semantic judgments or semantic and nonsemantic judgments) did not differ from Control subjects, while subjects making nonsemantic judgments alone (whether one or two) scored significantly lower than both the Control group and the semantic task groups. The crucial

observations that differences in recall and clustering were not found between the number of judgments but rather between the type of judgment add further support to the processing explanation of semantic analysis facilitating later memory. The superiority of the Control condition in this study is at odds with the finding of Fritzen (Note 1) that a similar control group did not differ in LD performance from a group which used both lists with a single semantic criterion. It is possible that the Hyde task, which involved free recall, is more sensitive to the formation of higher order memory units which takes place in the no judgment control. Kintsch (1968) has demonstrated, for example, that list organization affects recall but not recognition.

Rowe (1974) suggested that this processing of semantic features represents a level of processing independent of the processing of other types of features. Rowe's observation that presenting stimuli in varying or constant orthographic displays resulted in similar levels of correct frequency judgments, while frequency judgments were adversely affected by orienting subjects in a task which was nonsemantic in nature (e.g., writing down the number of consonants and syllables within each word). Rowe suggested that correct frequency judgments were possible only when semantic encoding is available to the subject. That is, semantic encoding is possible when the subject is not told to orient towards any particular task or when the orienting task is semantic in nature. Rowe espouses the Anderson and Bower (1972) notion that for each presentation of a particular stimulus, a list marker is associated with the word. Frequency judgments, as noted earlier, arise from the

mere counting of such markers attached to the word. When words are shown in varying contexts or when semantic encoding is precluded, Rowe suggests that frequency judgments are hindered as retrieval of list markers to count becomes more difficult since the markers are stored with a number of different stimuli or as list markers fail to become attached to words whose peripheral (nonsemantic) attributes were analyzed.

In a within-subject replication of Hyde and Jenkins (1969), Till and Jenkins (1973) also illustrated the importance of varying contexts and of the nature of the orienting task on recall and clustering. In Experiment I, it was observed that words which were previously rated on pleasantness were later recalled more often than words whose lengths were estimated or in which es were detected. Intentionality had no effect on the level of recall nor was clustering evident for either the semantic or nonsemantic tasks.

In a second experiment in which related words were used, it was found that clustering was higher for words which had been rated on a semantic task as compared to words rated on nonsemantic tasks. The authors also observed that clustering was higher for words which had been previously analyzed on the same task. This suggests that information which may be a product of this analysis is stored with each stimulus and that such information is later used by the subject to organize his recall. While the observation that semantic analysis enhances later recall more than nonsemantic analysis replicates the reports of earlier research, the clustering with task type may suggest that particular information stored with each stimulus differs as a function of task type.

VITA

Mary Mapes O'Connor was born on February 21, 1951, in Sheffield, Alabama. She attended public school in Richmond, Virginia. Following her graduation from Douglas Southall Freeman High School, she attended Virginia Polytechnic Institute and State University where she received her Bachelor of Science degree in 1973 and was selected as a member of "Who's Who Among Students at American Universities and Colleges". She is also a member of Pi Delta Epsilon, a national honorary journalism fraternity. During her two years as a graduate student at Virginia Tech, Mrs. O'Connor served as an administrative intern with the Student Personnel Division of the University. Upon graduation, she will receive the Master of Science degree. She is presently a social worker with the State of West Virginia.

Mary Mapes O'Connor

EFFECTS OF SEMANTIC JUDGMENTS AND PAIR COMPARISONS
UPON LIST DIFFERENTIATION

by

Mary Mapes O'Connor

(ABSTRACT)

The present study investigated the effects of semantic and non-semantic analyses of words on list differentiation (LD). It was suggested that the memory of when stimulus events occurred might be due to the storage of information about judgments made about words and that this information may be differentially available to subjects as a function of the number of prototypes, or instances of an orienting task's criterion. In the first experiment, subjects judged two lists of word-pairs on semantic or nonsemantic criteria. The results indicated that information about these judgments about words was stored with semantic task subjects and to a lesser degree with nonsemantic task subjects. In the second experiment, the effects of word-pair comparisons and single word ratings on a semantic criterion were examined. It was observed that single words were later assigned correctly to lists more often than word-pairs, although information appeared to be stored with the word-pairs as well. It was also found that information about judgments about words is stored with these words regardless of their subjective relatedness to the orienting task's criterion. The results of both experiments were interpreted in terms of Anderson and Bower's (1972) theory which suggests that information about the context in which a stimulus is presented may become associated to that stimulus.