DOCUMENT RESUME

ED 082 170 CS 000 763

AUTHOR Underwood, Benton J.
TITLE The Role of the Association in Recognition Memory.
INSTITUTION Northwestern Univ., Evanston, Ill.
SPONS AGENCY Office of Naval Research, Washington, D.C. Personnel and Training Research Programs Office.
PUB DATE Sep 73
NOTE 77p.; See related documents CS 000 761 and CS 000 762

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Association (Psychological); Cognitive Processes; College Students; Criteria; Educational Research; *Memory; *Paired Associate Learning; Reading; *Recall (Psychological); *Word Recognition

ABSTRACT

The purpose of the eight experiments was to assess the role which associations between two words played in recognition decisions. The evidence on weak associations established in the laboratory indicated that association was playing a small role, but that the recognition performance on pairs of words was highly predictable from frequency information. However, the use of strongly associated words indicated that the strength of the association per se was not a critical variable in recognition performance. A post hoc expansion of frequency theory was proposed. Some unexpected findings included criterion differences in making frequency judgments as compared with recognition decisions, and criterion differences in recognition tests on homonym pairs as compared with other classes of word pairs. (Author)

ED 082170

THE ROLE OF THE ASSOCIATION IN RECOGNITION MEMORY

Benton J. Underwood

Northwestern University

September 1973

Project NR 154-321
Sponsored by
Personnel & Training Research Programs
Psychological Sciences Division
Office of Naval Research
Arlington, Virginia
Contract No. N00014-67-A-0356-0010

CS 000 763

THE ROLE OF THE ASSOCIATION IN RECOGNITION MEMORY

Benton J. Underwood

Abstract

The purpose of the eight experiments was to assess the role which associations between two words played in recognition decisions. The evidence on weak associations established in the laboratory indicated that the association was playing a small role, but that recognition performance on pairs of words was highly predictable from frequency information. However, the use of strongly associated words indicated that the strength of the association per se was not a critical variable in recognition performance. A post hoc expansion of frequency theory was proposed. Some unexpected findings included criterion differences in making frequency judgments as compared with recognition decisions, and criterion differences in recognition tests on homonym pairs as compared with other classes of word pairs.

THE ROLE OF THE ASSOCIATION IN RECOGNITION MEMORY

Benton J. Underwood[1]

Northwestern University[2]

At the empirical level the present studies were concerned with the role that an association between two words plays in the recognition performance for the pair. At the theoretical level, the studies were prompted by a theory which states that recognition decisions are mediated by frequency information (Underwood, 1971). This theory has evolved from work in which the unit of analysis was the single word; the association between words has received little attention in the development of the theory. For the single word, the theory, as applied to recognition decisions between old and new words, simply assumes that these decisions are made by discriminations of phenomenal frequency differences between the old and new words. Associations enter this formulation in only a vague way. It is known that the background frequency of words (as indexed by word counts) has little if any effect on judgments of situational frequency, i.e., frequency induced in the laboratory (Underwood, Zimmerman, & Freund, 1971). This must mean that the frequency information induced in the laboratory is kept distinct from background frequency and, therefore, in the vague sense noted above, situational frequency is somehow associated with the laboratory context.

The broad perspective of memory into which frequency theory fits assumes that various types of information may constitute a memory for an event (Underwood, 1969). Frequency information is simply one of these types, but a type that assumes (by theory) the major function in recognition decisions involving single verbal units. The theory has never presumed that other types of information will always be irrelevant for recognition. Indeed, it has been concluded that when frequency information becomes invalid for recognition decisions the subject will turn to other types of information in memory to mediate his judgments (Underwood & Freund, 1970). A recent study by Broder (1973) shows a serious breakdown of the theory when frequency information and associative information are in conflict. So, then, frequency theory can only be viewed within a larger perspective; the theory deals with one type of information in memory and its role in recognition performance. It is a part of the theoretical-empirical task to establish the limits of the theory, limits at which the theory is no longer useful in predicting the outcome of recognition studies. The studies to be reported were undertaken to determine if associative information between words leads to a breakdown of the predictions based only on frequency information.

There is a rapidly growing literature dealing with word recognition and the associative relationships among the words. These studies are often spoken of as studies dealing with associative context. The outcome of these experiments are presumed to be relevant to the question of whether or not recognition involves retrieval. The idea is that if an association influences recognition, retrieval processes <u>are</u> involved. The central method of study, with several variations, is that of removing, on the test, the items which occurred with the target word during study. If performance falls (as compared with the condition in which the context remains on the test) it is said to implicate the loss of retrieval cues. Context may be added also, but if this influences recognition, the theoretical interpretation is not clear with regard to the role of retrieval in recognition.

An examination of the literature shows no consistency in the effect of associative context on recognition. This confusion may be sampled. Tulving and Thomson (1971) reported that both the addition of context and the removal of context degraded recognition. Subsequently, Thomson (1972) found little effect of context addition. The data of Cofer, Segal, Stein and Walker (1969) showed that removing context caused a fall in recognition scores. Light and Carter-Sobell (1970) found that context change did influence recognition (performance fell) but in a further study Light (1972) reported that recognition for words embedded in sentences for study but tested alone gave recognition scores that were quite comparable to those obtained when the word was studied alone and tested alone. Ellis and Shumate (1973) showed no loss in recognition of stimulus terms from a paired-associate list over 28 days although the ability to recall the response terms fell over this period. Wood (1969) also concluded that the association developed in paired-associate learning did not influence recognition, although Wolford (1971) interpreted his results to indicate that the association between words in a pair was involved in recognition.

It is not the purpose here to attempt to evaluate the possible reasons for the contradictory findings. The details of some of these studies will be considered in conjunction with the specific experiments to be reported. It is merely noted here that the empirical facts concerning the role of associations in recognition decisions are far from clear.

One further background comment is required. It was noted earlier that the studies to be reported were undertaken with frequency theory as the background orientation. This theory fits nicely into the conceptual notions of signal-detection analysis. If, for example, an old and new word are given on a forced-choice test of recognition, the apparent frequency of the old item constitutes the signal, the apparent frequency of the new word, the noise. Words which have the same input frequency will constitute a distribution of phenomenal frequency for old words, and the phenomenal frequency of new words will consitute a distribution of noise. Two presumably independent measures of behavior may be obtained from signal-detection analysis. One of these is discriminability or sensitivity between the signal and noise distribution (between the phenomenal-frequency distributions for old and new words in terms of frequency theory). The measure d' is used to express sensitivity differences. The other measure has to do with the criterion set by the subject for accepting old words and rejecting new ones. Presumably the criterion represents a form of response bias which may differ among subjects and among conditions of an experiment, and it is measured by beta in signal-detection analysis. As will be seen, some unexpected differences in the criterion settings or response biases were found. However, while accepting the conceptual notions of signal detection, it has seemed unnecessary to use the measures of signal detection (d' and beta) to reflect the behavioral manifestations of these concepts. Rather, other measures have been developed. Nevertheless, the relationships between the measures used and d' and beta will sometimes be reported.

The eight experiments to be described divide themselves into two subsets. The first five were concerned with transient or weak associations developed in the laboratory between the two words in a pair. The remaining three used pairs of words already strongly associated by cultural usage.

Experiment 1

The first experiment was planned as an analogue of a previous study using single words (Underwood, 1972, Experiment I). In that study, comparative recognition decisions were evaluated in conjunction with comparative judgments of frequency under exactly the same conditions. The outcomes were statistically equivalent. Furthermore, when errors were made they occurred for the same items for the group making recognition decisions and for the group making frequency judgments. The correlation between the error distributions was .74. This finding was deemed consistent with the theory that recognition decisions are based on frequency information. In Experiment 1, reported below, pairs of unrelated words were used as the recognition unit, with frequency judgments and recognition decisions being made by different groups. The empirical question is whether or not the recognition decisions are predictable from judgments of frequency as was true when single words were the recognition unit.

Method

The general method may be described first. The Ss were presented a series of pairs for study, each pair being presented on a slide for 5 sec. Before presenting the pairs the Ss were given a booklet, and the cover sheet of the booklet contained the test instructions. These instructions were studied by the S before the list was presented so that he knew exactly how he would be tested. One set of instructions informed the S that he would be asked to make recognition decisions, both absolute recognition (YES-NO) decisions for individual pairs, and comparative or forced-choice decisions between two pairs, one old and one new. The two types of tests were illustrated on the instruction sheet. Another set of instructions, given to other Ss, informed them that their memories would be tested for frequency information. These instructions further indicated that there would be two types of test items. In one case, single pairs were shown and the S was required to fill in a blank with a number to indicate the number of times he thought the pair had been presented. In the other case, two pairs were shown and the S was required to encircle the pair which had been presented most frequently. Thus, all Ss received the same study list, and all Ss made both absolute and comparative judgments, but for one group the instructions emphasized correct recognition and for the other group, correct frequency information.

Materials. The pool of words consisted of 240 five-letter words with Thorndike-Lorge (1944) frequencies falling between 11-30. This pool consisted of approximately 90% of all five-letter words falling in this frequency range.[3] From this pool, 72 pairs were formed randomly subject only to the restriction that the two words in a pair did not have the same initial letter. The 72 pairs were used in the study list, with 48 pairs presented once, 12 twice, and 12 three times. The pairs were assigned randomly to these frequency categories. Pairs presented twice were first assigned positions randomly, once in each half of the study list. Pairs presented three times were then assigned, each pair occurring once in each third of the study list. Finally, the 48 pairs occurring once were assigned randomly to the remaining positions. Two buffer pairs were assigned at the beginning of the list, and two at the end so there was a total of 112 positions in the study list.

Test booklets. The test booklet consisted of three pages plus the cover sheet of instructions. There were 84 test items, 28 on each page. Each page included 16 items containing pairs that had been presented once for study, four that had been presented twice, and four that had been presented three times. In addition, there were four pairs on each page which had not been on the study list (new pairs). Half of the items in each class (excepting the new pairs) were tested by a forced-choice procedure, half by an absolute test. Thus, across the three pages, 24 pairs that had been presented once for study were paired with new pairs to produce 24 forced-choice items, and 24 pairs were presented alone for YES-NO decisions. Pairs presented twice and three times for study were also equally divided between forced choice and absolute judgments.

For absolute recognition, the words YES and NO followed the pair with the S instructed to encircle the appropriate response. For the judgments of absolute frequency, a blank, rather than YES and NO, followed the pair with the S requested to write in a number to indicate the frequency with which he thought the pair had been presented during the study list. In the forced-choice items the S encircled the correct pair (recognition) or the pair which had occurred most frequently during study (frequency judgments).

As noted earlier, new pairs were required for the forced-choice items. The 36 new pairs required (24 for the pairs presented once for study, six each for those presented twice and three times) had been formed randomly from the pool and were paired randomly with old pairs. On a single form only half the study pairs could be tested by a given method (forced choice or absolute judgment). Therefore, a second form was constructed in which the method of testing was reversed for the pairs. As a consequence, when considered across all Ss, each study pair was tested by a forced-choice procedure (for recognition and for frequency information) and also by an absolute test (YES or NO recognition, and judgments of absolute frequency).

On each page of the booklet the type of test item was randomized. Thus, S might have three successive items requiring a forced-choice decision, then an item requiring an absolute judgment, then an item for a forced-choice decision, then three items for an absolute judgment, and so on. On the forced-choice items the old and new pairs were randomly assigned to the left or right position. The instructions required that the S respond to all items on each page, guessing if necessary.

Procedure and Ss. The data were collected by a group procedure. As the initial step, the booklets were distributed to the Ss with instructions not to open the booklet. The experimenter informed the Ss that they would be presented a long list of pairs of words and that they should try to learn the pairs as pairs. They were further told that some pairs would be presented more than once. Next, they were asked to study the instructions in order to understand how they would be tested. Any questions concerning the instructions were handled individually. The list was then presented for study. After the last slide was presented the Ss were asked to review the instructions (if necessary) and then proceed to the first test page. The testing was unpaced.

The Ss were tested in groups of varying size, but within each group both frequency-judging instructions and recognition instructions were represented as well as were both forms. The booklets were randomized within blocks of four (two forms and two types of instructions) before they were assigned to the Ss. A total of 152 Ss was completed, 76 having had the recognition instructions and 76 the instructions for frequency judgments. Within each group there were 38 for each form. Since the forms did not differ in any of the analyses, no further mention will be made of this balancing variable.

Results

Pairs presented two and three times were included merely to make the frequency-judging task a reasonable one, particularly for the absolute judgments. The errors made on these pairs were few in number. The basic interest was for the pairs presented once for study. The results for the forced-choice tests will be considered first.

Forced-choice tests. There were 24 items each of which consisted of a pair presented once and a pair which was new. The Ss were required in one case to encircle the pair which had been presented for study; and in the other case, to encircle the most frequently presented pair. For the

76 Ss given recognition instructions the mean number of errors was 2.59 (10.8%), and for the 76 subjects given instructions to choose the most frequently presented pair the mean number of errors was 3.30 (13.8%). The difference between the two means was not reliable, $t\ (150) = 1.73$, $p > .05$. The discriminability or sensitivity, therefore, may be judged equivalent for the two groups. This result is the same as found previously with single words (Underwood, 1972). In the previous study it was found that the errors fell on the same items under both sets of instructions. In the present study the number of errors made on each of the 48 pairs presented once for study was determined for each type of instruction. The product-moment correlation between the two arrays was .43, $p < .01$.

Absolute tests. For these tests, each S was given 24 pairs presented once and 12 new pairs. In one case YES-NO decisions were requested, in the other, absolute judgments of frequency. The data will be viewed in a number of ways.

The frequency estimates may be handled in exactly the same way as are the recognition decisions. If S indicated that a pair not presented for study had a frequency of one or greater, it would be eqivalent to a false alarm for a YES-NO recognition decision. Similarly, if S assigned a zero to a pair that had been presented once, it would be equivalent to a miss in recognition. Throughout the studies to be reported, it has been found that the sum of the misses and false alarms is a simple and meaningful measure of discriminability or sensitivity. This measure and d' correlate highly. In the present experiment the correlation for the 76 subjects given the recognition test was -.86, and for those making absolute frequency judgments, -.91. These are somewhat lower values than those to be reported for later studies. In view of these relationships, the sum of the misses and false alarms (M+FA) has been used as the basic measure of discriminability.

The mean M+FA for absolute recognition was 7.66, that for frequency judgments, 6.92. These means do not differ statistically (*t*=1.20). However, these values alone do not adequately reflect the outcome for the absolute tests. The full picture can be seen in Fig. 1. This figure implicates criterion differences between those *S*s who made recognition decisions and those who made frequency decisions. The *S*s in making recognition decisions infrequently made false alarms but correspondingly had many misses. The *S*s in making frequency judgments behaved in the opposite manner; they made many false alarms (assigning a new pair a value of one or more) but had few misses.

The difference in the criterion set by the two groups is reflected directly by the interaction (F = 69.66) as seen in Fig. 1. An analysis of variance of data such as Fig. 1 is based upon allows conclusions about both differences in sensitivity or discriminability and differences in the criterion or response bias. If there is a significant main effect between recognition and frequency judgments, differences in sensitivity would be indicated. A main effect of input frequency would simply indicate the overall balance between the tendency to produce false alarms and the tendency to produce misses. A significant statistical interaction resulting from a convergence of the lines between input frequencies zero and one (whether a crossover or not) indicates a criterion difference.

The above technique for determining sensitivity and criterion differences seems straightforward. However, some comments should be made about other measures used to reflect criterion differences. Beta, the measure of the criterion setting used in signal-detection analysis, is the ratio of the heighth of the ordinate of the normal curve for the proportion of hits and the ordinate heighth for the proportion of false alarms. The beta distributions were severely skewed for the present data. Ignoring this skewness, the mean beta for recognition was 4.95, and for frequency judgments, 2.44 (t = 3.59).

Another measure which could be used to reflect differences in the criterion is M-FA/M+FA. Thus, if one S had two misses and eight false alarms, and another S had eight misses and two false alarms, sensitivity is equivalent but the criterion differs for the two (-.60 for the first, and +.60 for the second). Scores may vary from -1 to a +1. For the present data the mean values were .72 for recognition and .08 for frequency judgments (t = 7.14). This measure is curvilinearly related to beta. If zero false alarms and zero misses are assumed to be one miss or one false alarm, the rank-order correlation between the two measures (beta and M-FA/M+FA) is essentially perfect. The logic of changing zeros to a positive value (the value could be less than one, e.g., .1) is to avoid too many extreme scores (-1 and +1) which result when zero scores are entered. This simulates the situation for <u>beta</u> where no ordinate value is zero.

The values for M+FA are essentially equivalent for the two groups. Therefore, the criterion difference resulting from the formula M-FA/M+FA must come from the M-FA values. In effect, this is simply another way of calculating an interaction. For the present work, therefore, any criterion difference between conditions seems given most directly by the interaction of the analysis of variance between input frequencies and type of test. However one analyzes the data, it is clear that there were differences in the criterion setting for Ss given the recognition task and those given the frequency-judging task. As was the case with forced-choice tests, sensitivity did not differ for the two groups on the absolute tests.

<u>Item correlations</u>. The number of misses for each of the 48 pairs presented once was determined for recognition and for frequency judgments on the absolute tests. The correlation between the two distributions was .63 ($p < .01$). This indicates a substantial relationship between phenomenal frequency and recognition decisions. This relationship emerges in another analysis. The mean frequency judgments for the 48 pairs presented once were correlated with the number of misses for recognition. The value was -.50 ($p < .01$).

<u>Correlations by Ss</u>. A given S made both absolute and comparative frequency judgments, or both absolute and comparative recognition decisions. The relationship between the two types of judgments may be

examined. For frequency judgments the correlation for the 76 Ss between M+FA for 24 items and number of errors on the 24 forced-choice items was .65. For recognition, the corresponding value was 0.59. Both values are reliable statistically ($p < .01$).

Each S under frequency-judging instructions made absolute frequency estimates for 24 pairs which had been presented once for study. For each S the mean and standard deviation of these judgments was determined and correlated with the number of errors made on the forced-choice frequency judgments. The correlation between the mean frequency judgments and number of forced-choice errors was -.09, but that between the standard deviation and number of forced-choice errors was .44 ($p < .01$). Thus, the less the variability of phenomenal frequency for items with a constant input frequency the better the discriminability on forced-choice decision.

Discussion

The discussion of the findings will be brief at this point. Overall, the results showed that frequency judgments and recognition decisions for pairs of words have much in common, and these results could be interpreted as consistent with the idea that phenomenal frequency differences play a major role in recognition performance. Alternative interpretations are undoubtedly possible, of course, but such interpretations must include an accounting of the commonality in errors produced by recognition and by frequency-judging instructions.

The finding that there was a difference in the criterion set (on the average) for recognition judgments and for frequency judgments was quite unexpected. It was therefore decided to undertake a further study, with somewhat changed procedures, to see if the finding could be replicated.

Experiment 2

Method

In this experiment all Ss were given exactly the same study list as was given in Experiment 1. The two changes, which occurred on the test, will be described.

In Experiment 1 each S made both absolute and comparative judgments. In Experiment 2, each S made a single type of judgment on all items. Therefore, there were four groups of Ss. One group made only absolute frequency judgments, a second made only forced-choice frequency judgments. A third group made YES-NO recognition decisions, and a fourth made forced-choice recognition decisions. Of primary interest were the judgments made on the 48 pairs presented once for study.

It will be remembered that in Experiment 1 only 12 new pairs were used for YES-NO recognition and for absolute frequency judgments. In Experiment 2, 48 new pairs were included. Since there were also 48 pairs presented once, the number of true rights and wrongs was more nearly equalized in Experiment 2 than was true in Experiment 1. These 48 new pairs also became the new pairs on the forced-choice tests.

The booklet for the forced-choice tests (recognition and frequency judgments) consisted of the instruction sheet and two test sheets with 30 items on each for a total of 60 test items. These 60 items consisted of the 48 pairs presented once (each accompanied by a new pair) and six items in which the study pair had been presented twice and six in which it had been presented three times. Thus, only half the items presented twice and half those presented three times were tested. The purpose in this was to keep the total number of pairs to which the subject was exposed on the test equivalent for the forced choice and for the absolute tests.

The test booklet for the absolute judgments was made up of the instruction sheet and three test pages, 40 items to a page, for a total of 120. These consisted of 48 new pairs, 48 pairs presented once, 12 presented twice, and 12 presented three times.

All other details of the procedure were exactly the same as for Experiment 1. The Ss studied the instruction sheet which described how they would be tested, after which the list was presented for study at a 5-sec. rate. Again, the test was unpaced. The four instructional conditions were randomized in blocks of four so that for each session each was about equally represented. Sessions were continued until 36 Ss had completed the test for each instructional condition.

Results

Forced choice. The mean number of errors made by the 36 Ss on the 48 forced-choice items under recognition instructions was 5.42 (11.3%), and under instructions to choose the most frequent pair, 4.92 (10.1%). The t was .60. The correlation of errors by items was .54 ($p < .01$). These results confirm those of Experiment 1.

Absolute tests. The mean frequency judgments for the pairs presented 0, 1, 2, and 3 times may first be noted. The means in corresponding order were .18, .91, 1.74, and 2.49. The relationship is linear. Again, however, in the analyses to follow, no attention will be given to the pairs presented more than once since the errors on these pairs were few in number. As was done for Experiment 1, false alarms and misses were determined for the groups making absolute judgments. The results are plotted in Fig. 2. The interaction, although less severe than in Experiment 1, was still present, $F(1, 70) = 5.99$, $p < .05$. There was no appreciable effect of type of instruction ($F = 2.18$), indicating that discriminability was equivalent for both frequency decisions and recognition decisions. There was a main effect of input frequency ($F = 17.54$) indicating more misses than false alarms when combined across both groups.

The sum of the false alarms and misses correlated -.93 with d' for the frequency judgments, and -.94 for the recognition data. The difference in the criterion set by the two groups, as given in the interaction in Fig. 2, was not reflected in beta differences, the F being less than one.

Item correlations. The number of misses was determined for each of the 48 pairs presented once for study. This was done separately for the 36 Ss making absolute frequency judgments (a miss being an assignment of zero) and for the 36 Ss making YES-NO recognition decisions. The correlation between the two arrays was .55 ($p < .01$). The correlation for the number of false alarms on the 48 new items for the two groups was .42 $p < .01$).

If frequency information is involved in recognition decisions, it should be possible to predict forced-choice recognition and frequency judgments from the absolute frequency judgments. In the forced-choice tests, two pairs were involved for each item, one old (presented once) and one new. For each of these pairs a mean frequency judgment was available from the group making the absolute frequency estimates. It would seem that the greater the phenomenal frequency difference between the two pairs in the forced-choice item, the fewer the number of errors which should be made on that item. However, the variability of the frequency judgments should also be considered. For whatever reason, items having the same input frequency differ in the variability of the frequency estimates when viewed across the 36 estimates from the 36 Ss. It was shown in Experiment 1 that Ss having the lower variability of frequency estimates with constant input frequency did better on the forced-choice items than did the Ss with the higher variability. The same outcome may apply to item variability when calculated across Ss. Therefore, it would seem that the clearest prediction would be that the greater the mean difference in phenomenal frequency between an old and new pair in a forced choice item, adjusted for differences in the variability of frequency estimates for those pairs, the better should be the performance on the forced-choice test.

The mean frequency difference for each old and new pair appearing in the forced-choice items was divided by the square root of the sums of the variances of the frequency estimates for the two pairs. A high value should indicate good discriminability between the two pairs. The correlation between these values and the number of errors on the forced-choice recognition test was -.61. For the forced-choice frequency judgments, the correlation was -.79. Clearly, the phenomenal frequency of items presented once, and the variability of the estimates reflecting phenomenal frequency, predict performance on the forced-choice tests with some accurancy.

As one further bit of evidence on the relationship between frequency information and recognition performance, the absolute judgments made on the 48 pairs occurring once were combined across Experiments 1 and 2. Therefore, there were 74 Ss who had made absolute frequency judgments and 74 Ss who had made YES-NO recognition decisions. The apparent frequency of a pair was given by the mean of the frequency judgments, and for each pair a standard deviation was available. It would be expected that the higher the mean apparent frequency and the lower the variability the better should be the performance on recognition. The coefficient of variation (M/σ) was used to derive a single measure for each pair. A high coefficient should indicate good discriminability. These 48 values were correlated with the number of errors on YES-NO recognition for the 48 pairs. The product-moment correlation was -.70.

Discussion

Both experiments have shown that discriminability or sensitivity for pairs of words was equivalent when Ss were asked to make recognition decisions and judgments of frequency. This outcome held for both forced-choice items and for absolute judgments. Furthermore, the particular items on which errors were made in recognition tended to be the same items on which the errors were made when judging frequency and frequency differences. Mean frequency judgments for pairs combined with the variability of those judgments, predicted errors in recognition performance on both YES-NO and forced-choice tests. Viewed as a whole, the results were entirely consistent with the theory which specifies that frequency information is the major attribute of memory involved in pair recognition. In short, these two studies indicate that when pairs of words constitute the unit of analysis of recognition the outcome is much the same as when the individual word is the unit.

An unexpected finding in both experiments was that Ss making absolute frequency judgments set a more lenient criterion than did those Ss making recognition decisions. The reason for this difference is not known. An examination of the changes in performance across pages of the tests showed only very slight decrements under any of the conditions of testing, and the differences in the criterion were present initially and remained relatively constant across pages. Perhaps background frequency influenced the frequency judgments and not the recognition decisions although, as noted in the introduction, other data would argue against this possibility. By way of looking ahead it may be noted that a later experiment will show that criterion differences were found to be present for different classes of pairs of words for the same S. For the time being, therefore, the problem concerning the reason for the criterion differences will be set aside.

The two studies reported have established nothing concerning the role of the association in recognition. They have merely shown that frequency theory does not break down when pairs of words are used, pairs which presumably have some associative relationship as a consequence of the study trial. To assess in a more analytical way the role of associations in recognition requires different types of studies. The immediately following experiments represent some of these different types.

Experiment 3

Assume the S is presented a series of pairs of words under instructions to associate the words in each pair. Let three of the pairs be identified as A-B, C-D, and E-F. On the test of recognition he is presented A-B (intact pair) and C-F (broken pair). His instructions are to respond YES to all pairs providing both words in the pair had been presented for study whether paired during study or not. If the number of misses is greater for the broken pairs than for the intact pairs, the association would seem to be clearly implicated in recognition performance. If the misses are equivalent, it would be concluded that the association is irrelevant for recognition performance. The purpose of Experiment 3 was to make this test.

In addition to the two types of test items noted above, two others were included in order to have pairs for which the correct response was NO. One of these types consisted of one old and one new word, and the other type consisted of new pairs.

Method

Materials. The pairs of words were from the same pool as used in the first two studies. The S was presented 54 pairs for study. Of these, 18 were tested as intact pairs, that is, both words were old and paired as they were for study (O-O). A further 18 were re-paired words or broken pairs (O-Ob). One word from each of the remaining 18 pairs was used in pairs in which a new word constituted the other member of the pair (O-N). Finally, the test included 18 new pairs (N-N pairs).

For purposes of balancing conditions against possible pair differences, each pair was used in all conditions equally often. Originally, 72 pairs were assigned randomly to four subgroups of 18 pairs each. Across four forms each subgroup was used once for each of the four test types as described above. Thus, there were four different lists presented for study and four different test forms. The only additional words required were the 18 used as new words for the O-N pairs. These 18 words served this same function across all forms although, of course, the old words differed for each form. The O-Ob test pairs never consisted of an interchange of the members of two pairs. In the presentation list, each type of test item (except N-N of course) was represented in each successive block of three pairs. On the test, item types were randomized within blocks and particular items within blocks were random with respect to presentation position during study.

One further variable was introduced in the test. Half of the O-O pairs were tested in the order presented for study (A-B), and half were tested in reverse order (B-A). This was also true for the O-Ob pairs and for the O-N pairs. The latter became N-O pairs when reversed. The particular nine pairs to be reversed was determined randomly.

Procedure and Ss. The pairs were presented for study at a 5-sec. rate by a slide projector. The instructions emphasized the learning of pairs as pairs and, although the Ss knew their memories would be tested, the nature of the test was not specified before presenting the list. After the study of the pairs the Ss were given a two-page booklet, 36 pairs to a page, with the words YES and NO after each pair. They were instructed carefully that YES was the appropriate decision if both words had been presented for study whether paired together during study or not. The instructions further emphasized that if one or both of the words in a test pair had not been presented for study, the appropriate response was NO. According to these rules there were 36 pairs for which the correct response was YES, and 36 for which the correct response was NO, although the Ss were not told this.

Subjects were tested in small groups until 18 were completed for each of the four forms. Forms did not prove to differ statistically on the test, and they did not interact with item types. The results will be considered for the 72 Ss combined.

Results

The overall results are shown in Fig. 3, where percent errors is used as the recognition measure. An error would be a false alarm for N-N and O-N, and a miss for O-O and O-Ob. It is to be noted that there were about 7% more misses on O-Ob pairs than for O-O pairs. The overall analysis showed a significant effect of item type ($F = 37.58$). A standard error of the mean difference derived from the error term indicated that any difference as large as 4.3% was reliable at $p = .01$. All differences between item types were larger than 4.3%. The fact, therefore, that type O-O results in fewer errors than type O-Ob may be taken as evidence that the association between the two words in a pair played a role in recognition.

The misses on the O-O pairs were more frequent than were the false alarms on the N-N pairs. The M+FA measure of sensitivity for these pairs correlated with d' across the four forms at the following levels: -.93, -.94, -.91, -.95.

The O-N pairs produced nearly 35% false alarms. A satisfactory accounting of this result was not found. Several subsidiary analyses were done to search for leads. The 72 Ss were divided into two subgroups based upon the total errors made on all item types. The relative relationship among the four item types were exactly the same for these two subgroups of good and poor Ss. There were two pages to the test form, and the Ss completed the first page before moving to the second. It seemed possible that testing effects might have differed for item types. However, an analysis showed no difference in performance on the two pages and no interaction between pages and item types, both Fs being less than 1.

It will be remembered that the position of the words in a pair during study was reversed for half the pairs on the test. This change had essentially no effect on the O-O pairs, the nonreversed pairs producing 20.7% errors, the reversed pairs, 19.6%. For the O-Ob pairs the corresponding values were 25.8% and 27.9%. But, for the O-N pairs the false alarms were 38.0%, and when reversed, (N-O) 31.6%. This is to say that when the old word was the first member of the pair, errors were more frequent than when the new word was the first member. However, even for the latter pairs, performance was poorer than was found on any of the three other item types.

It is difficult to see how criterion differences could exist for the various item types, since if the S could classify an item type he presumably could provide the correct response. Frequency theory might predict good performance on the O-N pairs on the grounds that the two words would represent a contrast in apparent frequency and this contrast would lead to a correct rejection. In one sense the results for the O-N pairs are in contradiction to the notion that an association facilitates recognition. The two words in the O-N pairs were not associated during study just as the O-Ob pairs were not associated. If the lack of an association leads to a NO decision for the O-Ob pairs it should do the same for the O-N pairs, but in the latter case the decision would be correct. It is possible, of course, that this effect was present and kept the number of errors from being more frequent than was actually found. For the time being, the mechanisms responsible for the heavy error rate on the O-N pairs cannot be specified.

One other fact will be reported. When words are paired randomly the resulting pairs should show a distribution defined in terms of the ease with which the words within the pairs could be associated. There is a possibility that a pair which is easily associated when presented for study will also produce errors (false alarms) when it is used as a new pair on a test. This does not seem to be the case. In the present experiment, the 72 pairs were used both as old pairs (O-O) and as new pairs (N-N). The number of false alarms made on the 72 pairs when they were used as new pairs did not correlate with the number of misses which occurred when they were used as old pairs ($r = -.03$).

Experiment 4

The results of Experiment 3 were interpreted as demonstrating that an association between two words, developed on a single study trial, facilitated recognition of the pair. Or, as is sometimes said, associative context aided recognition performance. Experiment 4 was another test of the role of associative context using a different approach, although one that is similar to those used by other investigators, particularly DaPolito, Barker, and Wiant (1972). These investigators presented a triad of words for study. On the test, the Ss made recognition decisions on single words, but other items were sometimes present in the display when the decisions were being made. Essentially any change at test produced an increase in misses. The Ss were not instructed to associate the words in the triads during study, and since the triads were presented for only 1 sec., associative formation between the items would be minimal. Still, the results showed that when the order of the three items as presented for study was changed on the test, a decrement was observed, suggesting either that item-order information was important or that associations between the item and its position within the triad was a part of the memory. The method used by DaPolito et al. makes context (when defined as the presence of an item or items on which no recognition decisions are to be made) seemingly incidental. To some degree it would seem that S could ignore the context words on the test. The fact that context effects were found in spite of this seems to speak strongly that context, whether known to be associative or not, does influence recognition.

In the present experiment, Ss were presented pairs of words for study under instructions to associate the words in each pair. On the recognition test, either single words or pairs of words were presented. In the latter case the recognition decisions were made only on the second word in the pair. Thus, the first word in the pair need not have been involved at all in the recognition decision although the Ss were told that it might help them in reaching a decision on the second. Of course, the major variable was the nature of the first word in the pairs.

Method

A description of the seven classes or types of test items will be given first, expanding the symbol system used in describing the pairs in Experiment 3. An old item is designated O, a new item, N, and (1) the test word was always the second word in a pair on the test, and (2) a word on the test always occupied the same position in a pair (first or second member) as it held during the study trial. The seven types of items were as follows:

Type O: old second word tested alone.

Type N: new word tested alone

Type O-O: intact pair

Type N-N: new pair

Type N-O: old second word with new first word

Type O-N: new second word with old first word

Type O-Ob: two old words from different study pairs.

Materials. The words in the pairs were all four-letter, monosyllabic words. They were drawn from a pool of 312 such words which constituted a random sample from the Thorndike-Lorge (1944) tables. The change in the class of words (from the class used in the first three experiments) was necessitated by the fact that other experiments being conducted in the department were using the words from the earlier experiments.

Study and test lists. Initially, 96 words were drawn randomly from the pool and formed into 48 pairs randomly. These 48 pairs constituted the study list, hence the old pairs, for all Ss. A second sample of 84 words was drawn randomly and these 84 words were always used as new words on the test form.

For balancing purposes, the 48 study pairs were arbitrarily divided into four blocks of 12 pairs each. Across four test forms, each block of 12 pairs was used once for each of the four item types on the test involving old words in the second position (O, O-O, N-O, O-Ob). This means, therefore, that on a test form there were 12 items representing each of the four types. The O-Ob pairs were never constructed by interchanging the members of two pairs.

The assignment of new items to particular functions on the test forms was done randomly, with a different random assignment for each form. There were 12 Type O-N test pairs, the O words being the first words in the study pairs which on the test became Type O. There were also 12 Type N pairs on the test, but 24 pairs were used to represent Type N-N. This made 48 true YES items on the test and 48 true NO items. On the test, the seven item types were block randomized, and test position, relative to input position, was orthogonal by halves. There were 48 items on each page of the two-page booklet.

The single items on the test always occupied the nominal second position of a pair. For example, the first four items on one of the test forms were as follows:

1. spry-wile YES NO
2. whey-ramp YES NO
3. jest YES NO
4. came-dell YES NO

Procedure and Ss. The 48 study pairs, preceded and followed by two filler pairs, were presented at a 5-sec. rate. Prior to presenting the pairs, the Ss were told only that pairs would be presented and that they should try to associate the words in each pair in preparation for a memory test. After the last pair has been presented, the booklets were distributed. The Ss were run in groups of varying size but in any group the four test forms were assigned randomly to the Ss. The instructions for the test involved five key points: (1) for a single word on the test form, the YES-NO decision was whether or not the word had been shown as a member of a pair; (2) for a pair of words, the decision was t be made only for the second word, i.e., had it or had it not been presented as a second word in a pair on the study list; (3) the first word in a pair might aid in making the decision on the second; (4) an old word always occupied the same position on the test form as it had in the pair in which it appeared on the study list; (5) all items were to be completed, guessing if necessary. The test was unpaced.

Groups of Ss were tested until 68 Ss had been completed, 17 for each of the four forms.

Results

The results, shown in Fig. 4, are quite unambiguous. There were only two levels of errors, differing about 8-10%, and these two levels were associated with change of context and with context constancy. For the misses, omission of the first word (0), including a new first word (N-0), or re-pairing old words (0-0b) all resulted in about the same increase in the number of errors when no change (0-0) was used as a reference. False alarms were also correspondingly influenced when an old word was used as a first word (0-N). A number of different ways of viewing the differences statistically all produced the same conclusions. Perhaps the most direct way is to ask about the differences among the four types for the misses, $F(3, 201) = 7.87$, $p < .01$, and among the three types for the false alarms, $F(2, 134) = 8.66$, $p < .01$. In both analyses, the deviation of one condition from the others is largely responsible for the conclusion that reliable differences were present.

It will be remembered that in Experiment 3, Types N-0 and 0-N produced more errors than did Type 0-0b. In the present data, where S was asked to make decisions on a single word rather than on a pair as was true in Experiment 3, the numbers of errors for these types were essentially equivalent. Also, in Experiment 3 the misses on Type 0-0 exceeded the false alarms on Type N-N, which was not true in the present experiment. This is to say that the two techniques (judging single words in pairs versus judging pairs) produced differences in detail, but both lead to the same conclusion. The presence on the test of the full associative context present during study facilitated recognition. Any changes made in the context caused an increase in misses.

Experiment 5

Experiments 3 and 4 gave evidence that an intact association between two words in a pair is a positive factor in the recognition of that pair. The present experiment represents a further test of the role of an association in recognition, a test which makes use of paired-associate learning.

It was mentioned earlier that when words are paired randomly one would anticipate that some of these pairings would consist of words which would be easy to associate and others which would be difficult to associate, with the bulk of the pairs lying in between the extremes. If an association between two words influences the recognition of the pair, it might be reasoned that old pairs on which few errors were made developed stronger associations than did pairs on which many errors were made. To say this another way, during a constant period of study, the strength of the acquired association would differ for different pairs. If the strength of the association is involved in recognition, it must be predicted that pairs on which few recognition errors were made would be learned more rapidly as paired associates than would pairs on which many errors were made. Experiment 5 was a test of this proposition.

Method

Materials. From among the 48 pairs presented once for study in Experiment 1, and tested by a YES-NO procedure, 10 pairs on which few recognition errors were made were selected along with 10 pairs on which the most recognition errors were made. As paired-associate lists, these will be called the Easy List and the Hard List. Both lists are shown in Table 1. Although these lists were selected by using the data of Experiment 1, the same pairs occurred under much the same conditions in Experiment 2. Therefore, the errors made in Experiment 2 are also shown. The reliability of pair difficulty is by no means perfect, particularly for the pairs in the Hard List, although some regression would be anticipated. Nevertheless, the mean errors were substantially different for the two lists even when gauged by the errors made in Experiment 2.

Procedure and Ss. The lists were presented for one study trial and five anticipation trials at a 2:2-sec. rate on a memory drum. There were 20 Ss. Half learned the Easy List followed by the Hard List, and half learned in the reverse order.

Results

As may be seen in Fig. 5, the list made up of pairs defined as easy by recognition errors was much easier to learn than the list made up of pairs on which many recognition errors were made. This was true regardless of the order in which the lists were learned and the overall difference across the five trials was highly reliable ($F = 28.09$).

The mean number correct for each pair, summed across the five trials, is shown in Table 1. Only two pairs from the so-called Hard List fall within the range of number of correct responses as shown for the pairs in the Easy List. Other information is included in Table 1. In Experiment 1, 38 Ss made absolute frequency judgments of these pairs, and in Experiment 2, 36 Ss did the same. The mean apparent frequency as derived from these 74 Ss for each pair is shown in Table 1. The difference in the mean frequency judgments for the two lists is significant, $t\ (18) = 3.69$, $p < .01$. Thus, the pairs in the Easy List had higher apparent frequency than did those in the Hard List.

As shown earlier, prediction of recognition errors from absolute judgments of frequency was quite good when means and variability were considered simultaneously by using the coefficient of variation (M/σ). These values also appear in Table 1 and it is to be noted that the two lists differ markedly on this measure. Rank-order correlations were calculated across the 20 pairs. The M/σ values correlated .86 with recognition errors and .56 with the number of correct anticipations in paired-associate learning. The correlation between recognition errors and paired-associate learning was .72.

Discussion

The present results indicate that the ease of associating two words in a pair is positively related to correct recognition. These results support those of Experiments 3 and 4 in showing that an association between the two words in a pair seems to be involved in recognition performance Insofar as apparent frequency has been shown to predict recognition errors it is not surprising that apparent frequency also has some predictive value for paired-associate learning.

The evidence that an association between two words influences recognition must be kept in theoretical perspective. The major information used in recognition decisions may well be frequency information -- apparent frequency of the pair as a unit and of the individual words, each as a unit. The difference between intact and broken pairs in Experiments 3 and 4 was only 7-10%; performance did not fall to chance when associative context was removed. Thus, although the results of the experiments thus far indicate that an association plays a role in recognition, the magnitude of the role is not impressive when overall recognition performance is considered.

Experiment 6

Two of the basic findings from the preceding experiments led to Experiment 6. Experiments 3 and 4 indicated that an association resulting from one presentation of a pair of words adds information which aids recognition performance. Experiment 5 showed that pairs on which few recognition errors were made were more rapidly acquired in a paired-associate list than were pairs on which many recognition errors were committed. These facts suggest that the associability of two words in a pair facilitated recognition performance, or, that facilitation of recognition of a pair by the association existing between the two words was directly related to the strength of that association. An obvious projection from these facts is that if very strongly associated pairs of words were tested for recognition, few if any errors would be observed. The present experiment deals with this possibility.

Many of the pairs used in Experiment 6 were strongly associated as a consequence of cultural usage. Other investigators have approached the problem of associative context by the use of similar materials. Tulving and Thomson (1971) presented S strongly associated pairs of words as well as weakly associated pairs as indexed by published word-association norms. On the recognition test the S was asked to make a decision on each word of a pair independently. Based on the results of this procedure it would be concluded that the strength of an association had little influence on recognition. There were 12% misses on the strongly-associated pairs, and 21% false alarms on new associated pairs. For weakly-associated pairs the values were 15% and 16%. However, these data came from only a small part of an extensive procedure of testing which involved changing words in pairs, testing single words, adding new words to words which during study had been members of pairs, and so on. It is possible that the varied test context, and testing the words in a pair independently, may not have allowed differences as a function of strong and weak associative context to manifest themselves. Thomson (1972) used strongly associated pairs in one experiment (Experiment 1) and pairs with no pre-experimental associative relationship in another (Experiment 4). The Ss made decisions on each word in a pair. The misses were 5% less for the strongly-associated pairs than for those which were weakly associated, with the false alarms about equivalent for new pairs. These data, like the Tulving-Thomson data, suggest only a small facilitating effect of the culturally established associative relationship. Such findings seem odd if one is to assume that associative context is a powerful factor in recognition memory. Further investigation seemed necessary.

In the present experiment the S studied pairs of highly associated words and pairs of initally nonassociated words. Recognition decisions were always made for pairs as pairs.

Method

Materials. From a variety of sources, 100 pairs of words were brought together. As may be seen in Table 2, there were 20 pairs in each of five classes. Four of these classes were assumed to represent pairs with strong, culturally established, associative relationships.

These four classes are the parallel associates, synonyms, antonyms and conceptual associates. The fifth class, homonyms, represent a special class. There is no reason to believe that the two words in a homonym pair are associatively related in the sense that the words in pairs in the other classes are associated. The distinguishing property of the homonym pairs for the present study was that the acoustic-articulatory response to the two words in a pair would be equivalent. The consequence is that the frequency of this response would be double that for the words in the other types of pairs. It was believed that this property would yield information not only relevant for frequency theory, but also relevant to an understanding of the mechanisms of recognition in the associated pairs.

No argument is to be made for the purity of the pairs in the classes, nor that the pairs are necessarily representative of all possible pairs which might be placed in these classes. But, taken as a whole, each class is assumed to be distinctly different from each other class.

In addition to the 100 associated pairs, 100 nonassociated pairs were formed. In constructing these 100 nonassociated pairs, 100 additional words were chosen, these words varying in length, frequency, and form class. Some examples are: assail, degree, lid, oath, worthless. These words were paired randomly with one word from each of the 100 pairs listed in Table 2. Half the time (within each class), the first word in the pair as listed in Table 2 was retained, and it occupied the first position in the pair, the second position being held by a neutral word. For the other half of the pairs, the second word in the pair was retained, the first position being occupied by the neutral word. For example, using the pair cup-saucer in the list of parallel associates, the nonassociated pair became cup-artist, and using bed-sleep, the nonassociated pair became utter-sleep. Thus, the highly associated pairs (to be called E Pairs) and the nonassociated pairs (to be called C pairs) always had one word in common.

Study lists and test forms. Each S was presented 50 pairs on the study list (plus two filler pairs at the beginning and two at the end). These 50 pairs were made up of five E Pairs and five C pairs from each of the five classes. There were 100 pairs on the test forms consisting of the five E and C Pairs from each class presented for study plus five E and C Pairs from each class not presented for study. Of course, no word appeared in both an E Pair and a C Pair on the study list or test. For any given S, only half of the E Pairs and half of the C pairs were used. Across four study lists and four test forms, each E pair served as an old pair and as a new pair, and this was also true for each C Pair. This was accomplished by dividing the pairs in each class into four subgroups of C and E Pairs of five pairs each and rotating functions (old or new) across forms.

For the study list the 10 item types (five E Pairs and five C Pairs from each class) were block randomized across the 50 positions. The test forms consisted of 100 pairs, 50 on each of two pages, with the words YES and NO appearing after each pair. The order of the pairs on the test forms was random, but with a different random order being used for each form.

Procedure and Ss. Prior to presenting the list for study, the Ss were fully informed concerning the nature of the test. They were told to try to associate the words in each pair on the study trial. The list was presented at a 5-sec. rate by a slide projector. After the list was presented the test booklets were distributed and the instructions for the test repeated. The Ss were informed that they must make a decision for each pair, guessing if necessary.

The Ss were run in small groups until 21 had completed each of the four forms. Since forms did not interact with the variables of experimental interest, the data have been summed across forms for the 84 Ss.

Results

Data were available on 20 different types of pairs, each type represented by five pairs. These types were: old E Pairs in each of the five classes; new E Pairs from each class; old C Pairs from each class, and new C Pairs from each of the five classes. For the initial analysis, the sums of the misses and false alarms (M+FA) was determined for the C Pairs and for the E Pairs for each class. (It may be noted that d' correlated .98 with M+FA when summed across the five classes for the old and new E Pairs). The means of the M+FA score are shown in Table 3. It will be noted that for only two of the five classes (parallel associates and synonyms) are there fewer errors for the E Pairs than for the C Pairs. For the other three, the direction is reversed, although in none of the five classes is the difference between E and C Pairs of appreciable magnitude. Nevertheless, an analysis of variance showed that class was a significant source of variance, F (4, 332) = 7.87, $p < .01$, as was also the interaction between class and E-C Pairs, F (4, 332) = 4.18, $p < .01$, but the difference between E and C pairs was not, F (1, 83) = 2.25, $p > .05$. These findings indicate that, overall, associated pairs are not better recognized than nonassociated pairs.

The measure of sensitivity, or discriminability, M+FA, has indicated that across the five classes the discriminability of pairs of highly associated words was not better than the discriminability for pairs with the minimal associative relationship that might have been established on a single study trial. The interaction between type of pair and class, however, indicated that discriminability differed as a function of class of pairs. An examination of the results for each class in more detail is indicated.

In presenting the results for the first two experiments, it was pointed out that criterion differences could be detected by a certain type of interaction when two conditions, both represented by old and new items, were plotted on the same graph. These requirements were met in the present data for each class since there were old and new E Pairs and old and new C Pairs. The data are plotted in Fig. 6 for each class, using input frequency (0 and 1) rather than old and new along the baseline.

The critical fact exhibited in Fig. 6 is that S may maintain different criteria for different classes of pairs even when those pairs are intermingled on the same test form. The results for the homonyms show that S has a very lenient criterion for accepting pairs which were not in fact on the study list. The marked interaction between input frequency and E-C Pairs is evident. The results for each of the five classes of pairs have been analyzed separately. Each F is based on 1 and 82 df, and an F of 3.96 is required for the .05 level, 6.96 for the .01 level. Input frequency was significant for all classes, and except for the homonyms (which shows the smallest F, 9.51), more errors were made on old than on new pairs. The E-C difference was significant for only the antonyms (F = 9.15) and homonyms (F = 4.02). The interaction between the two variables was reliable for parallel associates (F = 6.65), synonyms (11.15), and, of course, for the homonyms (F = 59.67). The interaction for the parallel associates and for the synonyms cannot be interpreted as representing a criterion difference for the E and C Pairs, since the differences in errors are small in magnitude for new (zero frequency) items.

An examination of performance on the two test pages separately showed no systematic changes for any class of item. The differences in the criterion imposed for the different classes, therefore, was not a consequence of a gradual change or shift as the S made successive judgments.

It will be remembered that one word in each C Pair was tied to a word from a pair in one of the five classes. There is no reason to expect that these pairs should differ in number of errors across classes unless the words in the different classes differed in recognition memoriability in some way. Table 3 did suggest, however, that small differences were present among the C Pairs representing the different classes. Statistically, the M+FA values for the C Pairs for the five classes was reliable, $F(4, 332) = 4.00$, $p < .01$. The greatest number of errors occurred for the C Pairs derived from words making up parallel associates and synonyms, and the fewest from the conceptual associates.

Discussion

Four classes of natural associates have given no consistent evidence that the associative relationship facilitated recognition performance. Overall, control pairs, where the associative relationship would be minimal after a single study trial, produced no more errors than did the associated pairs. Even if it is presumed that the Ss rehearsed the C Pairs more than the E Pairs on the study trial, it does not seem possible that the associative strength of the C Pairs would even approach that of the E Pairs. It therefore becomes apparent that no simple conclusion about the role of associations in recognition is possible. This is not to say that an interpretation of the findings of Experiment 6 as seen in Fig. 6 is not possible. A tentative interpretation will be given later. Before doing this, however, it will be useful to look at the results of some additional experiments.

Experiment 7

In this experiment Ss were again presented pairs of items from the five classes or types of pairs used in Experiment 6. On the test, only one word from each pair was tested, along with single words (control or new words) which had not been presented. The purpose of this procedure was not to test the effects of removal of associative context. Rather, the interest was centered in the relative recognition performance under these "impoverished" conditions for the five different classes. The interest was in particular directed toward the homonyms. It had seemed clear from Fig. 6 that a lenient criterion was involved in the judgments on the homonym pairs. This alone could result in few misses on the E Pairs. But still another factor would also keep the number of misses to a minimum. As discussed earlier, a homonym pair should double the frequency of the acoustic-articulatory response as compared with that for any single word in the pairs of the other classes. Insofar as frequency information enters into the recognition decisions for pairs, the misses should be less for the homonym pairs than for other types. Now, when a single word from a homonym pair (as presented for study) is presented for a YES-NO test, the acoustic-articulatory frequency should be the major basis for the decision. In fact, it should be of little consequence whether the pair or the single word is presented if frequency is the dominant attribute. Number of misses should be small in both cases. Consider next the case of a single new word being tested, a new word which is a nominal member of a homonym pair but a pair which was not presented for study. There seems to be no reason to expect that the S would perceive this word as having a homonym. Therefore, there should be no criterion problem that is specific to homonyms. Or, in terms of formal predictions for the experiment, new items (from nominal homonym pairs) should not produce more false alarms than single new words representing the other classes. And, as discussed above, misses should be fewer on the old words from homonyms pairs than on the old words from the other pairs.

Method

Study lists and test booklets. The S was presented 50 pairs for study, 10 pairs from each of the five classes. There were also two filler pairs at the beginning and end. On the test, one word from each of the 50 pairs was printed in a booklet along with 50 new words, representing one word from each of the 50 pairs remaining in the five classes. For the 50 pairs presented for study, half the time (within each class) the first word in the pair was tested and half the time the second word in the pair was tested. By using two study lists, and a single test form, each word served once as an old word and once as a new word. The study list was block randomized so that each type of pair occurred once in each block of five trials. On the test form (two pages), the 10 item types (five old and five new from each class) were block randomized and were random with respect to input order.

Procedure and Ss. Prior to presenting the list of pairs for study, the S were informed that the study list consisted of a series of pairs, and further, that on the test there would be single words, some from the pairs presented and some from pairs not presented. They would be required to make a YES-NO decision on each word. The list was presented by a slide projector at a 5-sec. rate. Immediately following the presentation of the last slide, the booklets were distributed, and the instructions for the test repeated. As was true in all of the previous studies, the test was unpaced and the S was required to complete all items on the first page before going to the second. Thirty-four Ss completed each form for a total of 68.

Results

Figure 7 shows the basic results, plotted in the same manner as Fig. 6. Two facts are to be noted. First, errors on the antonyms (both misses and false alarms) were considerably higher than for the other types. Second, the fewest misses were made on the homonym pairs. The statistical analysis shows that class or type of item, based on M+FA is reliable, $F(4, 268) = 18.72$, $p < .01$. More errors were made on old words (f=1) than on new words (f=0), $F(1, 67) = 6.18$, $p < .05$, and the interaction between frequency and type was also reliable, $F(4, 268) = 6,52$, $p < .01$. Sources of these effects are quite apparent in Fig. 7.

The number of misses on the homonyms was slightly under 14%. The class of items having the next fewest misses was conceptual associates 19%. A direct-difference t showed this difference to be reliable, $t(67) = 2.84$, $p < .01$. This may be interpreted to mean that input frequency was greater on the homonym pairs than on the other types. The fact that there were only slightly more false alarms ("yes" to new items) for the homonyms than for the corresponding items in three of the other classes is interpreted to mean that a criterion difference for the classes was of little consequence in this study.

Recognition of antonyms was clearly inferior to the recognition of the words in the other classes. There was a suggestion of this possibility in the previous experiment for the E Pairs (see Fig. 6). A part of the difference seen in Fig. 7 can be attributed to heavy testing effects for the antonyms. This is to say, performance decreased in accuracy from page 1 of the test booklet through page 2. This was quantified for each class of items by using M+FA, with 10 items on each page. The errors increased for the parallel associates (15.3% to 19.3%), for the synonyms (16.9% to 22.2%), and for the antonyms (20.7% to 31.5%). All three of these changes were reliable beyond the .05 confidence level. The homonyms showed an unreliable increase (14.9% to 16.5%) and the conceptual associates showed a slight decrease (17.6% to 16.2%). The reasons for these differential testing effects in this experiment are not clear. It should be noted, however, that as pointed out earlier, there were no systematic changes in test performance over pages for Experiment 6 where pairs were tested.

Experiment 8

This experiment will be briefly reported. It had as its major purpose the examination of the possibility that some peculiarity of the words used in the homonym pairs led to the small number of misses in Experiments 6 and 7. It is, perhaps, becoming apparent that some theoretical importance is to be attached to the findings reported thus far for the homonyms. The test made in the present experiment was simply to present single words from the pairs in each class for study and then test these single words for recognition along with new words from the classes. The critical question is whether or not the outcome will be the same for single words from the homonym pairs as for the single words from the other classes, particularly the classes other than antonyms. The antonyms appear to represent a special problem.

Method

The same test forms were used as in Experiment 7. The study list consisted of 50 single words (plus two filler words at the beginning and end of the series), 10 in each class, representing 10 of the pairs. These were in fact the same words as tested in Experiment 7. By using two different presentation lists, each of 20 words in each class served both as new words and as old words.

The Ss were fully instructed concerning the study list and the method of testing. There were 24 completed records for each of the two forms for a total of 48. All procedures were exactly the same as in the previous studies.

Results

The essential results are shown in Fig. 8, which is plotted in the same manner as the other recent graphs. It is to be noted that the results for the single words from the homonym pairs fall well within the limits set by the parallel associates, conceptual associates, and synonyms. This is to conclude, therefore, that there seems to be nothing peculiar about the words in the homonym pairs when recognition performance is observed for them as single words.

Figure 8 shows that the single words from antonym pairs produced many false alarms, as well as more misses than did the words from the other classes. Performance on the antonyms was largely responsible for the reliable class effect, $F(4, 188) = 8.46$, $p < .01$. There is a suggestion that a criterion difference might exist for the antonyms versus the other classes, but the interaction between old-new and type was less than one. It should be clear that the new items in the test were not, for any of the classes, the associated word as shown in Table 2. The false alarms cannot, therefore, be attributed to implicit associative responses elicited during the study trial which were represented in the new words on the test. If the false alarms are to be attributed to such implicit responses they arise from sources which cannot be easily identified. That some of the false alarms could be attributed to this mechanism is not to be denied, but it is not apparent why these should fall more heavily on the antonyms than on the other classes. The antonym problem will not be pursued further in this paper.

General Discussion

This research was initiated to determine the role which an association between a pair of words played in recognition. Given the knowledge of this role (if such existed) it would be possible to assess the degree to which frequency theory fails to accomodate it. The major purpose of this discussion is to make this assessment. The basic evidence from the experiments will be summarized initially.

1. Experiments 1 and 2 showed that parallel procedures in which recognition decisions were required in one case and frequency judgments in the other produced the same outcomes quantitatively. This was true for both absolute and relative judgments. Errors tended to fall on the same pairs for both types of judgments, and recognition errors were predictable from frequency judgments by combining the means and variance of the frequency judgments.

2. The paired-associate study showed that pairs on which few recognition errors were made were more easily associated than were pairs on which many errors were made. This would seem to implicate associability as a factor in recognition performance over and above frequency. Experiments 3 and 4 supported this conclusion in showing that a pair of words, both shown on the study list but in different pairs, were less likely to be recognized than were intact pairs.

3. Experiment 6, however, in which pairs of culturally associated words were used, provided no consistent evidence that recognition performance was better for associated pairs than for nonassociated pairs. If one assumes that the association in culturally associated words is

not qualitatively different from the associations established in the laboratory between initially neutral words, the findings with the culturally associated words are in direct confict with those summarized in point 2 above.

4. Quite unexpectedly, criterion differences were found between frequency judgments and recognition decisions, the criterion being more lenient for the former. Recognition performance on homonym pairs also reflected a much more lenient criterion than was found for control pairs and culturally associated pairs.

The results of Experiments 1 and 2 make it difficult to turn abruptly away from frequency as the basic discriminative attribute for pair recognition. Therefore, the attempt has been made to develop a coherent account of the basic findings, using an expanded theory of frequency discrimination. This expansion is post hoc, and consequently must be viewed with skepticism, although the formulation as a whole does rest on frequency theory as developed in recent years. Before describing this account, it seems necessary to show how associative information or associative strength between words in a pair cannot be taken seriously as a direct explanatory mechanism for pair recognition.

Experiment 6 showed that associated pairs and nonassociated pairs produced equivalent recognition performance. It would be quite possible to obtain such a result even if the recognition performance in the two cases was mediated by quite different types of information in memory. To be more specific, for the associated pairs the decisions may have been based exclusively on the associative relationship, while the decisions for the nonassociated pairs may have been based on frequency information. Two considerations argue against the idea that the associative information is critical in the recognition of the culturally-associated pairs of words. First, the new pairs for the associated pairs were also highly associated pairs. Therefore, some information other than associative information would be required to distinguish between the old and new pairs. Otherwise, the Ss should have accepted any associated pair as having been included in the study list. Still, it might be argued (and this leads to the second point) that the single presentation of the associated pairs so increased the momentary situational strength of the association for the associated pairs that they were distinguishable on this basis from the control pairs even though the latter were strongly associated by cultural usage. This does not seem to be a likely possibility. In Experiment 2, 12 pairs made up of unrelated words were presented twice during the study trial, and 12 pairs were presented three times. On the YES-NO recognition test there were 5.3% misses for the former pairs, 3.5% for the latter. Although comparing error rates from experiment to experiment is risky because of many possible differences in subject samples, number of words presented and tested, and so on, these values must be considered appreciably less than the misses for the culturally associated pairs presented once. It does not seem appropriate to conclude that a neutral pairs of words, presented twice, has a stronger associative link between them than does a pair of culturally associated words, such as table-chair, presented once. Based on associative strength alone, there is no reason to believe that the Ss should have made any errors on the culturally-associated pairs of words. Associative strength, at least as the term is understood as a consequence of multiple trials in the laboratory, does not seem to be directly involved at all in the recognition decisions. This conclusion was important in directing the explanatory efforts back toward frequency as the fundamental discriminative attribute for pair recognition.

In a recent study (Underwood & Zimmerman, 1973), the evidence pointed toward the fact that each syllable of a two-syllable word gained a small amount of subjective frequency which was independent of the frequency information for the word per se. This is to say that the subunits of larger units may have frequency representation in memory. This abstractive nature of the assimilation of event frequency has long been known to be a characteristic of memory in the developmental history of the individual (Underwood, 1971). It now seems evident that this same feature obtains to some degree in the laboratory. The implication of this fact for frequency theory as applied to recognition of pairs is that the frequency information in memory will consist of information about each word, and also about the two words as a unit.

Given that frequency assimilation for a pair may involve three event frequencies, it is reasonable to ask what the optimal frequency should be for each event if recognition errors are to be minimized following a single presentation of a pair. Obviously, this arrangement would be one in which each unit has a subjective frequency of one, and the pair as a unit has a subjective frequency of one. All three sources of information would lead the S to respond with YES. A reduction in the subjective frequency for any one of the three events should reduce the likelihood of a correct decision. Assumptions will now be made to reflect changes in the subjective frequency of each of the three events, hence changes in the likelihood of a correct decision.

It will be assumed that the stronger the association between two words prior to the study trial, the greater the likelihood that the subjective frequency of the pair as a unit will approach one, and the less the likelihood that the subject frequency of each word will approach one. The converse may be stated. If there is no associative relationship existing between two words prior to the study trial, the greater is the likelihood that the subjective frequency for the individual words will approach one, and the less the likelihood that the subjective frequency of the pair as a unit will approach one. In the extreme case, if no associative link is developed between the words in a pair during the study trial, the subjective frequency of the pair as a unit would be zero. The application of these notions to the data may now be examined.

1. In Experiment 3, re-pairing old words on the test resulted in an increase in the errors. This would be accounted for on the grounds that a re-paired item has no subjective frequency for the pair as a pair. The decision must be made on the basis of the subjective frequency

for the individual words. In Experiment 4, errors on the recognition of a single word from a pair presented for study increased (as compared with the presence of the intact pair), and the amount of increase was independent of the newness or oldness of the context word. The lack of pair-frequency information is believed responsible. In a further experiment, not reported in this series, individual words were presented for study. On the test, pairs of individual words were presented just as in Experiment 4 and with S always making a recognition decision on the second word. Just as found by Thomson (1972), the characteristics (old or new) of the first word in the pair had no effect on the recognition of the second. Since there was no pair-frequency information induced during study, no loss of accuracy should have been found by prefixing either old or new words to the test word.

2. Pairs on which few recognition errors were made were more easily associated (in a paired-associate test) than were pairs on which many errors were made. The easy pairs were also judged to have higher frequency than were the difficult pairs. The easy pairs represent pairs which approximate as closely as it is probably possible the maximum-frequency case. The pair was not associated prior to the study trial. But, for whatever reason, an association could be established readily. In this process, each word will receive a near maximum frequency (one), and in addition, the pair will approach this maximum because of the rapid formation of the association.

3. Why were not culturally associated words better recognized than non-associated words? It is now presumed that this result is a coincidence of the distribution of subjective frequency information. For

the associated pairs, the decisions are primarily based on pair frequency, for the nonassociated pairs, on element or single-word frequency. It would be suspected that the small differences in errors among the four classes of associates resulted from differences in the magnitude of the subjective frequency for the single words.

It should be repeated that the formulation is post hoc. However, it does have a number of testable implications. Although these will not be examined here, it can be seen that the basic idea would be to evolve procedures which would change the subjective frequency of one or more of the sources assumed by the theory to be involved in the recognition decisions.

The theory does not seem capable of handling cleanly the false alarm data for N-O and O-N pairs in Experiments 3 and 4. All that can be said is that since subjective frequency should be present for one of the three events held important by the theory, the S is led to a YES response. This is not considered a satisfactory account for such false alarms.

Finally, the result for the homonym pairs provided a special case for frequency theory, a case that touches on several issues. Frequency theory makes no assumption concerning the awareness of the S that he is using frequency information in recognition decisions. In response to questioning, Ss will sometimes respond that the correct item looked familiar, but more frequently given is the response: "it looked right." It will be remembered that the only distinctive difference found between recognition decisions and frequency judgments in Experiments 1 and 2 was that the Ss set a more lenient criterion in making frequency judgments than in making recognition decisions. Now, when the S is presented homonym pairs he may, with some level of intention, make his decisions on the basis of the frequency of the acoustic-articulatory responses he produced to a pair during the study trial. If this is the case, then whatever leads the S to set a lenient criterion for frequency judgments may likewise lead him to do the same in recognition of pairs of homonyms. That frequency of the acoustic-articulatory response is implicated seems clear by the results of Experiment 7. This study showed that relatively few misses were made on a single word from a homonym pair, a result that was not due to a lenient criterion. The S, in effect, had a subjective frequency of two for the single acoustic-articulatory response; for the other classes of items, the frequency would be appreciably less.

This series of experiments was initiated with the expectation that associative information in memory would cause a serious breakdown in the usefulness of frequency information as an explanatory mechanism. The evidence as viewed has led to the conclusion that it would be premature to abandon frequency theory in attempts to account for recognition performance for associatively related verbal units.

References

Broder, P. Multiple-choice recognition as a function of the relative frequencies of correct and incorrect alternatives. Unpublished doctoral dissertation, Northwestern University, 1973.

Cofer, C. N., Segal, E., Stein, J., & Walker, H. Studies on free recall of nouns following presentation under adjectival modification. Journal of Experimental Psychology, 1969, 79, 254-264.

DaPolito, F., Barker, D., & Wiant, J. The effects of contextual changes on component recognition. American Journal of Psychology, 1972, 85, 431-440.

Ellis, H. C., & Shumate, E. C. Encoding effects of response belongingness and stimulus meaningfulness on recognition memory of trigram stimuli. Journal of Experimental Psychology, 1973, 98, 70-78.

Light, L. L. Homonyms and synonyms as retrieval cues. Journal of Experimental Psychology, 1972, 96, 255-262.

Light, L. L., & Carter-Sobell, L. Effects of changed semantic ontext on recognition memory. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 1-11.

Thomson, D. M. Context effects in recognition memory. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 497-511.

Thorndike, E. L., & Lorge, I. The teacher's word book of 30,000 words. New York: Columbia University Press, 1944.

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. Attributes of memory. Psychological Review, 1969, 76, 559-573.

Underwood, B. J. Recognition memory. In H. H. Kendler and J. T. Spence (Eds.), Essays in neobehaviorism. New York: Appleton-Century-Crofts, 1971.

Underwood, B. J. Word recognition memory and frequency information. Journal of Experimental Psychology, 1972, 94, 276-283.

Underwood, B. J., & Freund, J. S. Testing effects in the recognition of words. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 117-125.

Underwood, B. J., & Zimmerman, J. The syllable as a source of error in multisyllable word recognition. Journal of Verbal Learning and Verbal Behavior, 1973, in press.

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.

Wolford, G. Function of distinct associations for paired-associate performance. Psychological Review, 1971, 78, 303-313.

Wood, G. Associations and recognition learning. Psychonomic Science, 1969, 17, 222-223.

Table 1

Easy and Hard Lists used in the Paired-Associate Experiment with Other Relevant Information about the Pairs

	Recognition Errors		Frequency Judgments		Correct in
Easy List	Exp. 1	Exp. 2	M	M/σ	PA
nymph-prune	1	4	.95	1.53	67
bacon-shrub	3	6	1.11	1.28	58
mound-alert	3	15	1.24	1.29	65
onion-ivory	4	4	.93	1.37	75
daddy-waver	5	3	.86	1.30	77
broom-folly	5	5	1.26	1.34	61
laden-cargo	5	8	1.09	1.33	55
drank-voter	6	6	.99	1.43	76
exalt-baron	6	3	1.05	1.52	62
wedge-canon	7	6	.97	1.26	53
Mean	4.5	6.0	1.05	1.37	64.9
Hard List					
hatch-focus	25	21	.80	.84	45
marsh-brute	20	13	.76	.96	23
award-birch	17	6	.94	1.07	54
dwelt-forge	17	4	.99	1.27	19
harsh-inner	16	17	.81	1.01	44
lance-chime	16	21	.73	.97	44
risen-stray	16	18	.86	.97	43
punch-exile	16	8	1.08	1.33	22

(continued)

Table 1 continued

thief-drill	15	10	.78	.96	66
flush-abide	15	11	.89	1.16	22
Mean	17.3	12.9	.86	1.05	39.2

Table 2

Critical Pairs of Words used in Experiments 6, 7, and 8

Antonyms	Conceptual Associates	Parallel Associates	Synonyms	Homonyms
day-night	canary-bird	cup-saucer	complete-entire	see-sea
bad-good	horse-animal	bed-sleep	spoken-verbal	pail-pale
lost-found	minnow-fish	bread-butter	hidden-concealed	plane-plain
hard-soft	maple-tree	income-money	empty-vacant	minor-miner
dirty-clean	python-snake	lamp-light	bungalow-house	peek-peak
boy-girl	waltz-dance	hammer-nail	capsule-pill	cent-sent
buy-sell	pansy-flower	needle-thread	chill-cold	weak-week
true-false	red-color	salt-pepper	correct-right	fare-fair
give-take	murder-crime	eight-nine	wicked-evil	sleigh-slay
high-low	water-liquid	scissors-cut	signature-name	tea-tee
king-queen	apple-fruit	spider-web	starved-hungry	stayed-staid
long-short	cancer-disease	table-chair	silent-quiet	sale-sail
slow-fast	bracelet-jewelery	ale-beer	grief-sorrow	wholly-holy
bottom-top	silk-cloth	army-navy	careful-cautious	mane-main
far-near	pliers-tool	candy-sweet	mad-angry	meat-meet
hate-love	emerald-gem	dock-boat	tiny-small	need-kneed
open-close	gnat-insect	hand-foot	central-middle	lode-load
rich-poor	east-direction	nurse-doctor	ancient-old	course-coarse
smooth-rough	uncle-relative	mail-letter	rural-country	prey-pray
under-over	oil-fuel	lock-key	boulder-rock	seen-scene

Table 3

Mean Errors (M+FA) for E Pairs and C Pairs in each of the Five Classes#

	Parallel Associates	Synonyms	Antonyms	Homonyms	Conceptual Associates
E Pairs	1.14	1.15	1.70	1.42	.95
C Pairs	1.32	1.32	1.23	1.10	.83

The standard deviations for these 10 disbributions varied between .94 and 1.39

Footnotes

[1]This research was supported by the Personnel and Training Research Programs, Psychological Sciences Division, Office of Naval Research, under Contract N00014-67-A-0356-0010, Contract Authority Identification No. NR 154-321.

[2]Requests for reprints should be sent to Benton J. Underwood, Department of Psychology, Northwestern University, Evanston, Illinois 60201.

[3]This pool of words was formed by Dr. Carl P. Duncan. Its use is greatly appreciated.

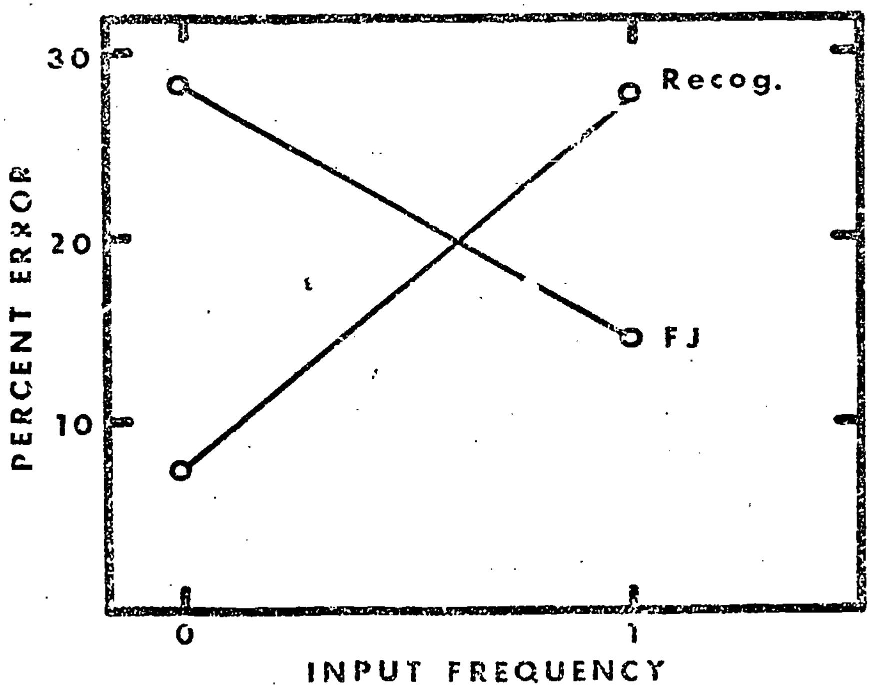

Fig. 1. Errors on pairs of words as related to input frequency. The test included both frequency judgments (FJ) and recognition (Recog).

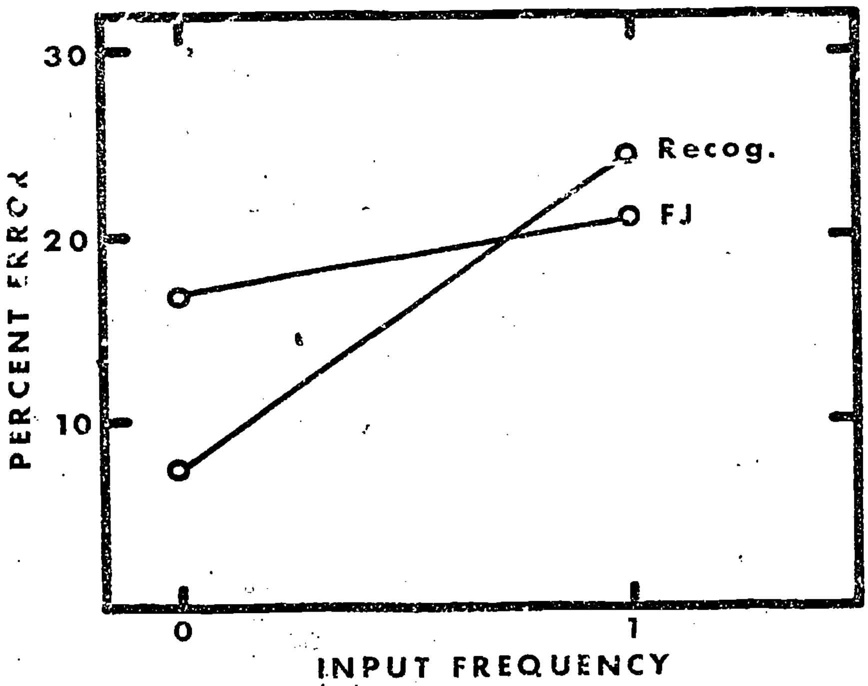

Fig. 2. Further evidence on the errors made on pairs of words as related to input frequency for frequency judgments and recognition.

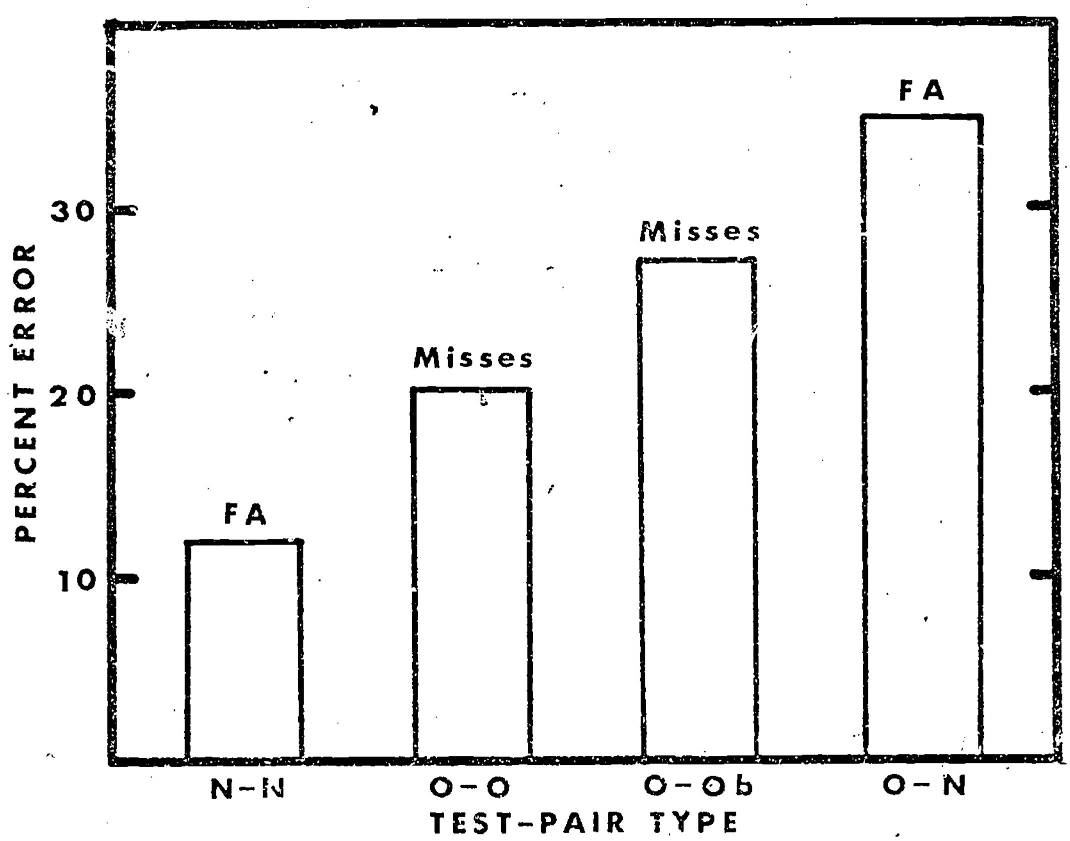

Fig. 3. Misses and false alarms (FA) for four types of test pairs. N refers to a new word, O to an old word, and b to a broken pair.

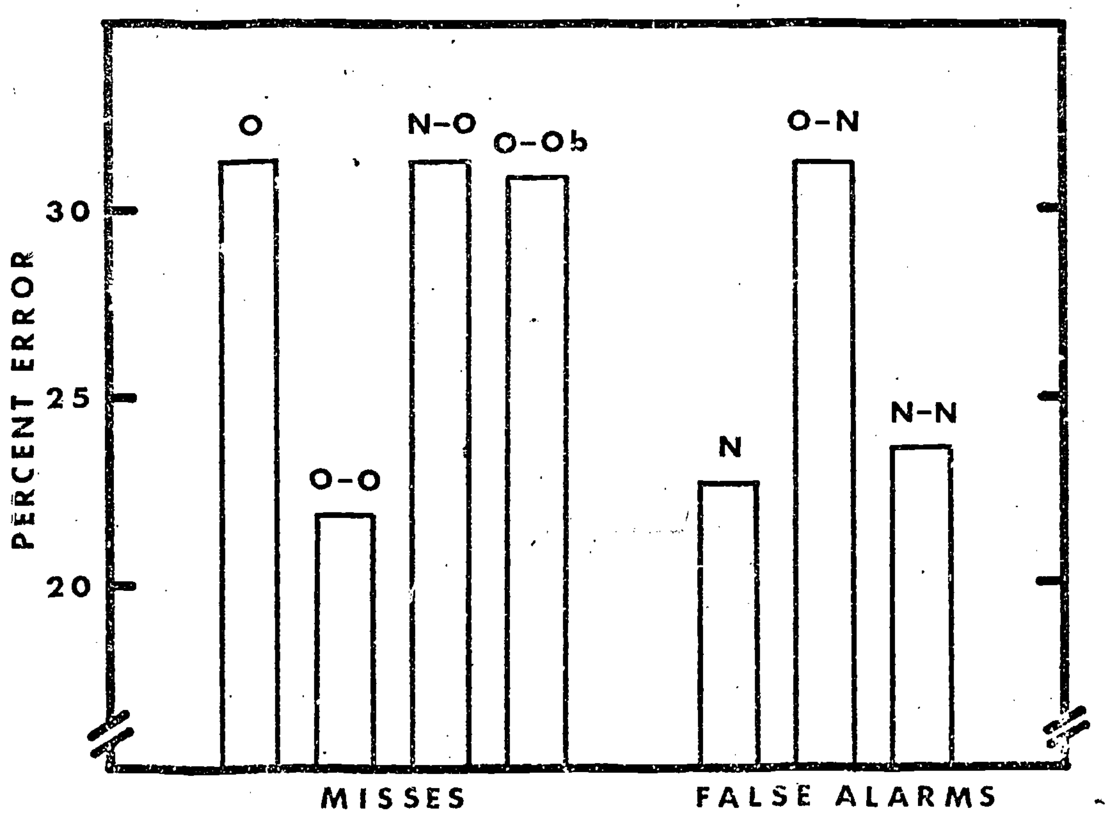

Fig. 4. Misses and false alarms for various types of test items. For pairs, the recognition decision was made only on the second word in the pair.

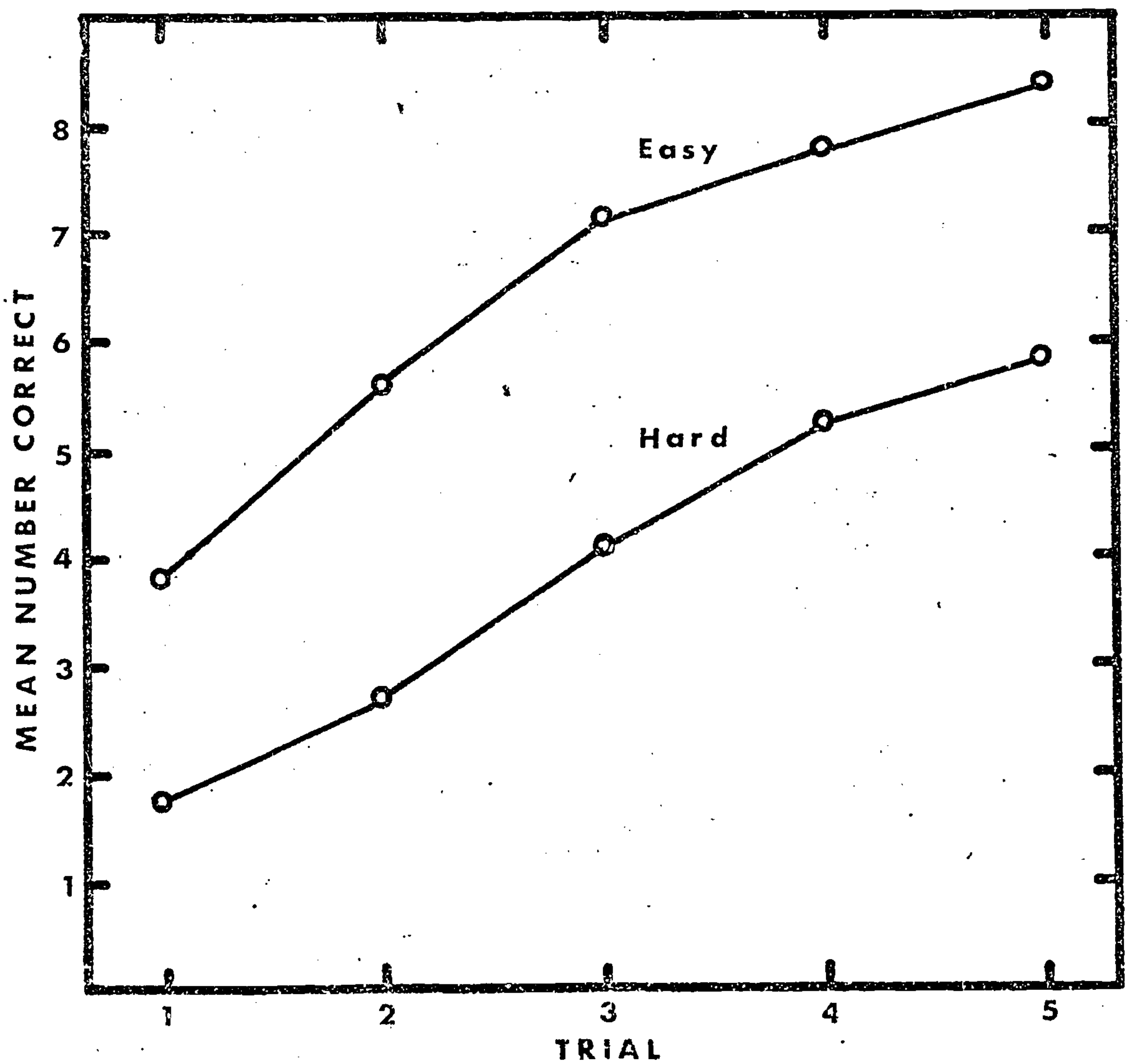

Fig. 5. Acquisition of paired-associate lists of 10 pairs each, identified as easy and hard by the number of recognition errors made on the pairs.

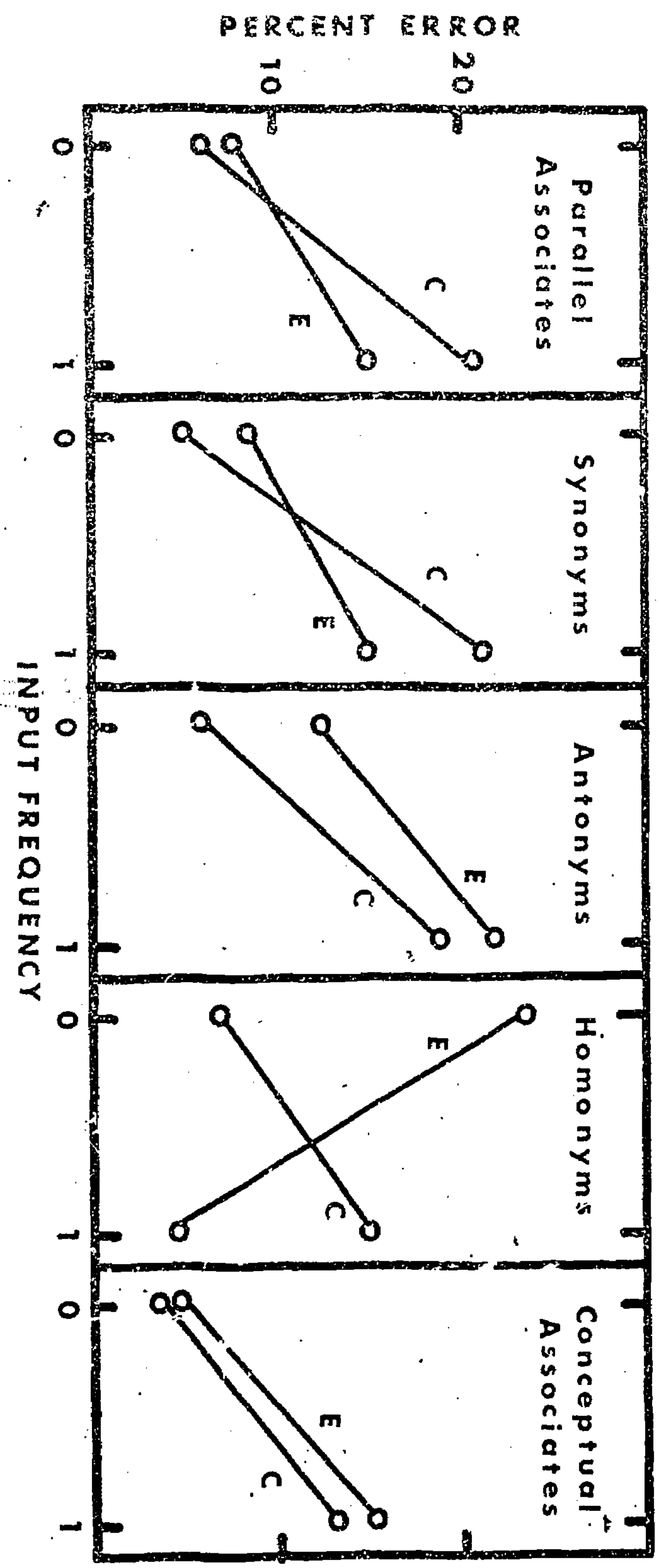

Fig. 6. Recognition errors on experimental (E) and control (C) pairs of words as a function of input frequency and class or type of pair.

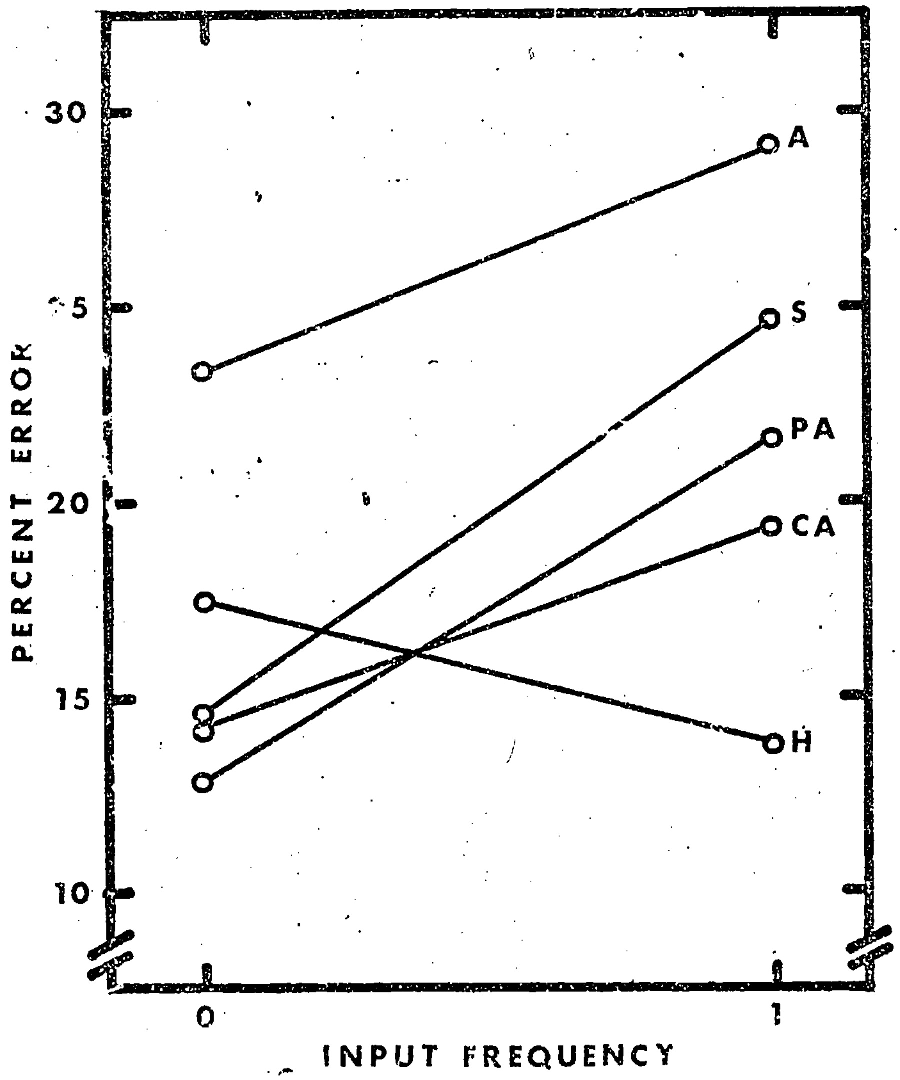

Fig. 7. Recognition errors on single words following pair presentation for antonyms (A), synonyms (S), parallel associates (PA), conceptual associates (CA), and homonyms (H).

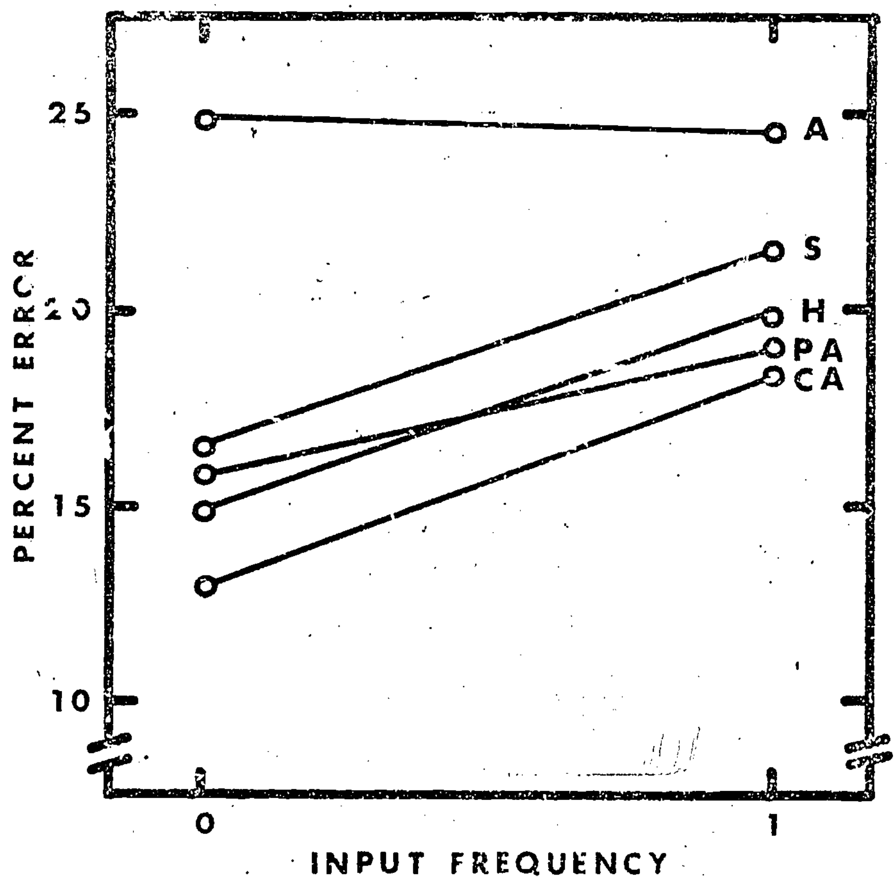

Fig. 8. Recognition errors on single words following presentation of single words from five classes of pairs.

DISTRIBUTION LIST

Navy

4 Dr. Marshall J. Farr, Director
Personnel & Training Research Programs
Office of Naval Research
Arlington, VA 22217

1 Director
ONR Branch Office
495 Summer Street
Boston, MA 02210
ATTN: C.M. Harsh

1 Director
ONR Branch Office
1030 East Green Street
Pasadena, CA 91101
ATTN: E.E. Gloye

1 Director
ONR Branch Office
536 South Clark St.
Chicago, IL 60605
ATTN: M.A. Bertin

6 Director
Naval Research Laboratory
Code 2627
Washington, DC 20390

12 Defense Documentation Center
Cameron Station, Bldg. 5
5010 Duke St.
Alexandria, VA 22314

1 Chief of Naval Technical Training
Naval Air Station Memphis (75)
Millington, TN 38054
ATTN: Dr. G.D. Mayo

1 LCDR Charles J. Theisen, Jr., MSC, USN
4024
Naval Air Development Center
Warminster, PA 18974

1 Mr. Lee Miller (AIR 413E)
Naval Air Systems Command
5600 Columbia Pike
Falls Church, VA 22042

1 Dr. Harold Booher
NAVAIR 415C
Naval Air Systems Command
5600 Columbia Pike
Falls Church, VA 22042

1 Commanding Officer
Naval Medical Meuropsychiatric Research
San Diego, CA 92152

1 Technical Reference Library
Naval Medical Research Institute
National Naval Medical Center
Bethesda, MD 20014

1 Dr. John J. Collins
Chief of Naval Operations (OP-987F)
Dept. of the Navy
Washington, DC 20350

1 Technical Library
Bureau of Naval Personnel
Dept. of the Navy
Washington, DC 20360

1 Dr. James J. Regan, Technical Director
Navy Personnel Research & Development Center
San Diego, CA 92152

1 Special Asst. to Chief of Naval Personnel
Bureau of Naval Personnel (Pers Oe)
Arlington Annex, Room 2403
Washington, DC 20370

1 Commanding Officer
Navy Personnel Research & Development Center
San Diego, CA 92152

1 Superintendent
Naval Postgraduate School
Monterey, CA 92940
ATTN: Library (Code 2124)

1 Technical Library
Naval Ship Systems Command
National Center, Bldg. 3
Room 3S08
Washington, DC 20360

1 Chief of Naval Training Support
Code N-21
Bldg. 45
Naval Air Station
Pensacola, FL 32508

1 Dr. William L. Maloy
Principal Civilian Advisor for
Education & Training
Naval Training Command, Code 01A
Pensacola, FL 32508

1 Mr. Arnold Rubinstein
Naval Material Command (NMAT-03424)
Room 820, Crystal Plaza #6
Washington, DC 20360

1 Dr. Martin F. Wiskoff
Navy Personnel Research & Development
Center
San Diego, CA 92152

1 Dr. John Ford, Jr.
Navy Personnel Research & Development
Center
San Diego, CA 92152

1 Director of Research
U.S.Army Armor Human Research Unit
ATTN: Library
Bldg. 2422 Morade St.
Fort Knox, KY 40121

1 U.S. Army Research Institute for the
Behavioral & Social Sciences
1300 Wilson Blvd.
Arlington, VA 22209

1 Mr. Edmund F. Fuchs
U.S. Army Research Institute
1300 Wilson Blvd.
Arlington, VA 22209

1 Chief, Unit Training & Educational
Technology Systems
U.S. Army Research Institute for the
Behavioral & Social Sciences
1300 Wilson Blvd.
Arlington, VA 22209

1 Dr. Stanley L. Cohen
Work Unit Area Leader
Organizational Development Work Unit
Army Research Inst. for Behavioral &
Social Science
1300 Wilson Blvd.
Arlington, VA 22209

1 Dr. Leon H. Nawrocki
U.S. Army Research Inst.
Rosslyn Commonwealth Bldg.
1300 Wilson Blvd.
Arlington, VA 22209

Air Force

1 Dr. Martin Rockway
Technical Training Division
Lowry Air Force Base
Denver, CO 80230

1 Maj. P.J. DeLeo
Instructional Technology Branch
AF Human Resources Laboratory
Lowry AFB, CO 80230

1 Headquarters, U.S. Air Force
Chief, Personnel Research & Analysis
Division (AF/DPSY)
Washington, DC 20330

1 AFHRL/AS (Dr. G.A. Eckstrand)
Wright-Patterson AFB
Ohio 45433

1 AFHRL (AST/Dr. Ross L. Morgan)
Wright Patterson Air Force Base
Ohio 45433

1 AFHRL/MD
701 Prince St. Room 200
Alexandria, VA 22314

1 AFOSR(NL)
1400 Wilson Blvd.
Arlington, VA 22209

1 Headquarters Electronic Systems Division
ATTN: Dr. Sylvia R. Mayer/MCIT
L.G. Hanscom Field
Bedford, MA 01730

Marine Corps

1 Mr. E. A. Dover
Manpower Measurement Unit (Code A01M-2)
Arlington Annex, Room 2413
Arlington, VA 20370

Coast Guard

1 Mr. Joseph J. Cowan, Chief
Psychological Research Branch (P-1)
U.S. Coast Guard Headquarters
400 Seventh St., SW
Washington, DC 20590

Other DOD

1 Lt. Col. Austin W. Kibler, Director
Human Resources Research Office
Advanced Research Projects Agency
1400 Wilson Blvd.
Arlington, VA 22209

1 Dr. Ralph R. Canter
Director for Manpower Research
Office of Secretary of Defense
The Pentagon, Room 3C980
Washington, DC 20301

Miscellaneous

1 Dr. Scarvia Anderson
Executive Director for Special Devpt.
Educational Testing Service
Princeton, NJ 08540

1 Dr. Richard C. Atkinson
Stanford University
Dept. of Psychology
Stanford, CA 94305

1 Dr. Ronald P. Carver
American Institutes for Research
8555 Sixteenth St.
Silver Spring, MD 20910

1 Century Research Corporation
4113 Lee Highway
Arlington, VA 22207

1 Dr. Kenneth E. Clark
University of Rochester
College of Arts and Sciences
River Campus Station
Rochester, NY 14627

1 Dr. Allan M. Collins
Bolt Beranek & Newman
50 Moulton St.
Cambridge, MA 02138

2 ERIC
Processing & Reference Facility
4833 Rugby Ave.
Bethesda, MD 20014

1 Dr. Victor Fields
Dept. of Psychology
Montgomery College
Rockville, MD 20850

1 Dr. Edwin A. Fleishman
American Institutes for Research
8555 Sixteenth St.
Silver Spring, MD 20910

1 Dr. Robert Glaser, Director
University of Pittsburgh
Learning Research & Development Center
Pittsburgh, PA 15213

1 Dr. Robert Tenneyson
Florida State University
Center for Computer-Assisted Instruction
Tallahassee, FL 32306

1 Dr. Henry J. Hamburger
University of California
School of Social Sciences
Irvine, CA 92664

1 Human Resources Research Organization
Division #3
P.O. Box 5787
Presidio of Monterey, CA 93940

1 Human Resources Research Organization
Division #4, Infantry
P.O. Box 2086
Fort Benning, GA 31905

1 Human Resources Research Organization
Division #5, Air Defense
P.O. Box 6057
Fort Bliss, TX 79916

1 Human Resources Research Organization
Division #6, Library
P.O. Box 428
Fort Rucker, AL 36360

1 Dr. Lawrence B. Johnson
Lawrence Johnson & Associates, Inc.
200 S Street, N.W., Suite 502
Washington, DC 20009

1 Dr. David Klahr
Carnegie-Mellon University
Graduate School of Industrial Admin.
Pittsburgh, PA 15213

1 Dr. Robert R. Mackie
Human Factors Research, Inc.
6780 Cortona Dr.
Santa Barbara Research Park
Goleta, CA 93017

1 Dr. Andrew R. Molnar
Technological Innovations in Education
National Science Foundation
Washington, DC 20550

1 Dr. Leo Munday, Vice President
American College Testing Program
P.O. Box 168
Iowa City, IA 52240

1 Dr. Donald A. Norman
University of California, San Diego
Center for Human Information Processing
La Jolla, CA 92037

1 Mr. Luigi Petrullo
2431 N. Edgewood St.
Arlington, VA 22207

1 Dr. Diane M. Ramsey-Klee
R-K Research & System Design
3947 Ridgemont Dr.
Malibu, CA 90265

1 Dr. Joseph W. Rigney
Behavioral Technology Laboratories
University of Southern California
3717 South Grand
Los Angeles, CA 90007

1 Dr. George E. Rowland
Rowland & Company, Inc.
P.O. Box 61
Haddonfield, NJ 08033

1 Dr. David J. Weiss
Dept. of Psychology
University of Minnesota
Minneapolis, MN 55455

1 Dr. Anita West
Denver Research Institute
University of Denver
Denver, CO 80210

1 Dr. Kenneth Wexler
University of California
School of Social Sciences
Irvine, CA 92664

1 Dr. John Annett
The Open University
Milton Keynes
Buckinghamshire ENGLAND

1 Dr. Charles A. Ullmann
Director, Behavioral Sciences Studies
Information Concepts Incorporated
1701 No. Ft. Myer Dr.
Arlington, VA 22209

1 Mr. Dexter Fletcher
Dept. of Psychology
Univ. of Illinois at Chicago Circle
Box 4348
Chicago, IL 60680