

Paired associate learning: Data on pair difficulty and variables that influence difficulty

BENTON J. UNDERWOOD
Northwestern University, Evanston, Illinois 60201

Two hundred pairs of five-letter words were produced randomly from all five-letter words in Thorndike and Lorge (1944). The difficulty of each pair was established for paired associate learning by having 50 subjects learn lists of 20 pairs. The difficulty of a pair was found to be highly reliable and was not influenced by the particular list in which it was learned. Frequency of response terms was positively related to learning, but the frequency of the stimulus terms was not. Two-syllable five-letter words were learned more rapidly than one-syllable five-letter words. Two 20-pair lists were constructed, one consisting of homogeneous pairs of average difficulty and the other of 10 very easy pairs and 10 very difficult pairs. As anticipated, performance on the latter list was initially better than that on the former, with the performance on the two lists converging over trials. The difficulty of the pairs as determined by paired associate learning was unrelated to misses on a recognition test, but the false alarms decreased as difficulty decreased.

This paper grew out of a simple research goal. There was a need in our laboratory for a pool of pairs of words with known levels of difficulty as determined by paired associate learning. One of the immediate needs for this pool was to carry out a study on the nature of the learning curve as a function of the composition of the pair difficulty within the lists. As happens frequently, other studies were suggested by the initial two studies, and it was found that in some cases factors could be isolated that were involved in determining pair difficulty. The relationship between recognition memory and pair difficulty was also examined.

In presenting the experiments, three sections will be used. In the first section, the procedures and results for determining pair difficulty are given. In the second section, various analyses of the results on pair difficulty are described, looking toward a determination of factors that influence pair difficulty. In the third section, three experiments are reported that grew out of the work on pair difficulty.

DETERMINATION OF PAIR DIFFICULTY

Method

Materials. Five-letter words were used to construct pairs. Thorndike and Lorge (1944) list 2,596 five-letter words, including those that occur with frequencies of less than one per million. The complete list of 2,596 words was compiled by C. P. Duncan,

This research was supported in part by NIMH Grant MH32362. Thanks are due Jerome Smeesaert for testing the subjects in the initial experiment, to William Mueller for testing the subjects in the other experiments, and to James Hall and Reid Hastie for suggesting certain analyses.

and this list constitutes the population sampled. On a random basis, 400 of the five-letter words were chosen and then randomly paired to produce 200 pairs. The only exclusions made were proper names, contractions, and words with unusual markings (e.g., accent markings).

The 200 pairs were assigned randomly to 10 lists of 20 pairs each. In addition, a practice list was constructed of 40 words drawn beyond the original 400. It might be supposed that the difficulty of a pair could be appreciably influenced by the characteristics of the other 19 pairs in the list. Therefore, the 200 pairs were assigned randomly to lists five different times, and these will be spoken of as representing five different contexts for each pair (A, B, C, D, E).

Subjects and Procedure. Each of 100 subjects learned five of the paired associate lists following the learning of the practice list. There were 10 different lists for each of the five contexts. Fifty of the subjects learned lists designated arbitrarily Lists 1-5, and 50 learned Lists 6-10. Five subjects learned the lists in the order 1 through 5, five in the order 5 through 1. The same was true for Lists 6-10. The purpose, of course, was to let practice effects fall about equally on all pairs. A group procedure was used for learning, with the size of the groups varying from one to six subjects. A block-randomized schedule was used to assign subjects to list groups (1-5 or 6-10) and to context.

A study-test method was used for learning with the pairs presented by a slide projector. The pairs were presented at a 2-sec rate, as determined by a peripheral timer. There were three study-test trials, with the order of the pairs being different for each study trial. For testing, the subjects were given a 19-page booklet consisting of a cover sheet and three test sheets for each of the six lists. For a given test, the sheet contained the 20 stimulus terms with a blank after each; the subjects were allowed 90 sec to fill in the blanks with the correct responses. After each test trial, the page was turned and the booklet was turned over, following which the next study trial was given.

Results

Difficulty values. Each pair was learned by 50 subjects, 10 subjects in each of the five contexts. As a mea-

sure of difficulty, the mean number of correct responses across three trials for 10 subjects was used, the means, of course, being the sum of the number correct across the five forms divided by five. These means varied between 7.8 for the most difficult pair (rouse-creel) and 27.2 for the easiest pair (kappa-fussy). As a measure of variability of pair difficulty, the standard deviation among the scores for the five forms was used. In the appendix, both the difficulty measure and the variability measure are given for each pair, with the order of the listing being from most difficult to least difficult.

It will be remembered that the 200 pairs were derived by drawing words randomly from the pool of all five-letter words and assigning the words randomly to pairs. What would be the nature of the distribution of pair difficulty to be expected? If it is assumed that there are a number of factors among the pairs that would determine difficulty, the random procedure used for deriving the pairs could result in a symmetrical distribution of pair difficulty. The most difficult pairs would be those in which the random assignment resulted in few positive factors and many negative factors, and easy pairs would be those in which there were many positive factors and few negative ones. There should be many pairs in the middle of the distribution in which the random assignment resulted in about an equal number of positive and negative factors.

These expectations seem to be supported, as may be seen in Figure 1, in which the number of pairs is plotted against the difficulty values. In this plot, all pairs for each 1-point interval were combined. Thus, between 7.0 and 7.9 there was one pair, between 8.0 and 8.9 there were four pairs, and so on. The fact that the tails of the distribution are not severely extended could mean that the number of different factors present in producing pair difficulty is quite limited.

Context effects. The five forms were said to represent five different contexts in which each pair was embedded. It is possible to determine whether these different contexts influenced pair difficulty. Differences in performance on a pair in different contexts could be due to two factors, namely, the context and to differences in the learning ability of the 10 subjects assigned

to each form. Nevertheless, by correlational analysis it can be shown that context effects were of little consequence. This analysis follows.

There were 10 subjects assigned to each form. The two subgroups of five subjects assigned to each form differed only in the order in which the lists were learned. The correlation across the 200 pairs for pair difficulty may be determined for the two subgroups of five subjects. Five such correlations could be calculated, one for each form. The values for these five correlations were .46, .40, .47, .30, and .46, with a mean of .42. These values represent the stability of pair difficulty when the pairs were learned in the same context. It is also possible to calculate correlations for the 200 pairs for subgroups of five subjects when the pairs were learned in different contexts. Five such between-context correlations were calculated of the many possible that could be calculated. Each form occurred twice in these correlations, in one case being represented by the five subjects given one of the orders of the lists in learning, and in the other being represented by the five subjects given the other order of learning. The resulting five correlations were .37, .44, .43, .43, and .38, with a mean of .41. Because this mean is essentially the same as the mean of the within-context correlations (.42), it is concluded that there was no measurable context effect in the learning of the pairs.

It is worth noting that in the above correlations pair difficulty was based on only five subjects. This is a clear indication of the reliability of pair difficulty. Of course, as the number of subjects used to determine pair difficulty is increased, the stability of pair difficulty should increase. Five correlations were calculated between forms, with 10 subjects determining pair difficulty for each form. Across the 200 pairs, the correlations varied between .50 and .70. Five correlations were also calculated when 20 subjects (two forms combined) determined pair difficulty, and these five correlations varied between .71 and .83. Further data on the reliability of pair difficulty will emerge later.

A final comment should be made about the variability of pair difficulty as reflected in the standard deviations reported for each pair in the table in the appendix. It might be expected that these standard deviations would be attenuated at the extreme levels of difficulty. There is no clear evidence in support of this. Pairs of average difficulty had about the same average variability as did those pairs at the extremes. The mean standard deviations for successive groups of 20 pairs from hard to easy were as follows: 2.69, 2.58, 3.02, 2.41, 2.63, 2.61, 2.65, 2.72, 2.77, 2.31.

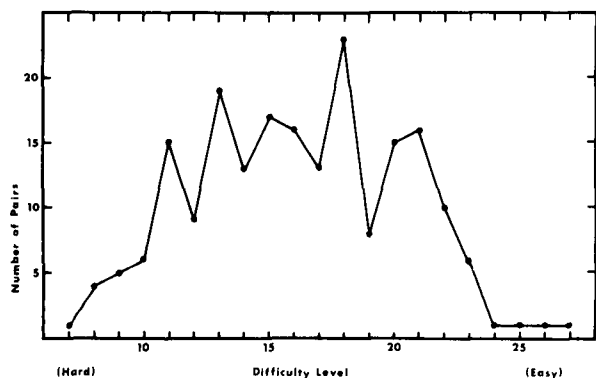


Figure 1. The distribution of pair difficulty.

SOME FACTORS INFLUENCING PAIR DIFFICULTY

Background Frequency

The words used in constructing the pairs varied widely in background frequency as indexed by Thorndike and Lorge (1944). At one extreme, the frequency of

some of the words was less than 1 per million, and at the other, the frequency was over 100 per million (AA words). From previous work (e.g., Underwood & Schulz, 1960), it would be anticipated that the frequency of the response words would be related to the difficulty of the pairs, whereas the frequency of the stimulus words would be of little consequence in determining pair difficulty. The present results are consonant with such expectations. AA words were assigned a value of 100 per million, and A words were assigned a value of 50 per million. Words occurring less frequently than once per million were assigned appropriate decimal values. For example, a word occurring four times per 18 million words was given a value of .22 per million by dividing 4 by 18.

Across the 200 pairs, response-term frequency and ease of learning correlated .31, a highly significant correlation statistically, although not striking in an absolute sense. For stimulus-term frequency, the corresponding correlation was .04. These findings, therefore, are quite in line with results previously reported.

Number of Syllables

In examining the pairs and the learning scores associated with them, it appeared that differences in difficulty might be associated with the number of syllables (one vs. two) of the five-letter words. Therefore, the 200 pairs were divided into four groups: 54 pairs for which both the stimulus term and the response term were of one syllable, 51 pairs for which the stimulus term had two syllables and the response term had one syllable, 38 pairs for which the response term had two syllables and the stimulus term had one syllable, and 59 pairs in which both the stimulus and response terms were of two syllables. The mean total correct responses across the three trials for the four groups were 74.71, 82.57, 81.84, and 91.58, in order. These data would seem to reflect a gross positive correlation between the number of syllables and the ease of learning. The difference between the two extreme groups was reliable ($t = 4.89$). Each of the two middle groups differed (at the .05 level) from the fourth group (two syllables for both terms) but did not differ reliably from the pairs in which neither term had two syllables. The above conclusions are conservative in view of the fact that the two-syllable words had lower background frequency than did the single-syllable words. Insofar as response-term frequency influences learning, the differences are somewhat underestimated for some of the comparisons.

Stimulus-Response Compatibility

Baddeley (1961) reported a study of paired associate learning of nonsense syllables in which the stimulus-response compatibility was varied. The compatibility referred to the relationship between the last letter of the stimulus term and the first letter of the response term. The more frequently this bigram occurs in the language, the greater is the compatibility. Baddeley

showed that ease of learning was directly related to compatibility.

The above analysis was made on the 200 pairs. That is, for each pair, the bigram made up of the last letter of the stimulus term and the first letter of the response term was identified in the Underwood-Schulz (1960) tables, the total count being used. Eight successive groupings of the 200 pairs were formed based on the bigram frequency, the Ns varying from 14 to 39 for the eight distributions. The mean total correct learning scores for the groupings from low bigram frequency to high bigram frequency were 85.6, 79.1, 86.7, 83.5, 83.7, 80.0, 86.1, and 82.4. Clearly, there was no relationship between bigram frequency and learning.

These results should not be considered to be in contradiction to Baddeley's (1961) results as described above. Baddeley used low-meaningfulness nonsense syllables, not words. The associative encoding processes for nonsense syllable pairs may be quite different from those used in encoding word pairs. The present data indicate only that letter-to-letter encoding is not likely with word pairs.

First-Letter Identity

It is not unusual in making up paired associate lists to avoid having identical first letters for the stimulus and response terms in a pair. It is presumed that first-letter identity will result in special forms of encoding, and this may be judged to be undesirable. It does not seem, however, that a test of the effect of first-letter identity has been made for word pairs. Because pairs in the present study were formed randomly, pairs with first-letter identity should not differ on factors from those pairs in which the first letters of the words in a pair were different. It may be asked, therefore, whether first-letter identity influences pair learning.

Of the 200 pairs, 24 had identical first letters. The mean total correct responses for these pairs was 85.54, and the corresponding means for the remaining 176 pairs was 82.20. Although the difference is in the "right" direction, it was not reliable ($t = .66$). Thus, there is no evidence that identical first letters of stimulus and response terms facilitates learning.

FURTHER MANIPULATIONS

In this section, three experiments will be presented. The first deals with the nature of the learning curve when pair difficulty is varied within the list. The second is a direct test of the finding that number of syllables is related to the rate of learning the pairs. The third experiment asks whether recognition memory for pairs is related to the associative learning of the pairs.

Predicting the Shape of the Learning Curves

Suppose a paired associate list is constructed in which half of the pairs are very difficult to learn and the other

half are very easy to learn. Such a list would be a heterogeneous list with respect to pair difficulty (List Het). Suppose further that a second list is constructed in which all of the pairs are about average in difficulty. This would be a homogeneous list (List Hom). What would the shapes of two acquisition curves be expected to be? First, it would seem that on the initial trial or two, performance would be higher on List Het than on List Hom. This would be expected because of the quick learning of the easy pairs in List Het. However, with more trials, the learning curves for the two lists should converge. They should converge because the learning of further pairs in List Het would be slowed because all pairs left would be the very difficult ones. In fact, it might be expected that the two curves would cross, because on later trials the comparison would be between the difficult pairs in List Het and the pairs of average difficulty in List Hom.

There is a factor that may work against the expectations just described. This has to do with strategies the subjects might impose in learning List Het. On the assumption that subjects will recognize two vastly different levels of difficulty of the pairs, some subjects might choose to try to learn the difficult pairs initially. They might implement this objective by displacing rehearsal to difficult pairs when being shown easy ones. If any appreciable number of subjects did this, it would work against the expectations described earlier that initial performance on List Het would be higher than that for List Hom. Nevertheless, it seemed worthwhile to undertake the experiment, checking in the data to determine if the displaced-rehearsal strategy was invoked by any appreciable number of the subjects. It could also be determined whether pair difficulty as determined by the scores on List Het remained the same as the difficulty determined in the original study. High reliability would not be expected if the subjects used a strategy of trying to learn the difficult pairs before learning the easy pairs.

Method

Lists. The construction of the two lists can be easily understood by referring to Figure 1 (the distribution of pair difficulty). List Het was formed of the 10 most difficult pairs and the 10 easiest pairs. List Hom was made up of 20 pairs from the middle of the distribution. These were the 20 pairs numbered 90-109 when the 20 pairs were rank ordered for difficulty, as in the appendix. The difficulty of these 20 pairs varied between 16.0 and 17.0, with a mean of 16.55. The symmetry of the distribution is indicated by the fact that the mean difficulty level of the 20 pairs used in List Het was 16.58.

Procedure and Subjects. The lists were presented for five study and test trials. On the study trials, each pair was presented for 1 sec by a slide projector, a rate that might discourage displaced rehearsal. The test trials were unpaced in the sense that 60 sec were allowed for recall on each trial. A sheet containing the 20 stimulus terms was given to the subjects, with instructions to fill in as many of the correct response terms as possible in the blanks provided after each stimulus term. Five different orders of the pairs were used on the study trials, and the stimulus terms were listed in five different orders on the test sheets.

Subjects were tested in groups of from one to six. A block-randomized schedule of the two lists assigned subjects to lists.

Testing continued until 40 subjects had been placed in each of the two groups. All subjects learned a common practice list before learning one of the two critical lists. The practice list was made up of 10 pairs numbered 50-59 and 10 pairs numbered 150-159; the numbers refer to the positions in the rank ordering of the 200 pairs.

Results

If any substantial number of subjects assigned to List Het tried the strategy of learning the difficulty pairs before the easy pairs, there should be a relatively small difference in learning the easy and difficult pairs. The acquisition curves for these two subgroups of pairs from List Het are plotted in Figure 2. It is apparent that the two classes of pairs differed widely in difficulty, making it extremely unlikely that many subjects set about to learn the difficult pairs before learning the easy pairs. Every one of the 40 subjects had more correct responses on the easy pairs than on the difficult ones.

The acquisition curves for the two lists are shown in Figure 3. The plot in the left panel is made up of the

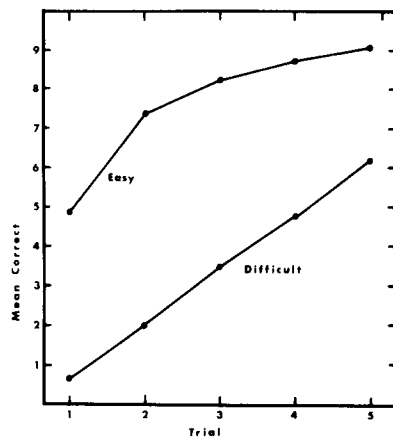


Figure 2. The acquisition of easy and difficult pairs in a mixed list with 10 pairs of each level of difficulty.

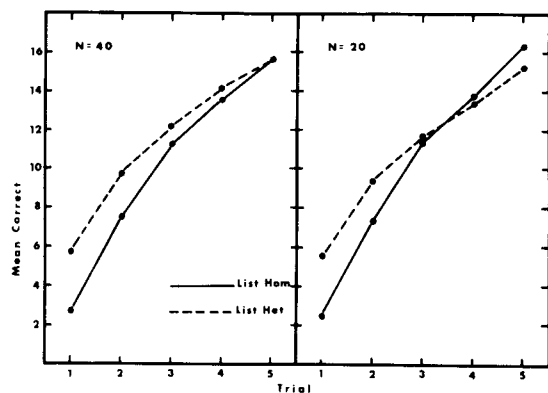


Figure 3. Acquisition curves for a list of pairs of homogeneous difficulty (List Hom) and a list of heterogeneous difficulty (List Het). The left panel is based on all 40 subjects in each group; each of the curves in the right panel is based on 20 subjects from the middle level of learning ability.

scores for all 40 subjects in the two groups. Performance on List Het is higher than that on List Hom on the initial trials, with the difference between the two decreasing across trials. An analysis of variance indicated that the two groups did not differ overall [$F(1,78) = 2.99$, $MSe = 58.18$]. The trials factor was a significant source of variance, of course, as was the interaction between trials and lists [$F(4,312) = 9.06$, $MSe = 3.54$].

The use of the data from all 40 subjects could be judged to be inappropriate for two reasons. First, some of the subjects in both groups correctly produced all 20 response terms before the fifth trial. This was true in spite of the rapid rate of presentation on the study trials. The implication is that this ceiling would necessarily produce a convergence of the two curves on the later trials. Opposed to this was the fact that slow-learning subjects given List Het did not reach the point at which the difficulty pairs entered substantially into the scoring on any of the trials. Thus, both factors would indicate that expectations might not be met because of the widely different rates of learning of the subjects. It seemed that further analysis was necessary.

All 80 subjects learned a common practice list before being given the critical list. The correlation between performance on the practice list and that for the group given List Het was .81. The corresponding value for List Hom was .76. Given these correlations, it seemed reasonable to select subjects from those learning the two lists so as to minimize both the ceiling effects and the effects of slow-learning subjects. This could be done by selecting average subjects from the middle of the distribution of the total correct responses on the practice list. For all 80 subjects, the range of scores for the practice list was from 14 to 87. Twenty subjects were selected from the scores on the practice list for the group that subsequently learned List Het and an equal number that subsequently learned List Hom. The range of practice-list scores for the first group was from 30 through 54, and the range for the second was from 28 through 50. For these two groups of 20 subjects each, the performance on the two critical lists was determined, and these results are shown in the right panel of Figure 3.

It can be seen that the major difference between the two plots in Figure 3 is the crossover of the two acquisition curves in the right panel that is not present in the left panel. Statistically, the results are the same for both plots, the analysis indicating that for the right plot the difference in the scores for the two groups did not differ [$F(1,38) = 1.07$, $MSe = 32.23$]. The interaction between trials and lists was reliable [$F(4,152) = 11.07$, $MSe = 2.68$].

In general, then, the data on the shapes of the learning curves conform to expectations based upon knowledge of the composition of pair difficulty in the two lists. If the acquisition curves are projected beyond Trial 5, it would seem that List Hom is an easier list

overall than is List Het, although it will be remembered that the mean difficulty of the two lists was equivalent. Whether this asymmetry is due to some negative interaction effects for the easy and difficult pairs in List Het or whether it is due to the nonequivalence of the intervals on the scale of pair difficulty is not known.

A secondary purpose of this experiment was to make further tests of the stability of the pair-difficulty values. The data in Figure 2 would seem to indicate stability, but more severe tests can be made. List Hom consisted of 20 pairs from the middle of the distribution of difficulty (Pairs 90-109), the range varying from 16 through 17 on the difficulty scale. Nevertheless, the correlation between the mean number of times each of the 20 pairs was given correctly in the original determination of pair difficulty and the number of times each was given correctly in learning List Hom was .62. This indicates that very small steps along the scale of pair difficulty are meaningful.

Number of Syllables

Data from the original study indicated that the number of syllables in the five-letter words was related to ease of learning; if both words in a pair had two syllables, learning was more rapid than if both words had one syllable. One might well speculate about the possible reasons for this finding, but it seemed necessary first to make a direct test of the role of number of syllables. This test, now to be described, used only the extreme cases, namely, pairs in which both words had two syllables and pairs in which both words had single syllables.

Method

Lists. Two lists of 16 pairs each were constructed. Because multisyllable words have, in the general case, lower word frequency than do single-syllable words, it was necessary to match items in the two lists for background frequency. Two lists of five-letter words were chosen so that frequency was equivalent, word for word, in the two lists. Words with frequencies less than one per million were not used, but the frequencies did vary between one per million and AA words. All words had a noun function. The 32 words in one list were paired randomly to produce a 16-pair list, and the 32 words in the other list were paired using the same random pairing used for the first list. The two lists will be identified as List 1S (one syllable) and List 2S (two syllables). Some of the pairs in List 1S were "roast-chain," "mouth-grave," and "dunce-thumb." For List 2S, three of the pairs were "rifle-cabin," "music-giant," and "diver-tiger."

Procedure and Subjects. All subjects learned both lists, half in the order of List 1S followed by List 2S, and half in the reverse order. Two study-test trials were given, the study trials being given at a 2-sec rate. Test trials consisted of the subjects trying to recall the correct response term to each stimulus term as given on a test sheet, with 60 sec allowed for them to write in as many correct responses as possible in the blanks following each stimulus term. There were two study-test cycles, with the presentation order of the pairs on the two study trials being different.

A group procedure was used, with from one to six subjects

being tested at a given time. Across successive groups, the order of the two lists varied. Testing continued until 40 subjects were completed; the two orders of the lists were represented by 20 subjects each.

Results

The mean total correct responses across the two trials on List 1S was 16.90, and for List 2S, 19.85. The difference was reliable ($t = 3.83$). There was no interaction between trials and lists. Thus, this evidence supports the finding of the earlier data, which constituted a post hoc analysis. It appears that multisyllable words are learned more rapidly as paired associates than are single-syllable words. Of course, there must still remain doubts about the true independent variable that produces this finding. It remains possible that some other characteristic of the words that is correlated with number of syllables is responsible for the observed differences in learning. Careful examination of the pairs has not suggested such a correlated factor, but it may be a subtle one.

A further study was carried out, the results of which place a restriction on the generality of the syllable effect. The two lists of 32 words used to construct the paired associate lists as given above were used as free recall lists. Two study-test trials were given, with the words presented at a 2-sec rate on the study trial and 60 sec allowed for recall. Twenty subjects learned the two lists, with 10 each assigned to the two orders. The mean total correct responses for the two trials were 24.50 for the list of two-syllable words and 23.60 for the list of single-syllable words. The difference was far from being reliable ($t = .90$). The failure of the number of syllables to influence free recall learning could argue against the possibility that some correlated factor was responsible for the difference observed when paired associates were used. That the findings were different for the two studies may also mean that the syllable effect occurs only when two words are to be associated with each other. Whatever, the mechanisms involved are not apparent to the present investigator.

Recall vs. Recognition

The question asked of this experiment was whether the difficulty of pairs as determined by paired associate learning (recall) will predict the recognition difficulty of the pairs. There were two purposes in asking the question. First, it will be determined whether the difficulty index of the pairs as given in the appendix can also serve as a difficulty index for recognition studies using pairs. Second, it is of theoretical interest to determine the relationship between recall and recognition using the same materials.

It is possible to list a large number of studies in which recall performance and recognition performance appear to be based on fundamentally different mechanisms.

However, most of these studies made use of single words rather than pairs. There is some evidence that when pairs are involved in recognition performance, there is a positive relationship between the difficulty measures for recognition and the paired associate learning of the pairs. In one study (Underwood, 1974), two sets of pairs were identified in recognition performance, one set consisting of easy pairs the other of difficult pairs. When these two sets of pairs were given in unmixed lists for paired associate learning, it was found that the easy pairs for recognition were also the easy pairs for paired associate learning. In the present study, the order is reversed, in that the difficulty level of the pairs in paired associate learning is known and the question concerns the relationship between these pairs of known difficulty and recognition performance for the pairs.

Method

Materials. Fifty pairs were used in the study list, plus two primacy and two recency buffer pairs. The 50 pairs were made up of five blocks of 10 pairs each, with the blocks differing widely in recall difficulty. The five blocks consisted of Pairs 1-10, 50-59, 95-104, 150-159, and 191-200 as listed in the appendix. A yes-no test was used that included 50 new pairs randomized with the 50 old pairs on the test sheet. The 50 new pairs were taken from various sections of the list of 200 as follows: 11-20, 45-49, 60-64, 90-94, 105-109, 145-149, 160-164, and 181-190.

Procedure and Subjects. The 54 pairs were presented for study at a 2-sec rate. The subjects were told that the words in the pairs consisted of both common and uncommon words. They were further told exactly how they were going to be tested after the single study trial. Immediately after the study trial, test sheets were passed out and the instructions were repeated. A decision was required on each pair by circling YES or NO. Subjects were tested in groups of from one to six; the testing continued until data for 40 subjects were collected.

Results

Overall, the misses averaged 24.6% and the false alarms averaged 18.9%. The critical data pertain to the relationship between pair difficulty in associative learning (as given by the values in the appendix) and the measures of recognition. Across the 50 pairs, the misses showed essentially zero correlation ($r = -.05$) with the associative learning values. There was no suggestion of nonlinearity in the scatterplot. As an extreme test, the misses for the 10 most difficult pairs were compared with the misses for the 10 easiest pairs. The mean number of misses for the former was 9.20 and for the latter, 7.70. The t was .96. On the other hand, the false alarms were seen to decrease as the pairs decreased in difficulty ($r = .46$). This correlation is probably a conservative estimate of the relationship, because there was some curvilinearity in the scatterplot. Overall, then, the data indicate that the number of misses in recognition of pairs is not related to the difficulty of learning the pairs as paired associates but the difficulty level does influence the number of false alarms, with the number being less for easy pairs than for difficult ones.

GENERAL DISCUSSION

The difficulty levels of the 200 pairs of five-letter words were found to be extremely stable in paired associate learning. Even small differences on the difficulty scale were found to be meaningful in predicting the rate of learning of the pairs by college students. It is not known, of course, whether or not this stability would be observed if lists consisted of 5 or 10 or 30 pairs, rather than the 20-pair lists used, but it seems quite unlikely that any serious disturbance of the difficulty of the pairs would be found to be based on list length (within limits, of course).

The data as a whole did not reveal critical factors determining pair difficulty. Response-term frequency had a small effect, as did number of syllables, but certainly these two factors account for only a small part of the difficulty variance. The data did reveal some factors that did not contribute to pair difficulty, including first-letter identity, stimulus-term frequency, and stimulus-response compatibility. It is probably also correct to conclude that idiosyncratic factors contributed very little to the difficulty variance. Furthermore, the context variable had no effect. That is, the difficulty of a pair was not appreciably disturbed by varying the other pairs being learned in the list. It was as if a pair has an inherent difficulty level that is not seriously altered by the nature of the other pairs in the list.

The critical question, of course, is what are the major factors that do determine pair difficulty. Why are "rouse-creel," "tinct-gauzy," and "spoor-crave" difficult pairs and "zebra-gruff," "totem-wives," and "kappa-fussy" easy pairs? Some differences may be produced by other known variables. For example, the concrete-abstract factor may be involved, but if one studies the pairs, this variable does not obviously "stick out." Indeed, I went over the list of 200 pairs many times seeking factors that might be related to learning, with only one success: the number-of-syllables variable. It was suggested that the distribution of pair difficulty (Figure 1) could indicate that the number of critical factors in determining pair difficulty is not large. It would seem that it would be possible to "spot" some of these, but this has not been true.

The recognition data address this same issue at a different level. Two matters must be considered. There is, first, the discrepancy between the data of the present study and those of an earlier experiment (Underwood, 1974). This earlier experiment showed that pairs differ-

ing widely on recognition performance also predicted the performance in paired associate learning. The present study, working from pair difficulty to recognition difficulty, showed no relationship between pair difficulty and misses on a recognition test. No solution to this problem has presented itself.

The second matter regarding the recognition results has to do with the fact that the misses were not related to pair difficulty, whereas the false alarms were related. Easy pairs had fewer false alarms than did difficult pairs. Such data make it difficult to maintain a uni-process theory of recognition. A post hoc explanation of this finding may be given. It is assumed that misses result from a failure of frequency discrimination and that this failure is unrelated to pair difficulty. The false alarms, then, could be influenced both by frequency information and by associative information, with the associative information viewed here as being "backup" information. It is as if the subject uses the associative information by saying, "Had that pair been in the study list, I would have remembered it." Thus, an easy pair is perceived as one that would have been remembered had it been studied, and this information is used to supplement the frequency information. To state this in signal detection terms, a subject's confidence is increased that the pair was not in the study list if the pair is perceived as being one that would be easy to learn as a pair.

Of course, the above interpretation assumes that the subjects can judge validly the difficulty of the pairs by the associative information, and there is no known evidence for this. It is also true that in the recognition test the subjects may not have made their decisions on the pair qua pair; rather, they may have reached a "no" decision by determining that one of the two words did not occur in the study list. So, to propose that associative information may be involved in producing the differences in the false alarms is clearly a speculation.

REFERENCES

- BADDELEY, A. D. Stimulus-response compatibility in the paired-associate learning of nonsense syllables. *Nature*, 1961, **191**, 1327-1328.
- THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Bureau of Publications, Teachers College, Columbia University, 1944.
- UNDERWOOD, B. J. The role of the association in recognition memory. *Journal of Experimental Psychology Monograph*, 1974, **102**, 917-939.
- UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning*. Philadelphia: Lippincott, 1960.

Appendix
Difficulty of 200 Paired Associates

	Mean	SD		Mean	SD		Mean	SD
1. rouse-creel	7.8	2.56	6. withy-clank	9.2	1.17	11. inner-strop	10.0	3.52
2. taker-clomb	8.2	2.14	7. trait-snell	9.4	3.61	12. rinse-smote	10.0	3.03
3. firth-leach	8.6	1.02	8. spoor-crave	9.8	4.26	13. aloud-shawl	10.6	3.88
4. tinct-gauzy	8.6	1.50	9. awake-spume	9.8	3.71	14. gorse-gules	10.6	2.33
5. dross-ephor	8.8	2.04	10. aside-burgh	9.8	.98	15. spear-gulch	10.8	3.25

	Mean	SD		Mean	SD		Mean	SD
16. nadir-shale	10.8	3.19	78. clamp-valet	15.2	3.25	140. drown-ditto	18.8	2.64
17. shard-plump	11.0	3.29	79. nizam-unlit	15.2	1.72	141. fairy-gourd	18.8	4.79
18. qualm-woful	11.0	1.55	80. ameer-shape	15.4	.80	142. fluid-forty	19.2	1.17
19. reedy-stept	11.2	3.19	81. found-finch	15.4	1.36	143. broom-newly	19.4	2.24
20. dingy-pease	11.2	3.49	82. lazar-voter	15.4	1.20	144. hollo-apron	19.4	3.14
21. tract-stork	11.2	2.71	83. river-pupal	15.4	2.80	145. mural-wiper	19.4	3.93
22. cruel-vouch	11.4	2.42	84. beget-farce	15.6	1.50	146. toxin-axial	19.4	1.96
23. sheer-stave	11.4	2.06	85. faint-rocky	15.6	3.50	147. gipsy-melon	19.6	3.61
24. cress-valor	11.6	3.61	86. broad-hawse	15.8	3.43	148. outer-knife	19.6	2.06
25. sadly-horde	11.6	3.56	87. humor-leash	15.8	2.14	149. twice-lease	19.6	3.26
26. sorry-stele	11.6	2.06	88. opine-soggy	15.8	2.93	150. donor-icing	20.0	3.03
27. anger-rowel	11.6	3.14	89. spill-brisk	15.8	1.47	151. flies-focus	20.0	1.67
28. velum-lasso	11.6	3.72	90. conic-about	16.0	2.28	152. cadet-slice	20.2	2.32
29. blent-boggy	11.8	2.14	91. merge-guard	16.0	3.35	153. dunce-build	20.2	3.60
30. locus-rebec	11.8	2.79	92. swear-bugle	16.0	2.76	154. troop-drive	20.2	1.83
31. tulle-float	11.8	2.32	93. loess-trial	16.2	.98	155. fugue-fifty	20.4	2.65
32. addle-squad	12.0	2.45	94. brief-lodge	16.4	1.85	156. lares-black	20.4	3.44
33. bairn-print	12.0	2.45	95. bushy-realm	16.4	4.67	157. poesy-pudgy	20.4	3.50
34. chore-codex	12.0	2.53	96. revue-sharp	16.4	3.20	158. durst-brown	20.6	2.15
35. quail-worse	12.4	2.42	97. stage-shade	16.4	2.33	159. major-testy	20.6	2.06
36. renal-lease	12.4	3.67	98. thank-delve	16.4	3.88	160. latex-chilly	20.6	2.06
37. whisk-dwell	12.4	1.36	99. bylaw-vigor	16.6	2.65	161. tiara-bloom	20.6	1.96
38. tapir-pesky	12.6	2.50	100. exude-skies	16.6	4.32	162. drain-opium	20.8	2.14
39. trunk-thane	12.6	2.06	101. perch-waive	16.6	3.38	163. blast-block	20.8	2.56
40. voile-lathe	12.6	1.62	102. store-hitch	16.6	2.87	164. moose-polar	20.8	3.76
41. grove-stoop	13.2	1.94	103. blurb-blear	16.8	2.56	165. fault-birth	21.0	4.15
42. nobly-molar	13.2	2.93	104. milky-metre	16.8	1.72	166. hazel-ennui	21.0	2.45
43. taste-crypt	13.2	3.71	105. wedge-clout	16.8	1.47	167. palmy-happy	21.0	2.61
44. delft-gross	13.4	3.26	106. motif-arbor	17.0	2.45	168. zoned-booth	21.0	3.03
45. fauna-plumb	13.4	3.61	107. poser-comma	17.0	2.97	169. ascot-cargo	21.2	2.79
46. idiom-squib	13.4	3.26	108. tooth-dowdy	17.0	3.16	170. wordy-level	21.2	3.54
47. quash-fiber	13.4	3.32	109. trine-udder	17.0	3.85	171. clung-hello	21.4	4.36
48. rheum-siege	13.4	2.33	110. brunt-blade	17.2	1.17	172. dress-civic	21.4	2.65
49. sewer-forgo	13.4	4.50	111. cross-diner	17.2	1.47	173. human-helve	21.4	3.32
50. torch-leant	13.4	3.01	112. julep-gauge	17.2	.75	174. hydra-peril	21.4	1.20
51. belle-wrack	13.6	2.73	113. steam-scrub	17.2	3.49	175. foyer-aisle	21.6	3.49
52. bosky-flyer	13.6	2.58	114. tight-covet	17.2	3.06	176. ninny-month	21.6	2.87
53. caulk-llano	13.8	4.58	115. amaze-tense	17.6	2.80	177. circa-rowdy	21.8	1.50
54. cover-extol	13.8	3.43	116. minor-gayly	17.8	3.31	178. daily-juror	21.8	1.33
55. downy-shuck	13.8	4.07	117. mufti-plant	17.8	3.66	179. haply-movie	21.8	2.14
56. exist-siren	13.8	1.47	118. usual-layer	17.8	2.56	180. slimy-bobby	21.8	3.54
57. prink-whelk	13.8	1.94	119. acute-ultra	18.0	2.53	181. amend-funny	22.0	1.67
58. scold-fungi	13.8	2.04	120. alive-glory	18.0	2.76	182. villi-under	22.0	1.10
59. twist-dizen	13.8	3.31	121. bravo-lowly	18.0	3.90	183. grass-slyly	22.2	2.99
60. spicy-heard	14.0	2.28	122. eject-prism	18.0	2.28	184. penal-jimmy	22.2	3.37
61. whose-spurn	14.0	1.67	123. heady-final	18.0	2.28	185. frost-scene	22.4	1.85
62. rapid-blunt	14.2	2.48	124. mango-flick	18.0	2.10	186. adult-begot	22.6	1.02
63. rosin-epoch	14.2	1.47	125. rural-heave	18.0	2.53	187. pupil-sober	22.6	3.01
64. scull-lupus	14.2	4.17	126. chaos-reach	18.2	2.64	188. crowd-photo	22.8	2.71
65. dower-demon	14.4	1.50	127. erase-grind	18.2	1.94	189. exile-dummy	22.8	3.06
66. gleam-serum	14.4	2.80	128. lunch-cling	18.2	1.94	190. onion-banjo	22.8	.98
67. route-creep	14.4	2.06	129. offal-knave	18.2	1.17	191. buyer-plane	23.0	3.03
68. tweak-polyp	14.4	2.42	130. eater-viand	18.4	3.32	192. embed-flute	23.0	3.03
69. weave-unsay	14.6	3.44	131. pshaw-pylon	18.4	3.20	193. fling-earth	23.0	2.10
70. bowls-grout	14.8	2.93	132. stalk-porch	18.4	3.61	194. zebra-gruff	23.0	1.67
71. hilum-crier	14.8	2.71	133. ulcer-chime	18.4	2.42	195. cello-chink	23.2	3.66
72. spice-dumpy	14.8	1.17	134. batik-lyric	18.6	3.32	196. thumb-comic	23.4	3.56
73. awoke-preen	15.0	2.00	135. befog-noway	18.6	2.42	197. adder-altar	24.2	1.94
74. carry-debit	15.0	2.28	136. cobra-stony	18.6	3.07	198. totem-wives	25.2	1.94
75. guilt-fated	15.0	3.63	137. flunk-egret	18.6	2.06	199. erect-floor	26.4	2.58
76. wafer-scorn	15.0	2.97	138. moody-greet	18.6	1.85	200. kappa-fussy	27.2	.98
77. calve-quean	15.2	2.79	139. spine-delay	18.6	4.22			

Note—The pairs are ordered from most difficult to least difficult, the difficulty level of a pair being determined by 50 subjects. The standard deviations were calculated from five scores, each score representing a different context for the pair. The mean difficulty level represents, in the abstract, the mean number of correct responses produced by 10 subjects in three trials. Thus, 30 is the maximum possible score.