The spacing effect: Additions to the theoretical and empirical puzzles

BENTON J. UNDERWOOD, SUSAN M. KAPELAK, and ROBERT A. MALMI
Northwestern University, Evanston, Illinois 60201

Four studies examined the MP-DP effect (spacing effect) in four quite different situations: recognition of letters, verbal discrimination, short free recall lists, and recall of MP items presented twice, with an intervening interval inserted to produce forgetting. MP-DP differences were found in all studies. Of particular interest were three interactions. Subjects with a low criterion of responding in the letter study lost the MP-DP effect over a 30-sec delay, and subjects with a high criterion did not. A clear MP-DP effect, but no lag effect, was found only with unmixed verbal discrimination lists. In free recall, a sharp lag effect was shown for words presented three times but not for words presented twice. A forgetting interval inserted between the two occurrences of an MP item did not appreciably aid its recall. The results were found to pose severe problems for current theoretical ideas about the spacing effect.

The four experiments reported here deal with two phenomena which may eventually yield to a common theory but which at the moment must be kept distinct as empirical phenomena. If a series of verbal units are presented for study, and if a unit occurs two or more times in the series, the practice schedule for these multiple occurrences may be massed or distributed. In massed practice (MP), the item occupies adjacent positions; in distributed practice (DP), successive occurrences are separated by at least one other unit. The spacing effect, or, as it will be called here, the MP-DP effect, refers to better recall following DP schedules than following MP schedules. The difference between MP and DP in recall is one of the most omnipresent effects found in the verbal learning laboratory (Hintzman, 1974).

The second phenomenon is known as the lag effect and refers to the relationship (usually direct) between recall and the amount of the separation (number of intervening units) of successive occurrences in DP. As Hintzman (1974) has pointed out, the lag effect lacks the universality found for the MP-DP effect. That is, the MP-DP effect may be produced without its magnitude being influenced by lag of DP repetitions.

No explicit theoretical tests were involved in the experiments to be reported. The past work on the MP-DP effect from our laboratory has led to a leaning toward an hypothesis which emphasizes attenuation of attention produced by the MP items (Shaughnessy, Zimmerman, & Underwood, 1974). However, the theory as developed thus far cannot be viewed as satisfactory, in that the most direct evidence available

indicates that attenuation will not account for the entire MP-DP effect, although it may account for the lag effect when found (Zimmerman, 1975).

Inductive theoretical development is normally facilitated by the discovery of situations in which a phenomenon can be made to vary in magnitude, particularly a variation to the extent that the phenomenon is no longer observed. This is to say that interacting variables are valuable in determining the contours of a theory. It was with this in mind that relatively novel situations were devised in an effort to discover interacting variables. The more specific notions guiding the design of the experiments will be given at the time each is introduced. The first two experiments to be reported involve recognition, the second two, recall.

EXPERIMENT I

A survey of previous studies indicates that the MP-DP effect is present in recognition memory (Allen & Garton, 1970; Hintzman & Block, 1970; Underwood, 1969; Winograd & Raines, 1972). Nevertheless, the number of different situations which have been examined for the MP-DP effect in recognition is far less than the number examined using recall. Three of the four studies cited above used long lists of words and classical word recognition procedures, the other (Winograd & Raines, 1972) using word recognition following the study of sentences. In the present experiment, the units presented for study were individual letters. Each list contained 17 different letters, and the subject was given 12 successive lists, with the letters chosen randomly for each list. Of course, within each list some letters were given under MP schedules and some under DP schedules. Assuming that our subjects knew the 26 letters of the alphabet, the task they faced was that of discriminating which particular 17 letters had been presented for study in the list of the moment. The letters were presented at a rapid rate
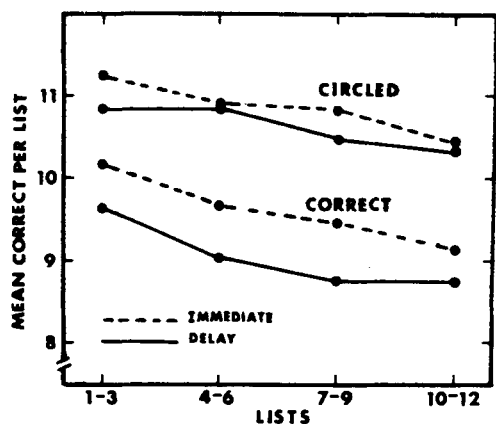
Figure 1. Mean number of letters recognized and mean number encircled as a function of list position and delay: Experiment I.

under the assumption that this would minimize organizational possibilities.

After a list was presented for study, the subject was given a sheet containing the 26 letters in order, and he merely circled those he believed had been in the list. Although these procedures correspond to those commonly associated with recognition tests, it seemed likely that the performance would reveal recall-like phenomena. In particular, it was expected that a recency effect would be manifest when the test was given immediately. Therefore, one group was tested immediately after the presentation of the last letter in the list, and a second group was tested after 30 sec.

Method

Lists. The lists were mixed in that both MP and DP items were included within each list. The 17 letters appearing in a list were assigned one of seven functions: two primacy, three recency, two occurring once within the body of the list, three shown twice under MP (MP-2), three twice under DP (DP-2), two presented three times under MP (MP-3), and two three times under DP (DP-3). Each list required 31 positions. Across the 12 lists, the letters were assigned to a class randomly subject only to the restriction that a particular letter was not allowed to serve in the same class more than twice. This restriction was not applied to the primacy and recency classes.

To prevent the subject from learning a particular pattern of repetition across lists, three different patterns were devised with four lists made up from each pattern. Across 24 subjects, blocks of three lists (one for each pattern) were systematically rotated, as were the lists within blocks, so that each list occurred twice in each of the 12 positions.

Items were assigned to positions within the list in a manner to minimize differences in position of last occurrence of MP and DP items. For letters presented twice, the average position of last occurrence for the DP letters was 19.5, for MP letters, 16.3. For letters presented three times each, the corresponding values were 18.8 and 18.5. The lag was not systematically varied for the DP items. Across the three list patterns, the lag for DP-2 items varied between 4 and 13, and for DP-3 items, between 2 and 11.

Procedure and subjects. It is obvious that a subject could get a perfect score by circling all 26 letters on the tests. It was necessary to stress in the instructions that only the letters which the subject was quite sure were in the list just presented should be circled.

The letter lists were presented on a memory drum at a rate of .91 sec, a value which includes change time. On the test, the subject was allowed 30 sec to circle the letters of his choice, after which the next list was presented. The subjects in one group were tested immediately after the list was presented (Group I), the subjects in a second group being given a 30-sec delay between study and test (Group D). The delay period was filled by requiring the subject to count backward by threes, the counting being initiated by the experimenter when a three-digit number was read to the subject. Thus, the interlist interval was 30 sec for Group I and 60 sec for Group D. There were 24 college students in each group, assigned by a block randomized schedule to group and to list order.

Results

General performance characteristics. It is first necessary to examine overall performance to determine the relationship between number of correct responses and number of letters circled. The data are plotted in Figure 1, with the lists grouped by threes. There is some decline across lists both in number of letters circled and number correct. Perhaps of greatest importance is to note that, on the average, approximately 11 letters were circled (17 were in each list) and that the discrepancy between the number circled and number correct was between one and two (1.26 for Group I and 1.57 for Group D). The maximum discrepancy observed for any subject among the 48 was 4.75, this subject having circled 17.75 letters, of which 13.0 were correct.

The statistical analysis of the data on which Figure 1 was based showed that the decrease across trials was reliable (using the .05 significance level), $F(3,138) = 4.05$, $MS_e = 24.94$, and that the decrease for number correct was greater than for the number circled, $F(3,138) = 3.97$, $MS_e = 2.11$. The difference between number circled and number correct was, of course, reliable, but the interaction between the groups and the two measures was not, $F(1,46) = 1.33$, $MS_e = 15.92$. Delay per se had no overall effect $(F < 1)$. The data indicate that sheer guessing was at a minimum and that there was a small decrement across lists, this decrement being somewhat greater for the number correct than for the number circled.

As a second preliminary step, performance on the two primacy letters, the three recency letters, and the two letters presented once within the body of the list was examined. For Group I the mean correct was 43.7%, 41.7%, and 68.3% for the primacy, body, and recency letters, respectively. The corresponding values for Group D were 48.7%, 41.5%, and 43.9%. The drop in the performance on the recency items was highly reliable, $t(46) = 6.93$, $\sigma_{diff} = 3.52$. These data indicate a delay effect quite like that found in free recall.

MP-DP. The percentages correct under the MP and DP schedules are shown in Figure 2. For comparison, the value for the letters presented once within the body of the list is shown, although this value did not enter into the statistical tests. It is obvious that the MP-DP effect was present for the lists of letters. Figure 2 suggests that the effect was attenuated with delay.
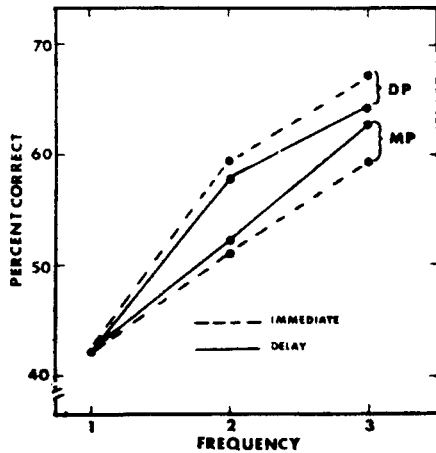
Figure 2. The effect of delay, MP-DP, and frequency on the recognition of letters: Experiment I.

However, statistically this was not supported. Overall there was no influence of interval (F < 1). Frequency was a reliable source of variance, $F(1,46) = 48.35$, $MS_e = 57.25$, as was the MP-DP difference, $F(1,46) = 17.71$, $MS_e = 93.20$. The interaction of MP-DP by Interval was, as indicated earlier, not reliable, $F(1,46) = 2.35$, $MS_e = 93.20$.

Figure 1 showed that performance decreased a small amount across lists. It may be asked whether the MP-DP effect changed in magnitude across lists. For the analysis all massed items were summed and all distributed items were summed, for each subject, for the first six lists, and separately for the second six lists. There was a reliable decrement across halves, $F(1,46) = 9.26$, $MS_e = 10.71$, and a reliable MP-DP difference, $F(1,46) = 19.07$, $MS_e = 8.29$, but the interaction of these two variables was less than one. Delay did not interact reliably with the other variables.

As seen in Figure 2, the MP-DP difference was less for Group D than for Group I, although the interaction was not significant statistically. Nevertheless, the reminiscence-like effect for the MP items was sufficiently intriguing to warrant further consideration. Suppose that, with a delay, the MP items, for whatever reason, more closely approached the criterion of acceptance than was true with no delay. If this did happen, it should be more apparent for subjects who had the most lax or low criterion for acceptance. Each group of 24 subjects was divided into two subgroups of 12 each, based on the criterion level. A subject was said to have a high criterion of acceptance if the discrepancy between number of letters circled and number correct was small, and was said to have a low criterion if this discrepancy was large. These discrepancy scores were determined for each subject, using the totals for all 12 lists. We then examined the number of MP items and DP items recalled for the 12 lists.

The outcome of this analysis is plotted in Figure 3, with the retention interval (Group I and Group D) along the baseline. For subjects with a high criterion,

the MP-DP effect was large and constant for both the immediate and delayed tests. On the other hand, for subjects with a low criterion, the MP-DP effect found on the immediate test was actually reversed on the delayed test. The statistical analysis showed only three reliable sources of variance, the MP-DP difference, $F(1,44) = 27.38$, $MS_e = 11.52$, the interaction of MP-DP by Criterion, $F(1,44) = 13.02$, $MS_e = 11.52$, and the triple interaction, $F(1,44) = 6.08$, $MS_e = 11.52$.

These data indicate that some characteristic of the subject, combined with the delay period, influences the MP-DP effect. Although the subgroups were defined in terms of criterion differences, there is no way to know if this characteristic is the critical one, or whether some other one correlated with it is responsible. Two auxiliary facts should be noted. First, the groups as distinguished by criterion differences did not differ in overall performance, whether viewed in terms of total correct responses or in terms of MP and DP items. Second, in searching for other differences between the groups, it was noted that performance on the recency items was better for the subjects with the low criterion than for those with the high criterion. This was true on both immediate and delayed recall, the values combined being 60.0% and 52.2% for low- and high-criterion subjects, respectively, $F(1,44) = 5.06$, $MS_e = 143.95$. Delay was highly reliable, of course, but did not interact with criterion groups. Whatever the fundamental characteristic which distinguishes these subjects, it seems to influence performance differentially over time only on particular items, primarily the MP items.

We have found that letters presented at a fast rate showed the MP-DP effect on both an immediate test and on a delayed test. Subjects who set a low criterion of acceptance of letters do not show the MP-DP effect after 30 sec. As will be true for all experiments, discussion will be delayed.

EXPERIMENT II

We turn next to recognition memory as exemplified in the verbal discrimination task. According to our
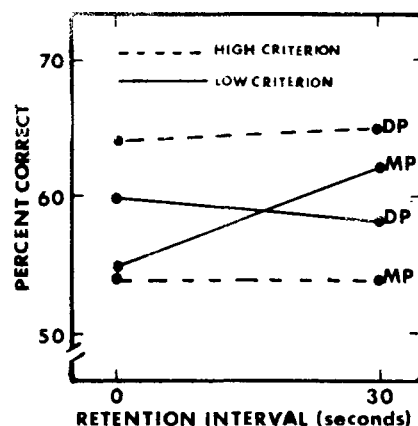

Figure 3. The effect of criterion differences and delay on the recognition of MP and DP items: Experiment I.

search, only one MP-DP study using this task has been reported. Ciccone (1973) reached the conclusion that the MP-DP effect was present; however, the procedures used were not typical of those for the verbal discrimination task. The subject was presented 12 pairs four times each on a single trial, during which he guessed which member of each pair was correct, and was given immediate feedback. In unmixed lists the four occurrences of each item were either massed or distributed (one occurrence in each quarter of the list). On the test, the 12 pairs were randomly mixed with 36 new pairs, and the subject was asked to indicate the pairs that had been presented for study. Following this, the subjects were asked to identify which word in the pairs checked had been called correct on the study trial. The subjects were given two successive lists using this procedure, but for both tests the same 36 "new" pairs were used as distractors. For both lists combined, the subjects having DP identified more of the old pairs (23.7/24) than did those having MP (20.3/24). Of those pairs that were correct, the correct word was identified 16.0 times following MP and 20.2 times following DP. Both sets of measures showed a reliable difference between MP and DP, hence the MP-DP effect. The effect of MP and DP on identifying the correct word in the pair (the essential requirement of verbal discrimination learning) is not completely clear from the data. Since the correct word could only be indicated for pairs identified as old, the values of 16.0 and 20.2 are not meaningful without adjusting for differences in initial recognition. When the adjustment is made, the values are 78.8% for MP and 85.7% for DP. The reliability of this difference is not known, but in recognition studies a difference of 7% is likely to represent a reliable difference. However this may be, it seemed to us necessary to examine the MP-DP variable, using as the critical test the ability to identify the correct word in a pair. This variable was studied in both mixed and unmixed lists, and lag was systematically manipulated.

Method

Lists. A pool of 96 two-syllable words with frequencies of from 1 to 10 in the Thorndike-Lorge (1944) tables was selected randomly from a larger sample drawn randomly from the tables. The 96 words were assigned randomly to two groups of 48 words, and then the words in each group were paired randomly. One group was always used as pairs presented twice for study, and the other group for pairs presented once for study.

The description of the DP unmixed list will be given first. With 24 pairs presented once and 24 presented twice, 72 positions were required in the study list. Quarters (18 positions) were used as a positioning unit in placing words within the lists. Of the 24 pairs presented twice, there were eight pairs at Lags 1, 3, and 7-8 (either 7 or 8). Two at each lag were assigned to each quarter. Six pairs presented once were also assigned to each quarter. The particular items assigned to quarters and, for the DP items, the particular items assigned to the lags were determined randomly.

The unmixed MP list was exactly the same as the unmixed DP list, except that all repeated (MP) pairs occupied adjacent positions within the quarter. The position of the second occur-

rence of the MP pair was identical to the position of second occurrence when the pair was used as a DP pair. This was made possible by changing the positions of some of the pairs presented once within the quarters.

The mixed lists were constructed to correspond to the unmixed lists. Half of the DP pairs (one at each lag) from the unmixed DP list were placed in corresponding quarters of the mixed list, and the other half of the pairs were used as MP pairs in the quarters. The lag of a given pair was the same for both mixed and unmixed lists. A second mixed list was constructed in which the DP pairs of the first mixed list became MP pairs in the second, and the MP pairs in the first mixed list became DP pairs in the second. Thus, in both mixed and unmixed lists, the same pairs were used as MP and DP pairs.

One word in each pair was randomly determined as the correct word. In presenting the list on the study trial, this word was underlined. For half the pairs, the underlined word was on the left, for the other half, on the right. The position of the underlined word remained the same on both occurrences of MP and DP pairs.

Procedure and subjects. The study lists were presented at a 2-sec rate on a memory drum. The subject was fully instructed concerning the nature of the list and of the test trial. After a single presentation of the 72-position lists, the test list of 48 pairs was presented, again at a 2-sec rate. The subjects' task was to call out the word in each pair that had been underlined on the study trial. Responses were required for all pairs. The test list positions were crossed by halves with regard to positions in the study list, in that half the pairs from each half of the study list were tested in each half of the test list.

Subjects were assigned by block randomization to one of the four groups representing the two mixed and two unmixed lists. Initially, 30 subjects were assigned to each of the unmixed lists and 15 to each of the mixed lists. A preliminary analysis indicated effects for the mixed lists that were statistically borderline. Therefore, two additional groups of 15 each were assigned to the mixed lists. Because the performance on the pairs presented once did not differ for these added subjects from those tested earlier, the data have been combined. Hence, results will be reported on 30 subjects in each of the four groups.

Results

It will be remembered that each pair presented twice occurred both as an MP pair and as a DP pair, and that when serving these different functions, occupied identical positions in the study and test lists. Under DP each lag was represented by only eight different pairs and, although these were assigned randomly, it is quite possible that the eight pairs as a group serving each lag could differ in difficulty. The use of precisely the same groups of pairs at the same positions under the MP schedule allows any variation due to differences in item difficulty to be assessed. These comparisons are made explicit in Figure 4, where the subgroups of MP items are plotted as a lag function to correspond to the lags given the subgroups under the DP schedule. A true lag function requires an interaction between lag and the MP-DP variable. The results for the unmixed lists are plotted in the left panel; those for the mixed lists are in the right panel. The performance on the pairs presented once (singles) is shown, their points connected by a dotted line to the points representing the performance for items presented twice with a lag of one. The reason for doing this will become apparent shortly.
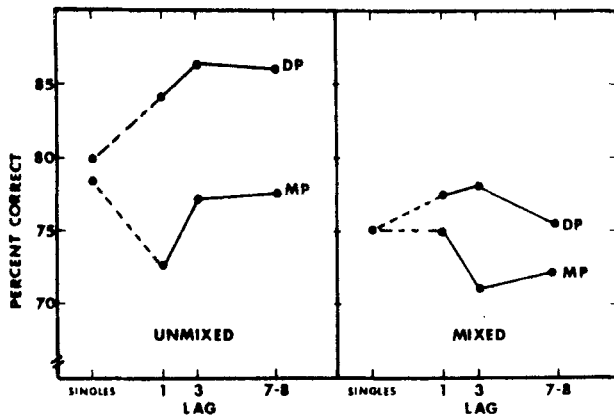
Figure 4. The effect of lag and list structure on verbal discrimination learning for mixed and unmixed lists: Experiment II. A lag is plotted for MP to show the recall for these items, which were identical to those in the DP conditions.

The results for the unmixed lists make it evident that there was the usual MP-DP effect, but no lag effect. What appears to be a small lag effect for the DP pairs is paralleled by a somewhat larger effect for the MP pairs and must, therefore, be discounted. The statistical analysis performed on the raw numbers correct showed the MP-DP difference to be reliable, $F(1,58) = 9.76$, $MS_e = 2.95$. The small differences as a function of "lag" were not reliable ($F = 1.55$), and the interaction was less than one. An examination of the 24 pairs presented twice showed that, for 21, the DP scores exceeded the MP scores when summed across subjects.

For the mixed lists in the right panel, the MP-DP effect is much diminished. Statistically speaking, there was no reliable source of variance among the six means. The only F greater than one was for the MP-DP variable, $F(1,59) = 3.06$, $MS_e = .66$. Of the 60 subjects, 24 showed better performance on DP pairs than on MP pairs, 19 showed the reverse difference, and for 17 the performances were equal.

Considering both experiments together, it would seem that there is an interaction between list type and the MP-DP variable. A statistical test was made by assigning randomly half the subjects from each mixed list to a DP group and half to an MP group. In the DP group, only the scores for the DP pairs were used and, in the MP group, only the scores from the MP pairs. A 2 by 2 analysis on scores representing percentage correct showed that the overall difference between the mixed and unmixed lists was reliable, $F(1,116) = 5.92$, $MS_e = 229.93$; but the F for MP-DP (3.53) and the F for the interaction (3.01) both fell short of the .05 level (3.93). The effect of list structure on the MP-DP difference, therefore, is not entirely clear statistically from these findings. It is apparent, however, that, if one wishes to produce a substantial MP-DP effect using verbal discrimination lists, the unmixed list structure should be chosen. Our results for such lists support the findings of Ciccone (1973) described earlier.

In Figure 4 it can be seen that the level of performance on pairs presented once did not differ appreciably from the level on pairs presented twice, for all cases except the DP pairs in the unmixed lists. In the latter case, the difference was reliable when the mean of the 24 pairs presented twice was compared with the mean of the 24 pairs presented once, $t(29) = 4.24$, $\sigma_{diff} = .33$. For the other three cases, the differences did not approach an acceptable level of significance. Thus, for 90 of the 120 subjects, the correct word in a pair presented twice was not consistently better discriminated than the correct word in a pair presented once. The 24 pairs presented twice and the 24 presented once had been determined randomly. However, to be sure that a sampling problem was not involved, we randomized the 48 pairs into a list, presented each pair once for study, and gave the test for recognition of the correct member of the pairs. The procedures were exactly the same as in the main experiment. For the 15 subjects used, the means were 18.47 and 18.20 (t = .26). The evidence in Figure 4 is therefore quite clear; for the MP pairs in the unmixed list, and for both MP and DP pairs in mixed lists, performance after two occurrences of a pair is no better than the performance after a single presentation.

EXPERIMENT III

In this experiment, we turn to the free recall task using relatively short lists, with lag varied systematically, and with the rate of presentation manipulated. The interest in the rate variable stems from two sources. First, the effect of rate on the MP-DP difference is not consistent in past studies. Waugh (1970) reports no MP-DP effect at a 1-sec rate, but a clear effect at a 4-sec rate, although there was no lag effect. Melton (1970), on the other hand, reported no interaction of Lag by Rate. Although, as Melton points out, the contradiction between his data and those of Waugh could lie in the method of presentation (aural vs. visual), other data which he reports make this doubtful. In still other studies (Underwood, 1970; Shaughnessy, Zimmerman, & Underwood, 1972), there was no interaction between rate and the spacing variable. It would appear that the preponderance of the evidence supports the conclusion that rate and spacing do not interact. Nevertheless, Waugh's results cannot be dismissed. One of the characteristics of her work is that subjects are tested across a number of short lists, and such a procedure may introduce certain factors, such as between-list interference, which are necessary for the emergence of the interaction between rate and the MP-DP effect.

The second reason for manipulating rate is a theoretical one. Assume that the subject has been given a number of lists so that he fully understands the demands of the task. In a given list, assume that the items are presented at a 1-sec rate. The knowledgeable subject
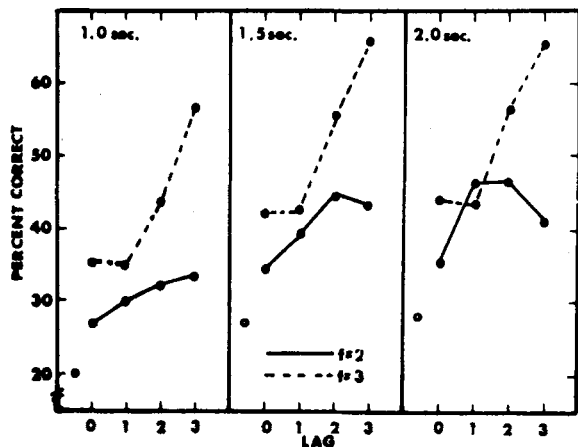
Figure 5. The influence of lag, frequency, and rate on free recall: Experiment III. The circles in the bottom left of each panel represent recall of words presented once.

may realize that during the study time allowed he is not able to encode the item to a level where he will be able to recall it. Under the MP schedule, the item is presented again immediately for an additional second. It would not seem unreasonable to presume that the subject would fully utilize the additional time for study. To say this in more theoretical terms, with a fast rate there should not be attenuation of attention and the MP-DP effect should be less than with a slow rate of presentation. Furthermore, the persistence of attention to an item on MP presentations may be more likely to occur when interlist interference is present, as might be the case when the subject is given a series of lists. The Waugh (1970) results, noted above, are in line with these expectations.

Method

Lists. Each subject was given eight lists, each list consisting of 20 different words. All words had four letters, and the 160 words were a random sample of a larger sample of such words taken from Thorndike and Lorge (1944). Within each list there were two primacy words and three recency words. In the body of the list there were three words for each of the following classes: singles (presented once), MP-2, MP-3, DP-2, and DP-3, where the numbers indicate the frequency of occurrence. Within each DP class, one word had a lag of one, another a lag of two, and the third a lag of three. With DP-3 items the lag was the same between successive occurrences, i.e., with a lag of two, there were two items between the first and second occurrences and the same number between the second and third occurrences. Each list required 38 positions to produce the repetitions.

Words were assigned to positions within the body of the list, using thirds as the placement unit. That is, of the three MP-2 items, one occurred in each third, and this was true of the other classes. Four different list patterns were used, such that lag length and list thirds were not confounded. Across the eight lists, each subject had 24 different words in each class, and eight different words for each DP lag. Furthermore, three different forms were constructed, with the assignment of words to classes being random subject to the restriction that a given word not serve in the same class more than once. As a consequence, summed across forms there were 72 different words represented in each class and 24 different words for each lag. Finally,

the eight lists for each form were ordered so that three different lists occurred at each of the eight positions.

Procedure and subjects. Three independent groups were used to implement the rate variable. The nominal three rates on the memory drum were 1.0, 1.5, and 2.0 sec. Measurements showed that, including change time, the true rates were 1.13, 1.71, and 2.22 sec.

Subjects were given the usual instructions for free recall, with added emphasis on the fact that the words would be presented at a rapid rate. Following the last word in a list, the word RECALL appeared as the signal for the subject to write as many of the words as he could, 90 sec being allowed for this. The study trial for the next list was presented immediately after the recall period of the previous list.

There were 30 subjects (college students) in each of the three groups, assigned to groups by a block randomized schedule.

Results

General characteristics of the learning. Overall performance was examined as a function of stage (list position) and rate. Rate was a reliable source of variance, $F(2,87) = 8.01$, $MS_e = 35.80$. Almost the entire effect was due to differences between the 1.0-sec rate and the other two longer rates. The mean numbers of correct responses per list were 6.95, 8.72, and 8.94 for 1.0, 1.5, and 2.0 sec, respectively. Performance increased a small amount across lists, $F(7,609) = 2.17$, $MS_e = 4.33$. This represented an increase of .25 items per list between the first four lists and the second four lists. Stage and rate did not interact $(F < 1)$. Primacy and recency effects were quite apparent, with the latter being of much greater magnitude than the former.

MP-DP. An examination of the difference in the recall of MP and DP items across the eight lists showed no systematic changes. The DP words were better recalled than the MP words for each successive list, but there was no systematic increase or decrease in the magnitude of the difference. Therefore, the scores were summed across lists to examine the influence of the three major variables. The recall scores were expressed as percentages and are shown in Figure 5. The three panels represent the three rates, the solid lines the words presented twice, and the dotted lines those presented three times. Lag is shown along the baseline and, of course, a lag of zero represents the MP items. The circles in the lower left corner of each panel identify the recall of words presented once within the body of the list. Because the data present a somewhat complex picture, the main effects of each variable will be noted, followed by the interactions.

As was true in overall recall, rate influenced performance on the MP and DP words, $F(2,87) = 7.51$, $MS_e = 1093.66$, but the difference was primarily produced by the 1-sec conditions. The overall values for the MP and DP words combined were 37.1%, 46.6%, and 47.7% for the three rates in order of magnitude.

Lag was reliable, $F(3,261) = 35.65$, $MS_e = 234.14$, as was also frequency, $F(1,87) = 116.96$, $MS_e = 184.44$. However, as may be seen in Figure 5, these two variables showed an interaction, $F(3,261) = 17.13$, $MS_e = 219.34$.

Essentially, with a frequency of two, there was an MP-DP effect but no lag effect, while, with a frequency of three, there was both an MP-DP and lag effect, the latter being of considerable magnitude for lags of two and three. There was no difference between MP recall (lag zero) and DP recall with a lag of one and a frequency of three, and this was true for all rates. The figure suggests that there is a change in the shape of the lag function across rates for words presented twice. However, this should be reflected in the triple interaction, but the F for this interaction was in fact less than one. Independent tests with the lag function for items presented twice confirm the fact that there was an MP-DP effect, no lag effect, and no interaction with rate. When the performance across the three rates was summed for the DP words presented twice, the result was a horizontal line. Waugh (1970) found no MP-DP effect at a 1-sec rate, found the effect at a 4-sec rate for items presented twice, but no lag effect. The present results confirm the latter finding at all rates, which is to say that we failed to replicate her results for the 1-sec rate.

The striking finding was that the lag effect in the present results was so critically tied to the frequency of presentation. Presenting an item for the third time with lags of two and three markedly increased the probability of recall. On the other hand, with a lag of one, the recall after three presentations was at the same level as that of MP recall.

EXPERIMENT IV

The purpose of the fourth experiment was to produce heavy forgetting of an item between its two presentations under a nominal MP schedule. An intuitive approach might suggest that, if an item presented once is forgotten over time so that it could not be recalled, a second presentation might have a very beneficial effect, if for no other reason than the subject would use the full exposure period during the second occurrence to study the item. Evidence presented by Melton (1967) would seem to deny this possibility. With lag as the basic variable, items were presented twice, and the subject was required to make a recognition decision for each word as it was presented for study. He found that words not recognized on their second occurrence were more poorly recalled than were those that were recognized. On the other hand, Bjork and Allen (1970) indicate that the second occurrence has a more pronounced effect on an item that is forgotten than it does for an item that is not forgotten. Tzeng (1973) presents data that confirm this finding. The present experiment does not examine this issue directly. Rather, our interest was in determining if it was possible to produce a recall level for an MP item comparable to that for a DP item by producing a memory loss for the first occurrence under both schedules.

The plan called for the production of forgetting over a filled interval. The first occurrence of the MP item would be just prior to interval, its second occurrence immediately after. The forgetting was produced by proactive interference from items presented earlier in the list. A pilot study was carried out in which lists of from 3 to 10 words were presented once each, at a 1-sec rate. The subject did not know for any given list how many words would be shown, but he knew that his task was always to recall as many as possible of the last three words presented just before a 20-sec retention interval was initiated. It was found that, with 8 to 10 words in a list, recall of the last three words was approximately 30% after the 20-sec filled interval. This seemed sufficient to establish that appreciable forgetting of the recently presented words would occur, and the design of the experiment followed from this fact.

The two critical conditions may be schematized as follows, with A representing the word of interest and x representing a neutral word.

Item Type	Sequence of Events
MP-F (forget)	x x x x x x x x x A—Interval— A x x x—Interval—Recall
DP-S (short lag)	x x x x x x x x A x—Interval— A x x x—Interval—Recall

As can be seen, the schedules for the two items differed only in that a neutral item occurred after the first occurrence of A in the DP schedule, whereas the first occurrence of A in the MP schedule was followed immediately by the interval. The second interval, between study and recall, was used to prevent excessively high recall of A and to eliminate recency effects.

In the above schedule, the DP item has a lag of one, and this might be viewed as unsatisfactory as a control for MP-F. Therefore, DP items were also used in which the first occurrence was in the early part of the list and the second occurrence was after the within-list interval. It seemed possible that the MP-F item might be recalled less well than the DP-S item but better than a "standard" MP item. In some lists, therefore, usual MP items were included in the early part of the list.

Method

Lists. The subject was given two practice lists followed by 18 experimental lists. Each list contained 12 different words and 14 positions, i.e., two words within each list were repeated. Three different list types, six of each, were needed to accommodate the different item types and to provide a sufficient number of observations for stability in the recall scores. In one type, the DP-S manipulation was involved; also included in each list was a DP-L (long lag) word. The first occurrence of the DP-L word occurred either in Position 5 or 6, and the second occurrence either in Position 12 or 13. In a second list type, the MP-F schedule was used, and this type also included a DP-L item in each of the six lists. The first occurrence of the DP-L items

Table 1
Recall Measures for the Various Item Types

Item Type	Number Possible	Mean Recall	SD	Percentage Recall
MP-F	6	2.80	1.12	46.7
DP-S	6	3.43	1.27	57.2
MP	6	2.38	1.42	39.7
RC	6	1.10	1.04	18.3
DP-L	18	10.23	2.50	56.8
Primacy	18	4.50	2.92	25.0
Recency	18	3.80	1.80	21.4
Others	138	25.20	11.71	19.5

in these lists was either in Position 2 or 3, while the second occurrence was in either Position 12 or 13. The six lists constituting the third type contained a normal MP word occupying Positions 5 and 6 or 8 and 9. A DP-L word also occurred in each of these lists, the first occurrence being in either Position 4 or 5 and the second occurrence in either Position 12 or 13. In addition, this list type was used to provide evidence on the recall of an item presented only once within the list but occupying the position just prior to the within-list interval. The concern was that some rehearsal of the item presented just prior to the interval might occur in spite of the fact that the interval was filled with counting backward by threes. If such rehearsal did occur, it would also inflate the recall for the MP-F item. This control item (RC, rehearsal control) was used to assess possible rehearsal effects by comparing its recall with other neutral words presented once in the list. Actually, the list type involving the DP-S items could also have been used for the RC control, but we chose to use only one such list in the rotation of items.

All words had five letters. From a pool of 240 such words, 24 were drawn randomly and assigned to four subgroups of six each. Four list forms were made up, such that across forms each subset of six words served each of the four critical functions (MP, RC, MP-F, DP-S) once. As a result, differences in item difficulty could not be involved in differences in recall. Since there were 18 different DP-L words, chosen randomly from the larger pool, it did not seem necessary to include them in the rotation scheme. All other words were assigned to lists and to positions within the lists on a random basis, including 24 words to the two practice lists.

The three list types were block randomized across the 18 positions, with the particular list from a given type assigned to a given block determined randomly.

Procedure and subjects. The words were presented on a memory drum at a 1.13-sec rate, including change time. After the words in the first 10 positions had been presented, a three-digit number appeared which served as the signal for the subject to count backward by threes. The instructions stressed the importance of doing this as rapidly as possible and, after each list, the experimenter notified the subject of the number of subtractions he had made. After the 20-sec interval, the word READY appeared and the additional four words were presented. Then a three-digit number appeared, and again the subject counted backward by threes for 20 sec. At the end of this interval, the word RECALL appeared, and the subject wrote the words from the list in any order he chose. The recall period was 30 sec. The study trial on the next list followed immediately. The practice lists were used to adjust the subject to the sequence of events.

A total of 40 subjects was tested, 10 on each of the four forms.

Results

The average number of words recalled per list across the 18 lists was 3.07 (25.6%). A plot for the 18 lists

showed a range of means from a high of 3.55 (29.6% to 2.55 (21.3%), but there was no systematic increase or decrease across the lists. The recall of the critical items was not influenced appreciably by the stage o practice, so the results for the 18 lists have been combined.

Recall measures for the various item types are shown in Table 1. In addition to the types described earlier recall for the first item (primacy), the last item (recency), and all other items presented once are shown in Table 1. Because of the different numbers of items involved in the classes, percentage scores are given in the last column.

The first clear fact shown in Table 1 is the no surprising result that items presented twice have a dis tinct advantage over those presented once. Summec across all conditions, items presented twice producec 52.3% recall, those presented once, 19.2%. Turning tc the more critical comparisons, the following results neec to be emphasized.

(1) The concern about rehearsal during the within-list interval was needless. The recall for this control (RC) was 18.3%, quite comparable to the overall recall of items presented once.

(2) The short lag DP items (DP-S) produced better recall than did the MP-F items. The difference between the two means was reliable, $t(39) = 2.23$, $\sigma_{diff} = .28$. Thus, a single neutral item inserted just after the first occurrence of the DP-S item, and coming just before the within-list interval, resulted in higher recall for the DP-S item than was found for the MP-F item, where the first occurrence occupied the position of the neutral item for the DP-S paradigm.

(3) The recall of the two classes of MP items (MP-F and MP) did not differ, $t(39) = 1.63$, $\sigma_{diff} = .26$, although the direction of the difference was as expected.

(4) The two classes of DP items (DP-S and DP-L) had almost identical recall (57.2% and 56.8%) and, of course, both were higher than the recall for the standard MP items by approximately 17%.

It would appear that the answer to the central question prompting this experiment is quite clear, namely, a DP-like recall was not created by inserting a forgetting interval between the two occurrences of the MP word.

GENERAL DISCUSSION

The critical results of the four rather diverse experiments will be summarized, maintaining the distinction between the MP-DP effect and the lag effect as made in the introduction.

(1) A rapidly presented series of 17 letters resulted in an MP-DP difference when the subjects on the test circled the letters from a list of all 26 letters. Subjects who set a low criterion for circling letters failed to show the MP-DP effect after a 30-sec delay, whereas the subjects with a high criterion showed a constant magnitude for the MP-DP difference.

(2) An MP-DP effect (but no lag effect) was found for unmixed lists in single-trial verbal discrimination performance. It was also found that, in lists where the MP-DP effect was small or missing (MP unmixed, both mixed lists), two presentations of a pair did not result in better performance than did a single presentation of the pair.

(3) In the free recall of short lists, the MP-DP effect was present, and there was no interaction with rate of presentation. A sharp lag effect occurred for words presented three times but not for words presented twice.

(4) The basic finding that DP yields performance that is superior to MP did not change when the MP item was presumably forgotten following its first presentation.

We know of no theoretical statements which will handle these findings along with the past findings of many investigators. The theory which uses attenuation of attention as a basic assumption is faced with problems presented by some of the above findings. The theory assumes that attenuation is directly related to redundancy between information being presented at the moment and information available in memory. The fourth conclusion above would not be predicted by this formulation. If the item is forgotten over the retention interval, there should not be redundancy when it is presented again after the interval. The fact that the lag effect in Experiment III was observed only with items presented three times is also a stumbling block for the attenuation of attention theory. With a constant lag, at the second occurrence of the DP item redundancy should be less than upon the third occurrence. In simple terms, on the third occurrence, the subject would be more likely to say "I know that word" than he should be upon its second occurrence. A lag effect by this theory might be anticipated for items presented twice, but the lack of a lag effect for such items and the sharp influence of lag with items presented three times would clearly not be anticipated by the attenuation theory. It would probably not be anticipated by any other theory either. For example, an approach which stresses context multiplicity for DP would be in the position of assuming that the context does not change as a function of lags beyond one for words presented twice, but does change when the word is presented three times.

The fact that the MP-DP difference was found with letters presented at a rapid rate might be handled by a "chunking" approach. It seems beyond doubt that DP items would have a higher probability than would MP items of being included in a group of two or three letters which formed words or their phonetic equivalents. An attempt was made to analyze for this possibility, and to relate it to performance, but it proved essentially impossible to carry out. Innumerable problems arose in specifying the chunks. Should letters in reverse order be used? How much leeway should be allowed in specifying phonetic equivalents? Is the two-letter unit

EG a chunk? Although we could not bring data to bear directly on the problem, the gross data present a strong argument against a chunking assumption. The MP-DP difference was essentially the same at both frequencies of two and three. A DP item should have a greater probability of falling into a chunk the greater the number of its occurrences, whereas this should remain constant with MP items, regardless of frequency. Yet, the number of correct items under both schedules increased equally between frequencies of two and three.

Perhaps the most interesting theoretical lead produced by the letter experiment was the finding that the MP-DP effect changed over a short delay for subjects identified as having a low or lax criterion for responding. It was as if the MP and DP items had been learned equally well but that the criterion set excluded the MP items on the immediate test but not on the delayed test. However, it should be emphasized that, although the subjects were distinguished on the basis of criterion differences, we do not know that this characteristic is responsible for the marked interaction between the MP-DP difference and time.

The MP-DP effect was clearly evident in unmixed verbal discrimination lists, but was at best a small effect for mixed lists. Perhaps the explanatory problem can be narrowed somewhat. This may be done by examining the reason why items presented twice did not result in better performance than did items presented once in all lists except the unmixed DP list. It may be assumed that, on each presentation of a pair, two frequency units accrued for the correct word and one for the incorrect. With a single presentation, therefore, the correct-incorrect ratio is 2:1. With two presentations the ratio would be 4:2. According to data presented by Hintzman (1969), the accuracy of distinguishing between the frequency of words having these two ratios is equivalent (Weber's law). And, these ratios could be the same for MP and DP schedules. The problem that remains, however, is why these ratios did not hold for the DP items in the unmixed list. It is possible to suggest several ways which would change the ratio, e.g., the subject might attend only to the correct word, but why these changes in behavior would occur with this list and not with the others is not at all apparent.

It was noted in the introduction that a major purpose of our studies was to try to discover situations in which the magnitude of the MP-DP effect and the lag effect varied as a function of some other variable. Three such interactions were discovered: (1) the interaction between delay and subjects having different criteria for responding; (2) the interaction with list structure in the verbal discrimination study, an interaction that was not, however, impressive statistically; (3) the interaction between lag and frequency in Experiment III. We have not been able to perceive a common process which might underlie these three interactions, or even two of them, so it is not evident that as a group the findings

will simplify inductive theory construction. We have great sympathy for Hintzman, Summer, and Block (1975). In a recent study on the spacing effect, they found that their successive theoretical notions could be rejected as rapidly as they could perform successive experiments to test them.

A final comment will be of a general nature. We believe that the evidence has accumulated to the point where one can no longer accept the proposition that a uniprocess theory will account for the diverse findings attending the spacing effect. It seems beyond doubt that MP and DP produce differences in the level of associative learning among items within a task. Why this occurs is one theoretical issue. Some evidence can be found to support a number of notions, notions which include attenuation of attention, encoding variability, context changes, and so on. So also, evidence can be found against these ideas. But, even if these ideas or some others prove to be satisfactory in accounting for the MP-DP effect in free recall and paired associate learning, there remain other tasks, such as the verbal discrimination task, where associative learning among the items is not involved in any causal way in the performance differences observed under MP and DP. Our theories must also be prepared to accept as an explanatory problem the fact that in nearly every experiment some subjects will not show the MP-DP effect. Our direct attacks on this matter thus far have failed because of the inability to demonstrate that the magnitude of the MP-DP effect is a reliable one for individuals. In three of the experiments reported here, multiple lists provided several measurements of the magnitude of the MP-DP effect for each subject. Without exception, the magnitude of the effect was found not to be a reliable one. This was the case even when scores reflecting the MP-DP effect were summed over lists in an attempt to provide stable estimates of the effect for each subject. However, this must result from a deficiency in our measurements, because it is difficult to see how so universal an effect could not have some reliability. Associative learning differences are highly reliable across subjects, and, if our theories about the MP-DP effects are to stress associative learning differences, the reliability of the MP-DP difference must surely emerge under appropriate conditions of measurement.

REFERENCES

ALLEN L. R., & GARTON, R. F. Manipulation of study trials in recognition memory. *Perception & Psychophysics*, 1970, 7, 215-217.

BJORK, R. A., & ALLEN, T. W. The spacing effect: Consolidation or differential encoding? *Journal of Verbal Learning and Verbal Behavior*, 1970, 9, 567-572.

CICCONE, D. S. Massed and distributed item repetition in verbal discrimination learning. *Journal of Experimental Psychology*, 1973, 101, 396-397.

HINTZMAN, D. L. Apparent frequency as a function of frequency and the spacing of repetitions. *Journal of Experimental Psychology*, 1969, 80, 139-145.

HINTZMAN, D. L. Theoretical implications of the spacing effect. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola symposium*. Potomac, Md: Lawrence Erlbaum, 1974.

HINTZMAN, D. L., & BLOCK, R. A. Memory judgments and the effects of spacing. *Journal of Verbal Learning and Verbal Behavior*, 1970, 9, 561-566.

HINTZMAN, D. L., SUMMERS, J. J., & BLOCK, R. A. What causes spacing effect? Some effects of repetition, duration, and spacing on memory for pictures. *Memory & Cognition*, 1975, 3, 287-294.

MELTON, A. W. Repetition and retrieval from memory. *Science*, 1967, 158, 532.

MELTON, A. W. The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior*, 1970, 9, 596-606.

SHAUGHNESSY, J. J., ZIMMERMAN, J., & UNDERWOOD, B. J. Further evidence on the MP-DP effect in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, 1972, 11, 1-12.

SHAUGHNESSY, J. J., ZIMMERMAN, J., & UNDERWOOD, B. J. The spacing effect in the learning of word pairs and the components of word pairs. *Memory & Cognition*, 1974, 2, 742-748.

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

TZENG, O. J. L. Stimulus meaningfulness, encoding variability, and the spacing effect. *Journal of Experimental Psychology*, 1973, 99, 162-166.

UNDERWOOD, B. J. Some correlates of item repetition in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, 1969, 8, 83-94.

UNDERWOOD, B. J. A breakdown of the total-time law in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, 1970, 9, 573-580.

WAUGH, N. C. On the effective duration of a repeated word. *Journal of Verbal Learning and Verbal Behavior*, 1970, 9, 587-595.

WINOGRAD, E., & RAINES, S. R. Semantic and temporal variation in recognition memory. *Journal of Verbal Learning and Verbal Behavior*, 1972, 11, 114-119.

ZIMMERMAN, J. Free recall after self-paced study: A test of the attention explanation of the spacing effect. *American Journal of Psychology*, 1975, 88, 277-291.