

# Persistence of the spacing effect in free recall under varying incidental learning conditions

JOHN J. SHAUGHNESSY  
*University of California, Berkeley, California 94720*

In three incidental learning experiments, an attempt was made to eliminate the processing deficit under massed presentation that is assumed to be responsible for the spacing effect in free recall, according to the attenuation of attention hypothesis. This was to be accomplished in Experiment I by requiring subjects to attend specifically to the total exposure duration of each item and in Experiments II and III by requiring subjects to rate the successive occurrences of repeated items on different semantic rating scales. The results of the three experiments consistently showed that these manipulations were ineffective in eliminating the spacing effect. Subsidiary analyses indicated that the activities involved in doing the semantic rating tasks do not provide direct access to retrieval cues useful for subsequent recall. Instead, it appears that, in order to perform the semantic rating tasks reliably, subjects must compare the to-be-rated item with previously rated items, and this comparison process may serve as the source of retrieval cues for subsequent recall.

Recall following a single study trial is better for items repeated according to a distributed presentation schedule (DP) than for items repeated according to a massed presentation schedule (MP). Several theoretical explanations have been proposed to account for this so-called spacing effect in free recall (for a review, see Hintzman, 1974). The present experiments can be viewed as further tests of one of these explanations, the attenuation of attention hypothesis, for which there is already some empirical support (e.g., Shaughnessy, Zimmerman, & Underwood, 1972, 1974). According to this hypothesis, the spacing effect is assumed to be the result of a deficiency in the processing of repetitions under MP.

Hintzman (1974) has argued that, if the processing deficits assumed to be present under MP are the result of voluntary processes, then it should be possible to eliminate the deficits through the use of instructional manipulations. On the other hand, Hintzman argues, if the processing deficits under MP are the result of involuntary processes, then the instructional manipulations should have no effect on the magnitude of the spacing effect. Hintzman and Summers (cf. Hintzman, 1974) tried to manipulate the attention given to the second occurrences of repeated items by telling subjects that certain of the MP and DP items would be worth 4 cents on the retention test, while the other MP and

DP items would be worth only 1 cent. They argued that the added incentive should serve to eliminate the processing deficit under MP, and thereby eliminate the spacing effect. The findings on a frequency judgment test showed an increase due to the incentive for both MP and DP items, but no reduction in the spacing effect.

The problem with motivational instructions of this type is that, even if the spacing effect does result from voluntary processing deficits, the amount of additional processing that subjects will be willing to expend in learning a given item will be dependent upon how well the subjects think they know that item. If subjects believe that they already know the MP items, they may see no need for any additional processing. The subjects may in fact think that they are more likely to obtain the incentive on the MP items, because these items appear to them to be easier. There is some recent evidence that subjects do overestimate how well they have learned MP items (Zechmeister & Shaughnessy, Note 1). If it is reasonable to assume that subjects' processing efforts are dependent upon how well they think they know the presented item, and if subjects do think they know MP items, then it will be difficult to eliminate the presumed processing deficit under MP in any intentional learning situation. Because of this difficulty, it seemed reasonable to assume that the elimination of the processing deficit under MP would require the use of an incidental learning task.

## EXPERIMENT I

The first attempt to control subjects' processing activities through the use of an incidental learning task proved to be unsuccessful for two reasons. First, subjects did not approach the task in the anticipated manner. Second, the level of recall under the incidental condi-

I want to thank Leo Postman for his helpful comments on an earlier draft of this paper. I also want to thank Gloria Shay for conducting Experiment II and Margaret McCullough and Robert Strawhorn for conducting Experiment III. Results of Experiment I were previously reported with Benton J. Underwood at the 46th Annual Meeting of the Midwestern Psychological Association, Chicago, May 1974. Requests for reprints should be sent to the Department of Psychology, Hope College, Holland, Michigan 49423.

Table 1  
Mean Number of Words Recalled and Mean Duration Judgments  
(in Seconds) for Once-Presented Items as a Function  
of Input Instructions

	Exposure Duration			
	1 sec	3 sec	5 sec	7 sec
	Recall*			
Duration Instructions	.50	.44	.38	1.03
Recall Instructions	.81	1.06	1.34	1.31
	Duration Judgments			
Duration Instructions	2.35	2.87	2.73	3.92
Recall Instructions	1.76	1.73	2.01	2.12

\*Maximum possible recall in any one cell is four.

tions fell well below that in the intentional conditions. Because of these difficulties, only a brief description of the procedures and results of this experiment will be provided.

The particular incidental task used in the first experiment was designed to require subjects to attend to each item for the entire presentation interval. This was to be accomplished by requiring subjects to attend specifically to the duration of the item's presentation. Subjects were instructed that they were to keep a cumulative record of the exposure duration of repeated items, and they expected a test requiring the estimation of the total exposure duration of each item. It was anticipated that subjects expecting this type of memory test would process each item throughout its presentation. Then, on the subsequent duration-judgment test, they would estimate how long they had processed each item.

Only half of the subjects expecting the duration-judgment test were given it; the other half were given an unexpected free recall test. If the subjects did process each item for the full presentation interval, there should be no difference in the processing time for MP and DP items. Thus, the spacing effect should be eliminated in the group given the free recall test. For purposes of comparison, two groups of subjects were included who were given free recall instructions prior to the presentation of the study list. Half of these were given a free recall test and half were given an unexpected duration-judgment test. Thirty-two undergraduate students were assigned randomly to each of the four independent groups, for a total of 128 subjects.

The study list consisted of a primacy buffer of seven positions, including one MP and one DP item; a recency buffer of four positions; four items presented once each in the body of the list for 1, 3, 5, or 7 sec; and four items presented twice either MP or DP, with each presentation being 2 sec. Eight different random assignments of items to list functions were used to make up eight forms of the study list. For the duration-judgment test, a single test form was used. It consisted of an alphabetical listing of the 33 study-list items and four distractor items that were not on the study list. Subjects were

instructed to estimate the *total* exposure duration of each item in seconds using whole numbers from zero to nine. Those subjects given the free recall test were allowed 5 min for recall.

The mean number of once-presented items recalled as a function of exposure duration and input instructions is presented in the upper half of Table 1. Overall recall in the group given free recall instructions was considerably higher than that in the group given duration-judgment instructions. Summing over the instructional variable, recall did increase with increasing exposure duration.

The mean number of twice-presented items recalled as a function of the MP-DP and instructional variable is presented in the upper half of Table 2. Again, the overall level of recall in the recall instruction condition was higher than that in the duration instruction condition. There was a significant spacing effect; and although the spacing effect was reduced in the group given duration instructions, it was clearly not eliminated. The critical predicted interaction of the MP-DP and instructional variables provided an  $F < 1$ .

For completeness, the mean duration judgments for once-presented items are presented in the lower half of Table 1 and those for MP and DP items are presented in the lower half of Table 2. Analysis of the judgments of once-presented items showed that: (a) overall, duration judgments were higher for subjects given duration-judgment instructions; (b) duration judgments increased with increasing exposure duration; and (c) the increase in the judgments with increasing exposure duration was more pronounced for subjects given the duration-judgment instructions. There was a significant spacing effect in the duration judgments for twice-presented items. The interaction of the input instruction and MP-DP variables again provided an  $F < 1$ . Finally, the duration judgments seem to have been influenced by both frequency and exposure duration, in that twice-presented items (presented for a total of 4 sec) were consistently judged as having been presented longer than once-presented items presented for 5 sec.

If it was argued that voluntary processing deficits under MP were eliminated by the present manipulation, then the presence of a spacing effect could be taken

Table 2  
Mean Number of Words Recalled and Mean Duration Judgments  
(in Seconds) for MP and DP Items as a  
Function of Input Instructions

	MP Items	DP Items
		Recall*
Duration Instructions	.88	1.16
Recall Instructions	1.84	2.31
	Duration Judgments	
Duration Instructions	3.20	3.83
Recall Instructions	2.45	3.32

\*Maximum possible recall in any one cell is four.

is evidence for involuntary processing deficits such as might result from the habituation process proposed by Hintzman (1974). Unfortunately, the present results do not allow such a definitive conclusion. Postexperimental questioning of subjects revealed that subjects who were given the duration-judgment instructions at input did not approach the task as had been expected. Subjects reported that they used the onset of a given word as a "start" signal; they then tried to keep track of the interval length using some technique, such as rhythmic counting, until the offset of the given word; finally, they tried to associate the word to the interval duration. As such, they were clearly not processing the to-be-remembered word for the entire exposure duration. Because subjects were required to keep a cumulative record of the total exposure duration for repeated items, a recognition decision would be required as each item was presented. It is reasonable to assume that this recognition decision may have taken longer for DP items than for MP items; thus, DP items may have been processed longer than MP items. This difference in processing time may have been responsible for the spacing effect in recall for subjects given duration-judgment instructions.

However it was that subjects approached the duration-judgment task, the recall results make it obvious that the processing of the to-be-remembered material under the duration-judgment instructions was not comparable to that under the free recall instructions. Recall following the incidental learning instructions was consistently lower than that following the intentional learning instructions. This difference contrasts sharply with the lack of a difference found in studies like those of Hyde and Jenkins (1973). In such studies, subjects required to perform so-called semantic rating tasks show recall essentially equivalent to that of subjects instructed to try to learn the items without being required to perform the rating task. These semantic rating tasks might serve as incidental tasks which could be used to eliminate the processing deficits under MP, while still providing a level of recall comparable to that found under intentional learning conditions. This possibility was examined in the next two experiments.

## EXPERIMENT II

In Experiment II, four independent groups of subjects were differentiated on the basis of the instructions they were given prior to list presentation. Three of the groups were given incidental learning instructions requiring the use of rating tasks similar to those used by Hyde and Jenkins (1973). The fourth group was an intentional learning condition given typical free recall instructions (Cond FR). All subjects were given a free recall test following list presentation and, of course, a spacing effect was expected in Cond FR. The reason-

ing underlying the prediction of the presence or absence of a spacing effect in each of the incidental learning groups will now be provided.

In the first condition (Cond F), subjects were required to rate each item and every repetition of an item as to its frequency of usage in printed English. In the second condition (Cond I), subjects were to rate each item and all repetitions of an item as to the ease with which they could form an image of the object or idea named by the word. It seemed reasonable to expect that, when the second occurrence of an MP item appeared, subjects would recognize the item as a repetition and would simply repeat their first rating of it without repeating the processing they had done to arrive at that rating. Upon the second occurrence of a DP item, however, it was assumed that, even if subjects did recognize the item as a repetition and try to recall the rating they had given the prior occurrence of it, this task would be very difficult due to the interference of rating the intervening items. Thus, it was expected that subjects would choose to process the item again in deciding on a rating for the item. If subjects did process MP and DP items along the lines just described, this would result in a processing deficit for MP items, so a spacing effect was expected to be present in Cond F and Cond I.

The third condition (Cond FI) required that subjects use both the frequency and imagery scales. Half of the once-presented items were rated on one scale; the other half were rated on the other scale. For MP and DP items, one of the scales was used for rating the first occurrence of a given item, with the other scale being used for rating the second occurrence of that item. Note that subjects in Cond FI never rated an item on the same scale twice. When the second occurrence of an item under MP appeared, the fact that the item had just been rated would be irrelevant to its present rating because a different scale was involved. The imagery and frequency scales were selected specifically because it was thought that the processing required to do the ratings would be quite different for the two scales. Also, care was taken in the selection of items, such that there was essentially a zero correlation across items between the normative rating of an item on the imagery scale and the normative rating of that item on the frequency scale. Therefore, it was expected that subjects would be forced to process each occurrence of an MP item. There was no reason to expect that subjects would not process each occurrence of a DP item. Insofar as there was no reason to expect a processing deficit under MP in Cond FI, it was expected that the spacing effect would be eliminated.

## Method

**Materials.** The study list presented to each subject consisted of eight once-presented items, eight items presented twice under MP (MP-2 items), and eight items presented twice under DP (DP-2 items). In addition, each study list contained a primacy

Table 3  
Mean Number of Items Recalled as a Function of Type of Presentation, Input Instructions, and Replication in Experiment II

		Once- Presented	MP-2	DP-2
Cond I	Replication A	2.38	2.67	3.92
	Replication B	2.50	2.67	3.96
	Overall Mean	2.44	2.67	3.94
Cond F	Replication A	2.00	2.75	3.96
	Replication B	2.13	2.79	4.25
	Overall Mean	2.06	2.77	4.16
Cond FI	Replication A	1.92	2.42	3.13
	Replication B	2.08	3.08	3.75
	Overall Mean	2.00	2.75	3.44
Cond FR	Replication A	2.96	3.25	4.63
	Replication B	2.46	3.46	4.46
	Overall Mean	2.71	3.36	4.54

Note—Maximum possible recall in any one cell is eight.

buffer of four words, including one MP-2 and one DP-2 item, and a recency buffer of four words. The same items were used for the primacy and recency buffers for all forms of the study list, and recall of these items will not be discussed. The two occurrences of DP-2 items were separated by a lag of either four or five other items, and the position of last occurrence of MP-2 and DP-2 items was controlled. The body of the study list was divided into four equal-sized sections, with each containing two once-presented items, two MP-2 items, and two DP-2 items.

There were six forms of the study list. One set of 24 words was used for three of these forms, and a second set of 24 words was used for the other three forms. Both sets of 24 words were drawn from the Paivio, Yuille, and Madigan (1968) norms. In each set of 24, 6 of the words were high imagery and high frequency, 6 were high imagery and low frequency, 6 were low imagery and high frequency, and 6 were low imagery and low frequency. High-imagery items were defined as those with ratings of 6 or greater on the I scale; low-imagery items were those with ratings less than 4. High-frequency items were those with Thorndike-Lorge frequencies of A or AA; low-frequency items were those with frequencies less than or equal to 20. The first form of the study list was prepared using the first set of 24 words, with two of the items from each imagery-frequency combination randomly assigned to serve as once-presented items, two as MP-2 items, and two as DP-2 items. One word of each imagery-frequency type was presented in the first half of the study list, and the other word of that type was presented in the second half. Two additional forms of the study list were then prepared using the same 24 words, such that, across the three forms, each word served as a once-presented, MP-2, and DP-2 item once and only once. The remaining three forms of the study list were constructed in a similar fashion using the second set of 24 words.

The rating sheet given to subjects in Cond F consisted of 50 numbered blanks along with a 9-point scale at the top of the page, with 1 marked low frequency and 9 marked high frequency. Subjects in Cond I were given a similar sheet with an imagery rather than a frequency scale. In Cond FI, both the frequency and imagery scales appeared at the top of the rating sheet. The particular scale to be used for each item was indicated by a parenthesized F or I next to each numbered blank. Half of the once-presented items were rated on each scale. Equal numbers of MP-2 and DP-2 items were scaled on imagery first and then on frequency, and vice versa. A second form of the FI rating sheet was prepared with the scale used for once-presented

items and the order of the scales for MP-2 and DP-2 items being reversed. Each of these forms of the rating sheet was used for half of the subjects in Cond FI.

**Design and subjects.** Two replications of this experiment were done, one at each of two universities. Hereafter, these will be referred to as Replication A and Replication B. In each replication, 24 undergraduate students were assigned to each of the four instructional conditions (Cond F, Cond I, Cond FI, and Cond FR), for a total of 96 subjects per replication. Within each condition, four subjects were assigned to each of the six forms of the study list. All subjects participated in the experiment in individual sessions, and all participated in partial fulfillment of a course requirement. Subjects were assigned to conditions according to a block-randomized schedule.

**Procedure.** Prior to the presentation of the study list, subjects were read instructions appropriate for their condition. Those subjects in Cond F, Cond I, and Cond FI were told that the experiment dealt with the study of the scaling of words on various dimensions. The particular scale or scales to be used were then described, and the instructions indicated that some items would be repeated but that subjects were simply to rate each item on the appropriate scale each time it was presented. No effort was made to instruct subjects to try to use the same rating for each of the occurrences of repeated items. Subjects were told that they would have only 5 sec to rate each item. Those subjects in Cond FR were given typical free recall instructions. Study list items were presented one at a time at a 5-sec rate using a memory drum. Following presentation of the study list, all subjects were given a written free recall test, and they were allowed 5 min for recall.

## Results and Discussion

The mean numbers of once-presented, MP-2, and DP-2 items recalled in each replication in each of the four instructional conditions are presented in Table 3. The first point to be made is that recall in Replication A in each of the instructional conditions was very similar to that in Replication B. Next, it can be seen that recall in the three incidental learning conditions was slightly, but consistently, lower than that in Cond FR. The magnitude of these differences, however, is much less than that between the incidental and intentional conditions in Experiment I. Third, the overall recall of MP-2 items was less than that of DP-2 items, and this difference was present in each replication in each instructional condition. Finally, although there was a slight reduction in the size of the spacing effect in Cond FI, this reduction was more a result of the lower recall of DP-2 items in this condition than of an increase in the recall of MP-2 items that had been predicted. Stated succinctly, the manipulation in Cond FI that was designed to eliminate the spacing effect failed to do so.

Recall results for once-presented and for repeated items were analyzed separately. Forms of the study list were included as a variable in both analyses, but effects involving this variable were not evaluated. In the analysis of once-presented items, there was a significant main effect of the instructional variable,  $F(3,144) = 3.74$ ,  $MS_e = .71$ . Post hoc pairwise comparisons among the four means showed that recall in Cond I was no different from that in Cond FR, but recall in these two conditions differed from that in Cond F and Cond FI,

which did not differ from each other. There was no effect of the replication variable, nor was there an interaction of the Replication by Instruction Variables (both  $F_s < 1$ ). The only significant effects in the analysis of recall of repeated items were a main effect of instruction,  $F(3,144) = 6.61$ ,  $MS_e = 1.92$ , and a main effect of the MP-DP variable,  $F(1,144) = 78.02$ ,  $MS_e = 1.54$ . The overall mean recall of MP-2 items was 2.89 and that of DP-2 items was 4.02. The predicted interaction of the MP-DP and instruction variables was not significant statistically,  $F(3,144) = 1.35$ .

The prediction that the spacing effect would be eliminated in Cond FI was based upon the assumption that the activities involved in doing the ratings would be responsible for subsequent recall. Specifically, it was assumed that the act of rating an item would provide access to information which would directly serve as a source of potential retrieval cues at the time of recall. If the rating activities did provide the potential retrieval cues directly, then these cues should have been comparable for MP-2 and DP-2 items in Cond FI. Thus, recall should have been comparable; clearly, it was not. The following analyses were done to determine if there was reason to suspect that the activities involved in doing the ratings were not directly responsible for the subsequent recall within the incidental learning conditions.

An unlikely, yet possible, reason for a lack of relationship between the rating activities and recall might be that subjects were not given sufficient time in the rating task to access the information necessary to do the rating sensibly. This might have been a problem particularly in Cond FI, in which subjects were required to use two rating scales. If subjects were unable to perform the rating task effectively, then there should be a very low correlation between the ratings provided for the items in the present experiment and the normative ratings for those items.

In Cond I, there were four subjects in each replication who rated each item when it served as a once-presented item, four subjects in each replication who rated each item as the first occurrence of an MP-2 item, and four who rated each item as the first occurrence of a DP-2 item. In that there were two different sets of 24 words used to construct the study lists, it was possible to derive six sets of mean imagery ratings given to 24 items in each replication. Thus, 12 different correlations could be determined between the mean imagery rating given an item in the present experiment and the normative imagery rating for that item. The average of these 12 correlations in Cond I was .90. This average correlation (and all other average correlations reported in this paper) was determined using Fisher's  $r$  to  $Z$  transformation (cf. McNemar, 1962). A comparable set of 12 correlations between the mean imagery rating for an item and its normative rating was calculated using the judgments in Cond FI, although

in this case the mean rating of each item was based on the judgments of only two subjects. The average of these 12 correlations was .74.

Comparable procedures were used to determine the relationship between the mean frequency ratings for the items in this experiment and the normative frequency ratings for these items. The average of the 12 correlations in Cond F was .70 and the corresponding value in Cond FI was .47. Because high-frequency items were selected as those with Thorndike-Lorge frequencies of A or AA, it was necessary to use the L count for the normative ratings entering into these correlations. It is possible that using this scale contributed to the lower overall correlations for the frequency ratings as compared to those for the imagery ratings. Although the correlations were slightly lower in Cond FI than in the corresponding single rating conditions, the absolute magnitude of the correlations lends little credence to the idea that subjects were unable to do the ratings sensibly in the present experiment.

The preceding analysis indicates that subjects did access information necessary to perform the rating tasks effectively. The next question to be asked is whether this accessed information was directly related to the subsequent recall. The analysis directed toward this question involved an examination of the item difficulty distributions in each of the instructional conditions. It could be argued that, in the act of performing the rating tasks, subjects in Cond F and in Cond I were required to access different and uncorrelated information about the study list items. If the item difficulty distributions in the two conditions are uncorrelated, then it could be argued that the information accessed in doing the ratings was critical in determining the subsequent recall.

Before assessing the degree of consistency across instructional conditions, some estimate is required of the consistency of the item difficulty distributions within each instructional condition. This was obtained by correlating the number of subjects in Replication A who recalled each item with the number of subjects recalling that item in Replication B. Two such correlations could be computed for each instructional condition, one for each of the two sets of 24 words used to construct forms of the study list. The average of these two correlations in Cond I was .76; in Cond F, .68; in Cond FI, .55; and in Cond FR, .52.

The next step was to determine the consistency of the item difficulty distributions across instructional conditions. The item difficulty distributions within each replication and within each set of 24 items for Cond I were correlated with the corresponding distributions in Cond F. The average of the four correlations of this type was .73—if anything, higher than those obtained within the instructional conditions. The average correlation of the item difficulty distributions in Cond I with those in Cond FI was

.66; the average correlation of Cond F and Cond FI was also .66. The average correlation of the item difficulty distributions in the three incidental learning conditions with those in Cond FR was .46. These findings are clearly inconsistent with the notion that information accessed in doing the ratings is directly responsible for subsequent recall.

### EXPERIMENT III

Experiment III was an extension of Experiment II. Two procedural changes were introduced. First, in order to make it easier for subjects to keep their place on the rating sheet in the incidental learning conditions, study list items were presented aurally rather than visually. This was done primarily for the benefit of subjects in the condition in which more than one scale was to be used. Second, both MP and DP items were presented at frequencies of three, as well as at frequencies of two. This necessitated the use of a third rating scale, and the pleasantness scale was chosen for this purpose.

There were three instructional conditions in Experiment III. In the first condition (same rating condition), subjects rated each item and every repetition of the items on only one scale. One-third of the subjects in the same rating condition rated all the items on the frequency scale; one-third of the subjects used the imagery scale; and one-third used the pleasantness scale. In the second condition (different rating condition), subjects were required to use all three scales, but they never rated the same word on a given scale more than once. In the third condition (free recall condition), subjects were given typical free recall instructions prior to the presentation of the study list. On the basis of the results of Experiment II, it was expected that there would be a spacing effect in each of the instructional conditions. In fact, the principal reason for doing Experiment III was to insure that the findings of Experiment II were not dependent upon the specific procedures used in that experiment.

#### Method

**Materials.** Thirty words were chosen from the Paivio et al. (1968) norms to serve as study list items. The words ranged from values of 2.13 to 6.90 on the I scale. The range of word frequencies represented was from 1 to AA on the Thorndike-Lorge count. The words were also selected to represent the full range on the pleasantness dimension as judged by the experimenter. An effort was made to insure that the position of an item on one of the scales was uncorrelated with its position on the other two scales. An additional 10 words were selected, with 5 designated to serve as a primacy buffer and 5 as a recency buffer.

The 30 critical words were rank ordered on the imagery scale. Each of the five highest imagery words was then randomly assigned to one of five item sets. A similar assignment procedure was used for each subsequent group of five words, resulting in five item sets of six words each, representing approximately an equal range on the imagery scale. Each of these five item sets was then randomly assigned to one of the five critical list functions. One set was used for the once-presented items; one

set was used for MP items presented twice (MP-2); one set for MP items presented three times (MP-3); one for DP items presented twice (DP-2); and one for DP items presented three times (DP-3). Four additional forms of the study list were then constructed, such that across the five forms each item served each list function once and only once.

The primacy buffer of the study list consisted of one of each of the five types of items to appear in the body of the list. The recency buffer contained five once-presented items. The same words served these functions across all five forms of the study list, and the recall of these items will not be discussed. Each half of the body of the study list consisted of three items serving each list function. Each DP item was yoked to an MP item of equal presentation frequency with respect to position of last occurrence in the study list. All DP lags ranged from 6 to 10 list positions. The entire study list consisted of 40 different words and 82 list positions.

The rating sheet for subjects in the same rating condition consisted of 82 numbered blanks. In the different rating condition, the scale to be used for each item was indicated next to each blank. Of the six once-presented items in the body of the study list, two were rated on imagery, two on frequency, and two on pleasantness. One of the items rated on each scale appeared in the first half of the study list; the other appeared in the second half of the list. For the MP-2 and DP-2 items, each of the six possible orders of two scales was used once, with the assignment of a particular order to a particular item being random. Three of the possible orders of the three scales were used twice for the rating of the MP-3 and DP-3 items, with the assignment of a particular order to a particular item again being random.

**Design and subjects.** Thirty undergraduate students participating in the experiment in partial fulfillment of a course requirement were assigned to each of the three instructional conditions (same rating, different rating, and free recall), for a total of 90 subjects. Six subjects in each condition were assigned to each of the five forms of the study list.

**Procedure.** Subjects participated in the experiment in groups of varying size, with all subjects in any given session receiving the same form of the study list. As far as was possible, an equal number of subjects in each session were assigned to each of the three instructional conditions. Prior to list presentation, the subjects were given written instructions appropriate for their condition. Those assigned to the same rating condition were instructed to rate each item and every repetition of the items on either the imagery, frequency, or pleasantness scale. They were provided with a 5-point scale on their instruction sheet, and they were instructed to use 1 for items low on the dimension and 5 for items high on the dimension. Those subjects in the different rating condition were provided with all three scales, and they were instructed to use the scale indicated on the rating sheet for each item. Those subjects in the free recall condition were given typical free recall instructions. Any questions the subjects had were answered privately.

Once all subjects had indicated that they understood their instructions, the study list was read at a 5-sec rate. In order to insure that subjects would not lose their place on the rating sheet, the number of the blank corresponding to each item was read prior to reading each word (e.g., 1-salad, 2-temple, etc.). Subjects in the free recall condition were instructed that they could ignore the numbers. Following the presentation of the last study list item, subjects were given blank sheets for a written free recall test, and they were given 5 min for recall.

#### Results and Discussion

Preliminary analyses were done comparing recall in the three subgroups within the same rating condition, which were differentiated on the basis of the scale used for rating all the items (frequency, imagery, or pleasant-

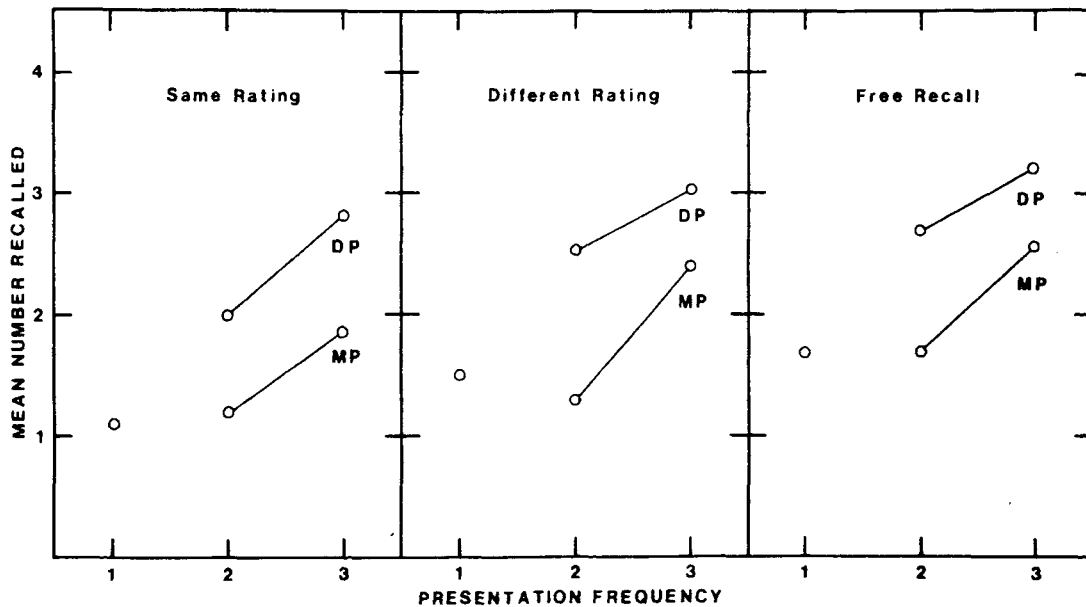


Figure 1. Mean number of words recalled in Experiment III as a function of presentation frequency, MP-DP, and input instructions.

ness). Recall of both once-presented and repeated items was comparable in the three subgroups, so only the results for the same rating condition as a whole will be reported. The mean number of words recalled as a function of presentation frequency, MP-DP, and input instructions is presented in Figure 1. The maximum possible recall for any one type of item was six. Separate analyses were done on the recall of once-presented and of repeated items.

Although the mean number of once-presented items recalled was less in the two incidental learning conditions than in the intentional learning condition, the differences were small and the overall effect of the instructional variable was not significant statistically,  $F(2,75) = 1.70$ ,  $MS_e = 1.46$ ,  $p > .10$ . The analysis of recall of repeated items showed a significant effect of presentation frequency, with the mean recall of items presented twice being 1.90 and that of items presented three times being 2.65,  $F(1,75) = 41.49$ ,  $MS_e = 1.22$ . As is clear in Figure 1, there was a consistent spacing effect in all three instructional conditions and at both presentation frequencies. The overall mean number of MP items recalled was 1.84 and that of DP items was 2.72, and this difference was reliable,  $F(1,75) = 97.53$ ,  $MS_e = .72$ . As was the case in Experiment II, there was no evidence of an interaction of the MP-DP by Instruction Variables,  $F < 1$ .

Correlational analyses similar to those done in Experiment II were done in the present experiment. The mean imagery and frequency ratings given to each item were determined for those subjects who used these scales in the same rating condition and for those items rated on these scales in the different rating condition. A single mean rating was obtained for each item on each scale

using the ratings given that item when it appeared as a once-presented item or as the first occurrence of an MP-2, MP-3, DP-2, or DP-3 item. The correlation of the imagery ratings in the same rating condition with the normative ratings was .95. The corresponding correlation for imagery ratings in the different rating condition was .92. The correlation of the frequency ratings in the same rating condition with the normative frequency ratings was .75, and the corresponding value in the different rating condition was .68. These results, like those of Experiment II, indicate that subjects in both the same rating and different rating conditions were able to access information necessary to do the rating tasks effectively.

As was done in Experiment II, the item difficulty distributions were compared across the instructional conditions. The total number of subjects recalling each item in the three subgroups of the same rating condition was determined. The average correlation among these three distributions of item difficulty was .53. An item difficulty distribution was determined for the different rating condition and for the free recall condition. The average correlation for the three item difficulty distributions of the same rating condition with the distribution from the different rating condition was .70. The average of these three distributions with the distribution from the free recall condition was .61. These results are very similar to those of Experiment II.

## GENERAL DISCUSSION

The following discussion will be limited to a consideration of the results of Experiments II and III. In both these experiments, an attempt was made to



eliminate the processing deficit under MP that is assumed to be responsible for the spacing effect according to the attenuation of attention hypothesis. The ineffectiveness of this manipulation could be taken as evidence favoring a nonvolitional processing deficit explanation of the spacing effect, such as the habituation explanation proposed by Hintzman (1974). As was mentioned in the discussion of Experiment I, acceptance of the persistence of the spacing effect as support for a nonvolitional processing deficit explanation requires that the manipulation result in control of any volitional processing that could be responsible for subsequent recall. If for no other reason than to demonstrate the difficulty of controlling volitional processing activities even in an incidental learning situation, a possible mechanism to account for recall following semantic rating tasks will now be described.

As Postman (1975) has pointed out, "the semantic tasks call for subjective ratings which cannot be either right or wrong." One consequence of this seems to be that, in order for subjects to use the semantic rating scales reliably, they have to compare the item to be rated with other previously rated items. This process is comparable to the process of displaced rehearsals under intentional learning conditions. One consequence of the comparison process may be the formation of incidentally acquired associations among list items, and it may be these associations that are responsible for the subsequent recall. It would seem reasonable to expect that the items that would occur to subjects for purposes of comparison with a given item would be items that are preexperimentally related to the given item. That is, in assigning a rating to the word "table," a likely candidate for comparison would be "chair," particularly if "chair" had appeared previously in the list.

Wallace (1970) has argued that the contiguous occurrence of items during study may be sufficient to result in clustering of these items in subsequent recall. Thus, the comparison process might be expected to lead to the clustering of associatively related items, and this is, in fact, what has consistently been found (Hyde & Jenkins, 1973). It would also seem reasonable to argue that the search for and utilization of preexperimental associations is a process involved in intentional learning situations. Insofar as the semantic rating tasks activate a comparison process that results in a similar utilization of these same preexperimentally established relationships, the finding of correlated item difficulty distributions in recall following semantic rating tasks and following intentional learning is understandable. In fact, it would be expected that a nonsemantic task involving subjective ratings would also activate the comparison process and lead to recall comparable to that under intentional learning instructions. Support for this prediction can be found in an experiment done by Eagle and Mulliken (1974), in which subjects required to rate the pleasantness of the sound of words recalled as

many words as did subjects given intentional learning instructions.

There is a source of comparison that is available for repeated items that is not available for items presented only once. A repeated item can be compared not only to other list items but also to any prior occurrences of that same item. This attempt to reinstate the processing done on previous occurrences of the item may be responsible for the beneficial effects of repetition. The process may be comparable to what Hintzman, Summers, and Block (1975) have called a study-phase retrieval or to what Jacoby (1974) has implemented with his "looking back" instructions. There would be no need for subjects to try to reinstate prior processing of items under MP, but the reinstatement process could very well be involved in the rating of DP repetitions, and this could lead to the spacing effect in recall.

One final source for the spacing effect in recall following semantic rating tasks should be mentioned. It is possible that recall of DP items exceeds that of MP items because of processing activities that occur after these items have been rated. That is, DP items are more likely to serve as comparison items for subsequently rated words. The counterpart of this in the intentional learning situation would be displaced rehearsals of an item occurring after its last list presentation. Rundus (1971) has shown that DP items do receive more displaced rehearsals after their last presentation in the study list than do MP items. If DP items are involved in comparisons with a greater number of subsequently presented items than are MP items, this greater opportunity for incidentally acquired associations could lead to a spacing effect in recall. Further, the nature of the rating done on the various occurrences of the MP and DP items themselves would have little effect on whether or not any one item would be compared to any other item. Thus, the size of the spacing effect would be expected to be, and was, the same regardless of whether the same or different scales were used for each occurrence of MP and DP items.

#### REFERENCE NOTE

1. Zechmeister, E. B., & Shaughnessy, J. J. *When you know that you know and when you think that you know but you don't*. Paper presented at the 15th Annual Meeting of the Psychonomic Society, Boston, November 1974.

#### REFERENCES

- EAGLE, M. N., & MULLIKEN, S. The role of affective ratings in intentional and incidental learning. *American Journal of Psychology*, 1974, **87**, 409-423.
- HINTZMAN, D. L. Theoretical implications of the spacing effect. In R. L. Solso (Ed.), *Theories of cognitive psychology: The Loyola symposium*. Potomac, Md: Lawrence Erlbaum Associates, 1974.
- HINTZMAN, D. L., SUMMERS, J. J., & BLOCK, R. A. Spacing judgments as an index of study-phase retrieval. *Journal of*



- Experimental Psychology: Human Learning and Memory*, 1975, **1**, 31-40.
- HYDE, T. S., & JENKINS, J. J. Recall for words as a function of semantic, graphic, and syntactic orienting tasks. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 471-480.
- JACOBY, L. L. The role of mental contiguity in memory: Registration and retrieval effects. *Journal of Verbal Learning and Verbal Behavior*, 1974, **13**, 483-496.
- MCNEMAR, Q. *Psychological statistics* (3rd Ed.). New York: Wiley, 1962.
- PAIVIO, A., YUILLE, J. C., & MADIGAN, S. A. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph Supplement*, 1968, **76**(1, Part 2).
- POSTMAN, L. Verbal learning and memory. *Annual Review of Psychology*, 1975, **26**, 291-335.
- RUNDUS, D. Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology*, 1971, **89**, 63-77.
- 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.
- WALLACE, W. P. Consistency of emission order in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 58-68.

(Received for publication July 14, 1975;  
revision accepted October 28, 1975.)