# COMPARISON OF OUTPUT ORDER IN IMMEDIATE AND DELAYED FREE RECALL

472
1226-5600

by

LUIS MORALES FLORES JR.

B.A., University of the Philippines, 1966

---

A MASTER'S THESIS

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Psychology

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1973

Approved by:

Sam C. Brown
Major Professor

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

# LIST OF FIGURES

# INTRODUCTION

## Priority in Free Recall

The experimental procedure of free recall has frequently been used to study memory processes. The procedure involves presenting the subject (S) with a list of verbal units (e.g., words, nonsense syllables, anagrams, etc.), then asking S to recall the list (verbally or in writing) in any order he chooses. That S is free to determine the order that items are output from memory differentiates the method of free recall from other experimental procedures. This unique characteristic allows for an evaluation of the relationship between the order that items are presented and the order they are recalled, which is the major basis for investigating organizational processes or strategies that Ss employ during learning (Tulving, 1968).

An increasing number of studies have focused attention on the determinants of output order during recall. One such determinant is item strength defined in terms of probability of recall. For a given item (relative to other items in the list), the greater the probability of recall for that item, the greater its strength. Item strength or probability of recall has been shown to be dependent on at least two factors: (1) serial position, i.e., the position an item occupies in the list during presentation, and (2) degree of learning, i.e., the number of times an item had been previously recalled. Studies attempting to relate item strength and output order have come to different and apparently conflicting conclusions depending on which determinant of item strength is considered.

Previous studies that have focused on serial position have concluded that items are recalled in terms of decreasing item strength. In single trial free recall (a trial consisting of one presentation of the list and one attempt at recall), Murdock (1962) has shown that items appearing at the end of the list during presentation have the highest probability of correct recall. Items falling in beginning positions have the next highest probability of recall, followed lastly by middle items. The higher probability of recall for beginning and end items are referred to respectively as the primacy and recency effects of free recall learning. Typically, end items are recalled first, followed by beginning items and lastly by middle items, an output order reflecting decreasing item strength. This relationship between output order and serial position has been demonstrated in a number of investigations (e.g., Deese, 1957; Deese & Kaufman, 1957; Bousfield, Whitmarsh & Esterton, 1958).

The second factor known to influence item strength is the degree to which an item has been learned. Underwood (1964) has shown that items recalled correctly on one or more previous trials (hereafter referred to as old items) have a higher probability of recall than items not previously recalled correctly (hereafter referred to as new items). If items are output in decreasing strength we should expect old items to be emitted earlier than new items in the recall sequence. This is not the case, however. In a series of experiments, Battig, Allen and Jensen (1965) showed that new items were emitted before old items. This phenomenon, referred to as the priority of recall of new items (PRNI), has been demonstrated in subsequent investigations (e.g., Mandler & Griffith, 1969; Roberts, 1969; Brown & Thompson, 1971), and suggests that items are output from memory in terms of increasing strength. Battig et al. (1965) explained their findings in terms of a strategy whereby during input Ss presumably pay special attention to items not previously recalled. Then at recall these new items are output earlier and before old items to prevent their being lost (or forgotten) due to possible interference from the attempted recall of old items. Tulving and Arbuckle (1963), and Smith, D'Agostino and Reid (1970) have shown that the act of recalling an item interferes with the recall of subsequent items.

It is not surprising that a controversy has arisen because of this contradiction to the principle that items are recalled in terms of decreasing strength.[1] But the contradiction appears to be more apparent than real. Previous investigations which provided evidence to support the increasing strength notion looked only at serial position. On the other hand, investigations which provided evidence to support the increasing strength notion looked only at degree of learning. Brown and Thompson (1971) considered both factors simultaneously in the learning of a single list. Their results showed that (a) regardless of serial position new items were emitted before old, and (b) end items were emitted before middle items.[2] More specifically, the relationship generated between item strength and output position was curvilinear. New end items which had intermediate strength as measured by probability of correct recall were emitted first in recall, followed by new middle items with lower item strength and finally by old items with the highest item strength. The curvilinear relationship was even more pronounced when the location of overt errors in the recall sequence was examined. Overt errors which were assumed to have minimal or the lowest item strength were the very last items emitted. Like Battig et al. (1965), Brown and Thompson (1971) interpreted their findings in terms of interference theory:

> . . . it is proposed that Ss adopt a minimal-interference strategy whereby recall of the list is maximized by recalling items in an order which minimizes output interference and forgetting. More specifically, Ss recall newly acquired items first because these are the ones most susceptible to output interference and forgetting. The finding that new items in end positions are usually recalled before new items in middle positions can probably be attributed to the greater availability of the new end items owing to the fact that these were the last items seen. . . . Old items can easily be postponed until new items have been recalled because they are better learned than new items and therefore more resistant to output interference and forgetting. Finally, the occurrence of errors late in recall indicates that Ss postpone until the very last attempts to retrieve items for which they are least certain. It seems reasonable to assume that Ss spend a considerable amount of time thinking about these difficult items. If Ss were to take this much time at an earlier point in the recall sequence, they might forget some other items which otherwise would be recalled correctly (p. 447).

The Brown and Thompson (1971) study has clarified the contribution of the many variables that influence output order in free recall learning. However, other studies similarly attempting to answer the question of how verbal units are recalled from memory, but involving different procedures, have failed to find evidence of a priority strategy. Free responding studies are a case in point. Here, for example, Ss are asked to recall as many members of a taxonomic class (e.g., names of animals) as they can remember, or to free associate to single stimulus words. These studies have shown that words most frequently recalled by the group (i.e., high item strength words), are also those which tend to be emitted first and with a shorter latency of emission than less frequently recalled words (see Osgood, 1953); response latency has also been used as an index of item strength (e.g., Bousefield & Sedgewick, 1944; Underwood & Schulz, 1960).

One purpose of the present experiment was to specify more precisely the conditions under which the priority effect occurs. One major difference between free recall learning and other recall situations is that in free recall Ss are shown the items to be remembered while in free responding situations they are not. Thus, the free responding situation is more similar to everyday situations where the specific to-be-recalled material is not presented to S before attempted recall from memory. A comparable free responding situation can be created in a free recall experiment by adding a final free recall (FFR) test. Suppose Ss were successively presented with different lists of words using the typical free recall procedure and then asked to recall all words from all lists without again seeing the words. How would items be emitted during FFR? If the priority strategy is a long term memory phenomenon we would expect Ss to recall (or attempt to recall) items in some manner reflecting increasing item strength. On the other hand, the PRNI effect may be specific only to the acquisition of verbal material and may not occur during FFR. If the latter were the case we would expect final recall to reflect output in terms of decreasing strength. Thus, the present study would enable us to determine whether Ss retrieve information from memory in the same or different ways when the characteristics or demands of the task change.

A second purpose of the present experiment was to determine whether the PRNI effect occurs in the free recall learning of categorized lists of words (i.e., lists composed of subsets of exemplars, each classifiable in terms of a common label or name). Perhaps no other organizational phenomenon has received as much attention as clustering which is the tendency of conceptually related items to be recalled adjacent to one another under conditions where exemplars from different categories are randomly presented for study. Of special interest is the manner in which organization (i.e., clustering) develops as learning progresses. Previous investigations have shown that clustering increases with successive trials (Shuell, 1969). Why is it that Ss do not show perfect clustering during early trials? The answer may be that during early trials a considerable number of items lack sufficient strength to readily lend themselves to an organizational scheme based mainly on conceptual relatedness. Clustering requires that many of the newly acquired items be postponed in recall. It may be that Ss initially emit items in terms of item strength (i.e., new items first) in order to reduce output interference. Then as the items become better learned and more resistant to output interference they are incorporated in conceptually related clusters during recall. Thus, another purpose of the present experiment was to specify the relationship between degree of learning (i.e., item strength) and the development of organization (clustering) across trials.

## Comparison of Measures of Priority

Since a major concern of the present investigation was the specification of the relationship between order of free recall and item strength, this section concerns itself with an evaluation of currently used measures of priority and proposes another index of priority in free recall.

The concept of priority implies that new items are emitted early in recall and before old items. Previous investigations have employed basically one of two measures of priority. One is the standard recall rank (SRR) score developed by Battig et al. (1965) and the other an "observed minus expected" (O-E) score introduced by Postman and Keppel (1968) and Shuell and Keppel (1968). Both scores may be computed only when both old and new items occur in a protocol. Before proceeding further, it is important first to define how each measure is derived.

### The $\overline{SRR}$ Score

For a given recall protocol an SRR score for each item is obtained by the following relationship:

$$SRR_i = -\left(\frac{R_i - Mdn_R}{\sigma_R}\right), \quad \text{where,} \qquad (1)^3$$

$R_i$ = item *i*'s output rank position with the item recalled first assigned a rank of 1, the second 2, etc.,

$Mdn_R$ = median output rank for all *i* items, and,

$\sigma_R$ = standard deviation of the total number of ranks.

The SRR score is essentially a z-score in that each item is expressed in terms of its distance from the median in standard deviation units. Items recalled above the median rank take on positive values and those below, negative values. The algebraic mean of the deviations for new items ($\overline{SRR}$) constitutes the index of priority.

The O-E Score

The second measure of priority is the difference between the observed and expected number of new items occurring in some specified segment of the recall protocol, usually the first quarter. If new items occur randomly in different positions of the output sequence then one-fourth of the total new items recalled would be expected to occur in each quarter of the protocol. If the O-E difference is positive then more new items are recalled in the first quarter than expected by chance.

Limitations of the $\overline{SRR}$ and O-E Scores

Both measures suffer from the limitation that they vary with characteristics of recall unrelated to relative amounts of priority. Specifically, the score for maximum positive priority (all new items recalled first) and maximum negative priority (all new items recalled last) changes as a function of the total number of new items recalled (N) and in the case of the O-E score, also with the total number of items recalled (T). Consider the case when all new items are recalled first (i.e., maximum positive priority).[4] The following five hypothetical recall protocols (A-E) are all examples of maximum positive priority. All protocols are of the same length (12) but differ in number of new and old items recalled. The SRR score for each item is given in the last column. The scores are the same for items with identical output ranks because each protocol is of the same length.

| | Output Rank | Recall Protocol A | B | C | D | E | SRR |
|---|---|---|---|---|---|---|---|
| First quarter of recall | 1 | n | n | n | n | n | 1.59 |
| | 2 | n | n | n | n | o | 1.30 |
| | 3 | n | n | n | o | o | 1.01 |
| Second quarter of recall | 4 | n | n | o | o | o | .72 |
| | 5 | n | o | o | o | o | .43 |
| | 6 | o | o | o | o | o | .14 |
| Third and fourth quarters of recall | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| | 12 | o | o | o | o | o | -1.59 |
| | O-E | 1.75 | 2.00 | 2.25 | 1.50 | 0.75 | |
| | $\overline{SRR}$ | 1.01 | 1.16 | 1.30 | 1.44 | 1.59 | |

The size of a quarter of recall for all protocols is three ($Q = 3$), and the number of new items recalled ranges from one to five. As can be seen from the bottom two rows, both $\overline{SRR}$ and O-E scores for new items yield different values for protocols of varying numbers of new items recalled even when all of the new items are emitted before old items. In the case of $\overline{SRR}$, the scores increase as N decreases. This occurs because $\overline{SRR}$ is the average of the SRR scores for all new items. Decreasing the number of new items while maintaining maximum positive priority serves to eliminate lower SRR scores and therefore results in a larger $\overline{SRR}$ score. Parenthetically, it is important to note that the decrease in the number of new items as one goes from Protocol A to E is not unlike what would be expected to occur over the normal course of learning. Previous investigations (e.g., Battig et al., 1965; Brown & Thompson, 1971) using the $\overline{SRR}$ measure have reported an increase in priority over trials. Clearly, such an increase could be an artifact of the $\overline{SRR}$ score.

The relationship between O-E scores and number of new items recalled is different. Here, the score for maximum positive priority will be highest when $N = Q$ (Protocol C). Increasing N beyond Q (Protocols A and B), or decreasing N beyond Q (Protocolc C, D and E), yields progressively smaller O-E scores. When $N \leq Q$ it is possible to adjust for differences in number of new items recalled by dividing the O-E score by N. However, the not infrequent case where $N > Q$ cannot be adjusted in this manner.[5]

Recall protocols of different lengths is not a problem for the $\overline{SRR}$ but is for the O-E measure. Output protocols with the same number of new items but of different lengths will yield the same $\overline{SRR}$ score. This is a property of the z-score transformation. In the case of the O-E measure the picture is more complex. Holding N constant and varying T will yield the same O-E score only if $N \leq Q$ in each protocol. To illustrate, consider Protocols F and G:

| | Quarter<br>1 2 3 4 | | | |
|---|---|---|---|---|
| Protocol F | nn/oo/oo/oo/ | Q = 2, | N = 2, | E = .5 |
| Protocol G | nno/ooo/ooo/ooo/ | Q = 3, | N = 2, | E = .5 |

In each instance $N \leq Q$. Hence, both protocols produce the same O-E score of 1.50. Now consider Protocols H and I:

| | Quarter<br>1 2 3 4 | | | |
|---|---|---|---|---|
| Protocol H | nn/no/oo/oo/ | Q = 2, | N = 3, | E = .75 |
| Protocol I | nnn/ooo/ooo/ooo/ | Q = 3, | N = 3, | E = .75 |

In this situation $N > Q$ for Protocol H. The O-E score for Protocol H is 1.25, while for Protocol I it is 2.25. Thus, the O-E scores for maximum positive prority will be the same for protocols of different lengths, if for all protocols $N \leq Q$. When $N > Q$, O-E scores will differ from one another in a manner related to the magnitude of the difference between N and Q: the greater the difference between N and Q, the lower the O-E scores.

<u>The Relative Priority (RP) Score</u>

To overcome the shortcomings of the $\overline{SRR}$ and O-E measures we have developed a relative measure of priority (RP) where, for any given protocol, maximum positive priority is set at plus one, maximum negative priority at minus one and no priority at zero. (No priority is defined as the case where the sum of the algebraic ranks of new items from the median output rank equals zero.) In other words, RP interpolates (by linear transformation) scales based on protocols with different values of N and T to a common scale that ranges from 1 to -1. Hence, the RP score is invariant with respect to factors unrelated to relative amounts of priority (i.e., N and T). The formula for the RP score is as follows:

$$RP = 1 - \frac{2(OP - MaxP)}{MaxN - MaxP}, \quad \text{where } 0 < N < T, \text{ and,} \tag{2}$$

$OP = \sum R_i$ = observed priority (OP) is the sum of the rank output positions of new items with 1 being assigned the first item recalled, 2 the second item recalled, etc.,

$MaxP = \sum_{i=1}^{N} R_i$ = sum of the rank positions if all new items were recalled first,

$$MaxN = \sum_{i=T-N+1}^{T} R_i \quad = \text{sum of rank positions if all new items were recalled last,}$$

N = total number of new (n) items recalled, and

T = total number of items recalled.

Applying formula (2) to recall Protocols A-E and F-I will yield in each instance an RP score of 1. Had all new items in each protocol been recalled last, the RP score would be -1. For ease of calculation the following computational formula may be used:

$$RP = 1 - \frac{2OP - N(N + 1)}{N(T - N)} \quad , \text{ where } 0 < N < T, \text{ and,} \qquad (3)$$

OP, N and T, are defined as in (2).

It may be noted that for no priority, both $\overline{SRR}$ and RP give identical scores. However, the closer to maximum positive or maximum negative priority a protocol is, the greater is $\overline{SRR}$ biased by different numbers of new items. The amount of bias reflected by O-E is difficult to specify because O-E varies with both number of new items and total number of items recalled. The amount of bias will depend largely on the extent that N is greater than Q.

<u>Output Position of a Single Item</u>

A limitation of the RP measure is that it does not provide a score to describe the output position of an individual item, as does the SRR measure. When this is of major concern to the investigator an equivalent output position score for individual items (RPI) is given by:

$$RPI_i = 1 - \frac{2(R_i - 1)}{(T - 1)} \quad , \text{ where } i = 1,2,3,\ldots,T, \text{ and,} \qquad (4)$$

$R_i$ = item *i*'s output rank position.

Formula (4) has all the properties of SRR with the added advantage of having fixed upper and lower bounds equal to 1 and -1 respectively, for protocols with any values of N and T (where $0 < N < T$). Thus, when item *i* has an RPI score of 1, this indicates that it was the first item emitted. An RPI score of -1 indicates it was the last item emitted, etc.

Individual priority scores may be useful, for example, when the output position of new items presented in recency positions is of concern. One could entertain the hypothesis that new items presented in recency positions during study will be output before other new items (e.g., Brown & Thompson, 1971). For each protocol, a mean RPI score could be calculated for new items presented in recency positions and compared with the mean RPI score for other new items. This procedure, however, may result in the same biases as discussed in relation to $\overline{SRR}$: RPI scores averaged in this manner do not express the actual amount of priority as a proportion of the total priority possible. To avoid these biases RP scores could be computed separately for new items presented in recency positions and new items presented in non-recency positions.[6]

Meaning of a Single RP Score

The problems involved in the interpretation of the RP score generally applies to the interpretation of any statistic. Consider, for example, Protocols J and K. Both protocols contain four new items,

| | Output Rank Position | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | |
| Protocol J | o | n | o | n | o | n | o | n | o | RP = 0 |
| Protocol K | n | n | o | o | o | o | o | n | n | RP = 0 |

the ranks of which sum to 20. Both, therefore, yield an RP score of zero. However, the distribution of new items in each protocol is different. In Protocol J, new items are distributed evenly in the recall sequence, while in Protocol K they are grouped at the beginning and end of the output sequence. Clearly, the RP score does not differentiate between these two occurrences. It therefore behooves the investigator to examine individual protocols for systematic patterns of recall such as those in Protocol K. If, for example, the data indicate that *S*s are emitting some new items at the beginning and others at the end of the output sequence, then separate RP scores for the first and second half of recall may be computed. Other cases may of course warrant different applications of RP.

## Conclusion

It has been shown that the RP score is free of the limitations of previous indices and provides an uncontaminated measure of relative amounts of priority in free recall. However, the RP measure may be applied to any type of item. It should be emphasized that the RP measure provides *only* an empirical index of the relative output location of a subset of items in recall. It does not specify the mechanism or subject strategies underlying the recall order. Nonetheless, we believe that such specification will rest ultimately upon the use of a measure which accurately describes the phenomenon to be explained.

# METHOD

## Subjects

Forty-one male and female students enrolled in Introductory Psychology courses at Kansas State University served as Ss. They were tested in small groups ranging from two to nine Ss in size. The Ss received course credit for their cooperation and were not necessarily naive to verbal learning experiments. The data of two Ss were discarded because of failure to follow instructions.

## Groups and Procedure

Two groups of 18 and 21 Ss learned in succession four categorized lists of words and four non-categorized lists, respectively. The Ss in both groups were otherwise treated identically. Each list was practiced for four trials. The study portion of each trial consisted of the individual presentation of the words at a 1.5 second rate via a Kodak Carousel projector. Following study there was a 30 second (unfilled) interval before Ss were asked to write down as many words of the list as they could remember. Booklets were provided for this purpose and Ss were instructed to use a separate page for each trial. The Ss were told to begin writing at the top of the page and to write the words down in the order they came to mind. Two minutes were allowed for recall, after which the next trial was immediately begun. The items in each list were presented in a different randomized order on every trial, and the order of list presentation was counterbalanced (i.e., for each group, a different sequence of the four lists was used such that each list appeared once as a first, second, third and fourth list).

Immediately following the learning of the fourth list, Ss were given an FFR test in which they were asked to write down, in any order they wished, as many of the words they could remember from the four lists previously shown them. Ten minutes were allowed for completion of FFR. The Ss were not told at the beginning of the experiment the number of lists that would be presented, nor of the FFR test.

Materials

Non-categorized Lists -- The stimulus materials were 112 unrelated common nouns (e.g., cheek, attitude, market, lad, milk, etc.) rated A or AA in the Thorndike-Lorge (1944) word-frequency count. Four different lists of 28 words were constructed by assigning randomly one word to each list.

Categorized Lists -- Sixteen categories were selected from the Battig-Montague (1969) norms. Seven words (nouns) were selected from each category on the basis that mean rated frequency be approximately equivalent across categories. Words that could be obviously classified under more than one category heading were not used. From these materials four nonoverlapping lists, composed of four categories each, were constructed. Appendix I lists the categorized and non-categorized lists. (Equivalence in mean rated frequency between categorized and non-categorized lists was not possible. Only eight or nine words out of the 28 that made up each categorized list had Thorndike-Lorge ratings of A or AA. However, words with the highest rating possible were chosen to make up the rest of the categorized lists.)

# RESULTS

## Acquisition

For each S a tabulation was made of the number of words recalled correctly on each trial of free recall learning. Means were then computed separately for Ss learning categorized and non-categorized lists. The results are shown in the left portion of Figure 1. In all but one instance (List 1, Trial 3), Ss who learned categorized lists recalled more words than Ss who learned non-categorized lists. Pooled across lists and trials the Categorized group recalled a mean of 345.2 words compared to 293.5 for the Non-categorized group. This difference was significant beyond the .05 level; $F(1,37) = 12.91$, the criterion of significance for all analyses to be reported.

As expected, both groups showed an overall significant increase in number of words recalled across trials, $F(3,111) = 425.74$. Averaged across groups and lists the means for Trials 1-4 were 14.1, 19.9, 22.3 and 23.5, respectively. However, the superiority of the Categorized over the Non-categorized group decreased across trials within the learning of each list as evidenced by a significant Trials X Groups interaction, $F(3,111) = 7.60$, and a nonsignificant Trials X Groups X Lists interaction ($F < 1$). Pooled over lists the difference between the two groups in number of words recalled correctly on Trials 1-4 were 4.4, 4.0, 2.2 and 2.2, respectively.

There also was a significant Groups X Lists interaction, $F(3,111) = 5.00$, attributable to a slight decrease in words recalled across

THIS BOOK CONTAINS NUMEROUS PAGES WITH DIAGRAMS THAT ARE CROOKED COMPARED TO THE REST OF THE INFORMATION ON THE PAGE.

THIS IS AS RECEIVED FROM CUSTOMER.

# ILLEGIBLE

# THE FOLLOWING DOCUMENT (S) IS ILLEGIBLE DUE TO THE PRINTING ON THE ORIGINAL BEING CUT OFF

ILLEGIBLE

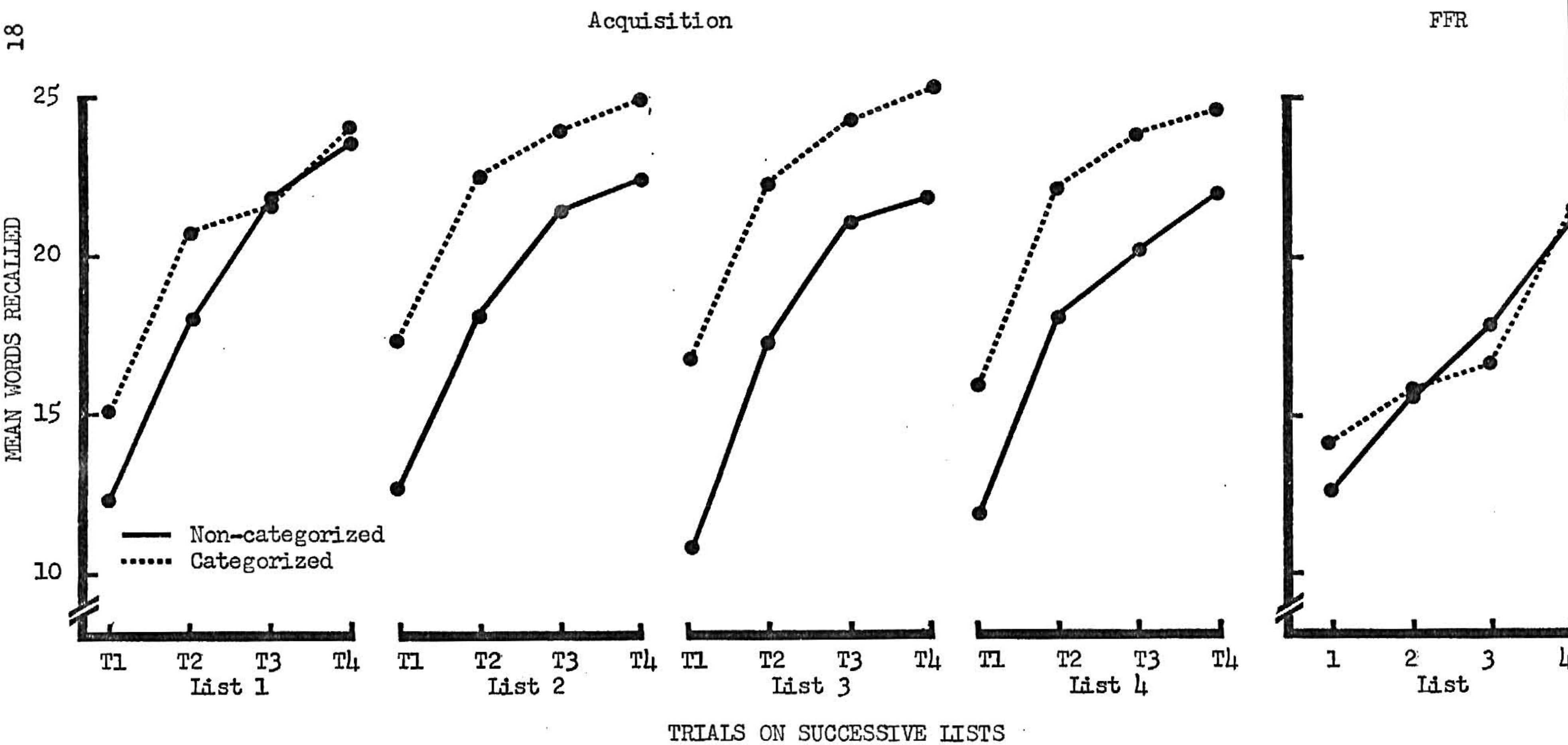


Fig. 1. Mean number of words recalled correctly on each trial (T) from each list during acquisition and from each list during FFR. Data are presented separately for categorized and non-categorized lists.

List 1-4 by the Non-categorized group (19.0, 18.6, 17.8 and 18.0, respectively) but an increase (from List 1 to 2) by the Categorized group (20.4, 22.3, 22.1 and 21.6, respectively). Interestingly enough, the learning-to-learn phenomenon, (i.e., increase in performance from List 1 to List 2), occured only for categorized lists. Both the main effect of Lists and the interaction of Lists with Trials were non-significant ($F$'s $\leq 1.27$).

Final Free Recall (FFR)

During FFR Ss were requested to recall as many words as they could remember from all lists. The right portion of Figure 1 shows the mean number of words recalled from each list, separately for the Categorized and Non-categorized groups. The data clearly show that the closer to FFR a list was studied the more items recalled from that list. The effect of Lists was highly significant, $F(3,111) = 13.71$. However, neither the difference in mean words recalled between the Categorized (17.0) and the Non-categorized groups (16.8), nor the Lists X Groups interaction were significant ($F$'s $< 1$).

It was surprising that during FFR the Categorized group did not recall more words than the Non-categorized group because during acquisition the Categorized group produced more total correct responses. For example, consider only words recalled from List 4 (i.e., the list where the smallest differences occurred between acquisition and FFR). Here we find that the Non-categorized group did not recall significantly more words during Trial 4 of acquisition (21.9) than during FFR (21.0), $F(1,20) = 1.81$. However, the Categorized group recalled significantly fewer words in FFR (21.5) than on Trial 4 (24.5), $F(1,17) = 8.58$. To examine the reasons for such relatively poor FFR performance by the

Categorized group the number of exemplars recalled from each category during FFR and during Trial 4 was determined separately for each list. Averaged across Ss the mean number of exemplars recalled from a category during Trial 4 of acquisition was 6.0, 6.2, 6.3 and 6.1 for Lists 1 to 4, respectively. For FFR the values were 5.4, 5.7, 6.0 and 6.0, respectively. An ANOVA showed that the Categorized group recalled significantly more words per category during Trial 4 (6.2) than during FFR (5.7), $F(1,17) = 11.11$. While there was a slight increase in the mean number of exemplars recalled from a category in FFR from List 1 to 4, the main effect of Lists failed to reach significance, $F(3,51) = 2.53$. Neither was the interaction between Lists and Time of Test (i.e., Trial 4 vs. FFR) significant, $F(3,51) = 1.14$. Thus, these results indicate that the decrease in performance during FFR was at least partly attributable to fewer exemplars recalled per category.

Another factor that accounted for the poorer performance during FFR was number of categories recalled. On Trial 4 of acquisition (for all lists) every S recalled at least one exemplar from all four categories. During FFR the number of categories recalled from List 4 was comparable (3.7) to the number of categories recalled during Trial 4 of acquisition. In fact, only five Ss failed to recall all categories from List 4 during FFR. Hence, the decrement in FFR for List 4 would appear largely the product of a decrease in number of exemplars recalled from each category. However, the number of categories recalled from other lists during FFR was considerably less than four categories. Specifically, the mean number of categories recalled from Lists 1-4 were 2.6, 2.8, 2.8 and 3.7, respectively. An ANOVA comparing these means was highly significant, $F(3,51) = 4.60$. Thus, the two analyses indicate that the decrement in performance of the Categorized group during FFR was attributable both to a decrease in number of exemplars recalled from a category and particularly for Lists 1, 2 and 3, to the number of categories recalled.

Clustering in Categorized Lists

To determine whether clustering increased over trials during acquisition, ARC (adjusted ratio of clustering) scores formulated by Roenker, Thompson and Brown (1971) were computed. The ARC score is a measure of the proportion of actual category repetitions above chance for a given protocol. The mean ARC scores on each trial are given in Figure 2, separately for each of the four lists. As can be seen there was an increase in clustering across trials, but, in agreement with Thompson (1972), the effect was limited largely to List 1 learning. Pooled across lists the ARC scores for Trials 1-4 were .76, .88, .84 and .83, respectively. The main effect of Trials was highly significant, $F(3,51) = 4.48$. Collapsing across trials, the main effect of Lists was also significant, $F(3,51) = 9.96$. The mean ARC scores for List 1-4 were .71, .80, .86 and .86, respectively. These means show an increase in clustering from Lists 1-3. While the interaction of Lists with Trials was not significant, $F(9,153) = 1.43$, a separate test of the interaction between the effect of List 1 against the combined effects of Lists 2 to 4 was significant, $F(3,153) = 2.68$. These results indicate that the increase in clustering across trials and across lists were due largely to List 1 performance.

Priority During Acquisition

To specify the relationship between output order and item strength, four types of items were identified in the recall protocol of each S.

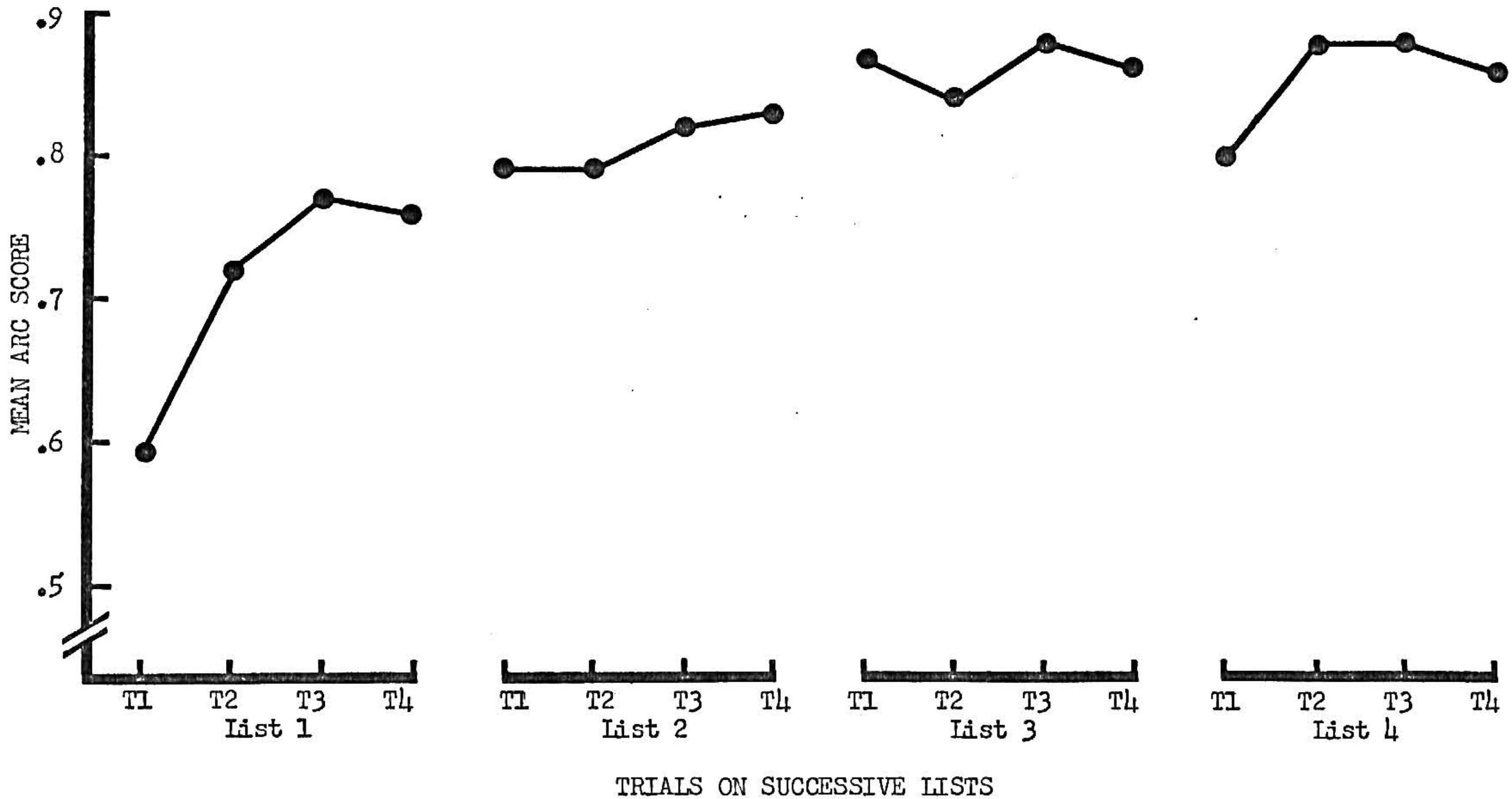


Fig. 2. Mean ARC (clustering) scores on each trial (T) from each categorized list.

These were new items recalled for the first time and old items recalled either once, twice or three times on previous trials. The four types of items can be ordered along a continuum ranging from low item strength (new items) to high item strength (thrice recalled old items). Relative priority (RP) scores for new and for once recalled old items were computed separately for Trials 2, 3 and 4. For twice recalled old items, RP scores were computed on Trials 3 and 4, and on Trial 4 for thrice recalled old items. No RP scores were computed on Trial 1 because by definition all items output during the first trial are new and therefore the concept of priority does not apply. The results are presented in Figure 3, separately for Categorized and Non-categorized groups. As can be seen, new items were emitted before old items in every instance except one (List 3, Trial 4 for the Categorized group). Moreover, there was a tendency on each trial for items to be emitted in terms of increasing strength, although on Trial 4 there were some reversals in the order of emission of twice and thrice recalled old items.

The examination of priority as a function of lists and trials was limited to new items. Because the number of _Ss_' protocols included in the computation of the mean RP score decreased as trials progressed, two separate ANOVA's (analysis of variance) were performed. The first was a Groups X Trials analysis averaged across lists and the second a Groups X Lists analysis averaged across trials. This prevented the elimination of a large proportion of data from the analysis. As anticipated, the Non-categorized group showed more overall priority (.37) than the Categorized group (.20). However, this difference fell slightly short of statistical significance, $\underline{F}(1,29) = 3.63$. Both Categorized and Non-categorized groups showed an overall increase in

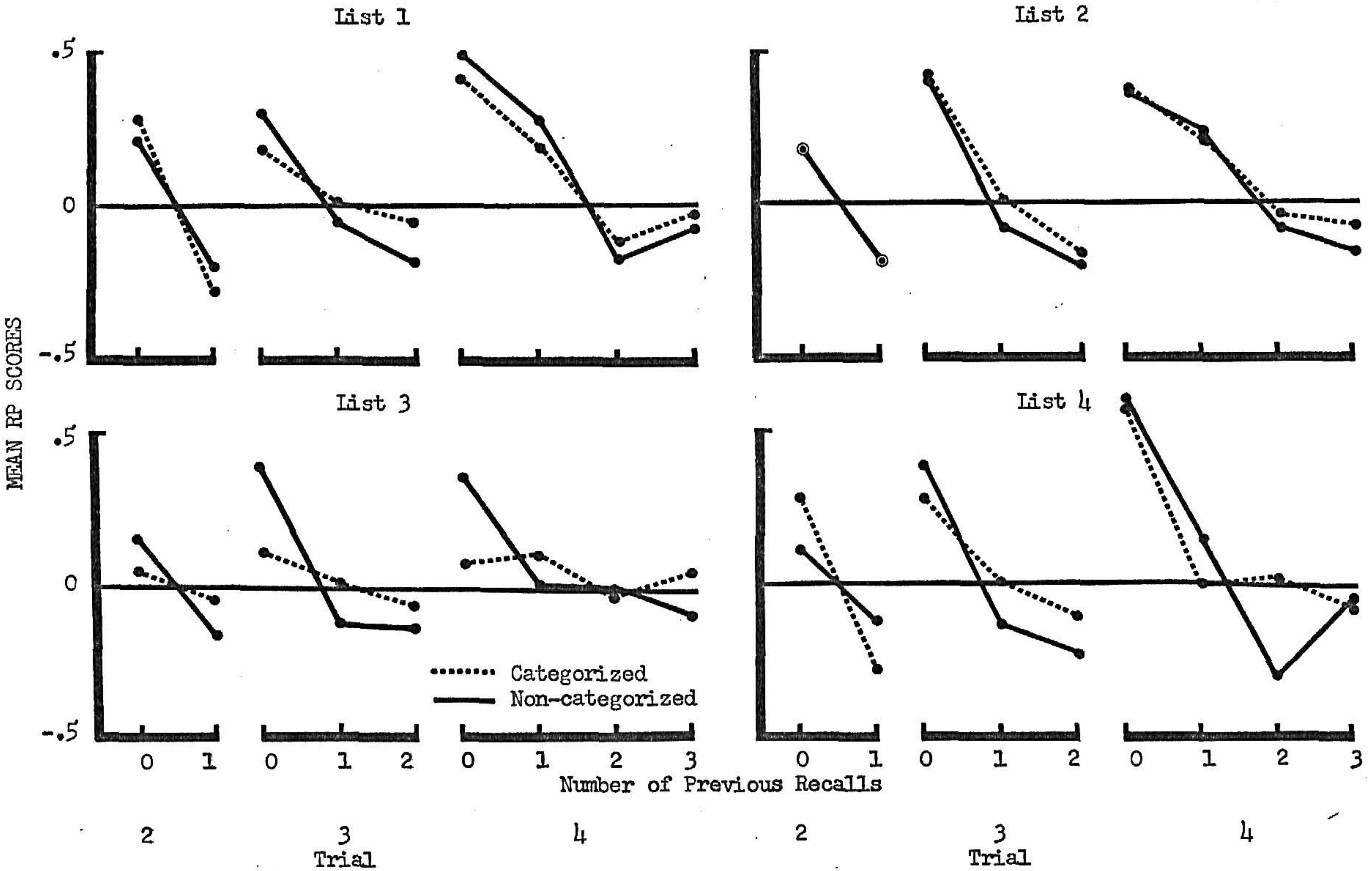


Fig. 3. Order of item recall (RP) during acquisition as a function of number of previous recalls. Data are presented for the categorized and non-categorized lists.

priority across trials, $F(2,58) = 10.99$. Averaged across lists and and groups the means were .15, .30 and .40 for Trials 2, 3 and 4, respectively.[7] The main effect of Lists, the Groups X Trials and the Groups X Lists interaction were not significant ($F$'s $< 1.54$).

Priority in Categorized Lists

One major hypothesis of the present study was that as clustering increases across trials in acquisition priority should decrease. This does not seem to be the case. On List 1 clustering increased from Trial 2-3 and decreased slightly on Trial 4.[8] Priority decreased slightly from Trial 2-3 and increased on Trial 4. On List 2 clustering increased from Trial 2-4, but priority increased (Trials 2-3) then slightly decreased (Trial 4). An examination of the remaining lists shows clustering to remain high and that trial-to-trial changes in priority and clustering did not systematically covary. These results suggest that priority and clustering are essentially independent. The task therefore becomes a matter of explaining how Ss employ seemingly contradictory strategies. An examination of individual protocols showed that one factor was the number of new items recalled. Averaged across lists the mean number of new items recalled on Trials 2, 3 and 4 were 7.9, 3.0 and 1.5 words, respectively. This is in marked contrast to the number of words recalled on Trial 1 (16.3 averaged across lists). The decrease in number of new items on Trials 2, 3 and 4 could enable S to maintain both high levels of clustering and priority. With an average of three new items recalled on Trial 3, for example, an S could easily recall all these new items first and still have a high clustering score. This could occur by clustering new items from the same category together and/or by following the last new item emitted with old items from the same category. Thus, the fewer number of new items (relative to old items) permitted S to have both high priority and clustering scores.

A final point of interest is the order of emission of items within clusters. For purposes of analysis only clusters of three or more words, and containing both old and new items were examined. The proportion of new items occurring in the first half of the cluster was determined (interpolating when necessary), and averaged across different cluster sizes and trials for each S. The mean proportion of new items emitted in the first half of the cluster was .57, .57, .50 and .57 for Lists 1-4, respectively. None of these scores, nor the average across trials (.58), proved to be significantly greater than chance (.50), $F$'s $\leq 3.22$, df = 1/(16 or 17). Thus, there was no evidence to indicate that Ss employed a priority strategy within clusters.

Output Order During FFR

Of major concern to the present experiment was an examination of the output order of items during FFR. Inspection of the FFR data revealed that Ss tended to recall together items from the same list. For both Categorized and Non-categorized groups, the items from List 4 tended to be recalled first, followed by items from Lists 3, 2 and 1 in that order. Because Ss recalled together items from the same list, attention was focused on the order of emission of words belonging to each list. Specifically, for each list, RP scores were computed for new items (those recalled in FFR but not during acquisition) and for old items recalled once, twice, thrice and four times during acquisition, disregarding intervening words recalled from other lists. The results are shown in Figure 4, separately for the Categorized and Non-categorized groups. Data from non-categorized lists will be considered first. In

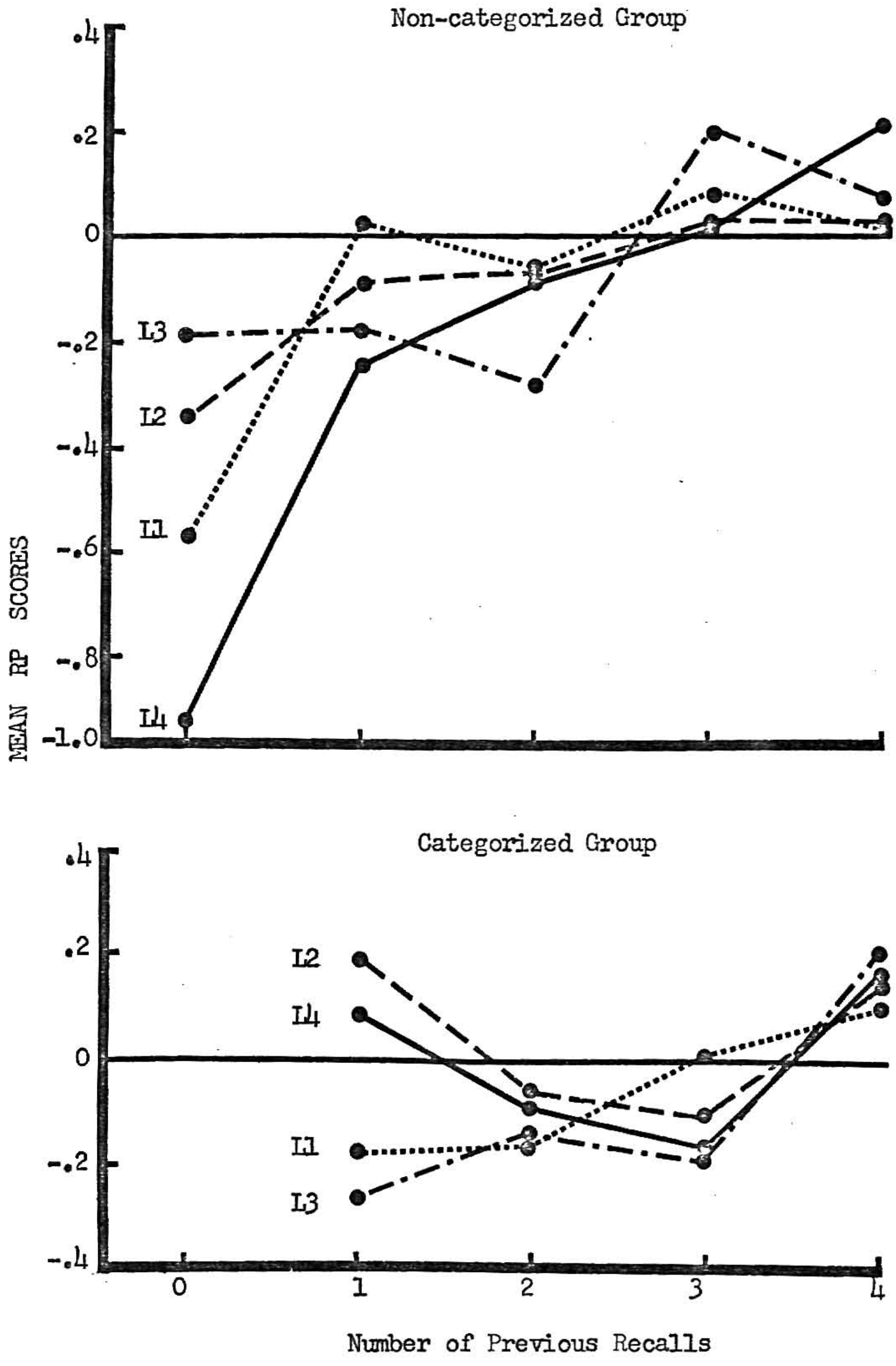


Fig. 4. Order of item recall (RP) in FFR as a function of number of previous recalls during acquisition. Data are presented separately for each list (L).

marked contrast to output order during acquisition, new items in FFR were recalled after and not before old items. Instead of emitting words in terms of increasing strength, Ss recalled in terms of decreasing strength. Items of high item strength (i.e. those recalled thrice and four times in acquisition) were recalled first followed by items of lower strength (i.e., those recalled once and twice previously). New items were emitted last. Collapsing across lists, the mean RP score for new items and for items recalled once, twice, thrice and four times was -.54, -.14, -.13, .09 and .09, respectively. All of these means were significantly greater or less than zero ($t$'s $\geq$ 1.74, df = 12).

Output order during FFR for the Categorized group presents a more complicated picture. To begin with, of the total number of items recalled, only one was a new item. Therefore, new items were excluded in the analysis of FFR. The RP scores for the Categorized group are given in the lower portion of Figure 4. Averaging across lists, items previously recalled in acquisition four times yielded a mean RP score (.16) significantly greater than zero, $t$ = 3.02, df = 72. Items recalled previously thrice yielded an RP score (-.12) significantly less than zero, $t$ = -2.32, df = 72. This order of recall is reflective of decreasing item strength. The data are less clear for once and twice previously recalled items. Items for Lists 1 and 3 continued to show an output order in terms of decreasing strength. Items from List 2 and 4 did not. The reasons for the discrepancy are not clear, although it should be noted, that the RP scores for items previously recalled once and twice from any list are based on a very small number of observations (the largest eight and the smallest four), thereby questioning the sensitivity of these items as a basis for drawing conclusions about output order. Moreover, clustering during FFR was exceedingly high. Of a total of 1226 items recalled by all Ss in FFR, only 42 of these did not appear in a cluster (i.e., were not adjacent to at least one other exemplar from the same category). Such high clustering would work against the detection of an output relationship based on item strength.

The shift in order of recall is more clearly seen when acquisition and FFR are directly compared. Collapsing across lists, Figure 5 gives mean RP scores for items recalled on Trial 4 of acquisition and also for items recalled in FFR for both the Categorized and Non-categorized groups. For acquisition separate entries are provided for items previously recalled 0, 1, 2 and 3 times prior to being recalled on Trial 4. For FFR the data are presented separately for items recalled 0, 1, 2, 3 and 4 times prior to being recalled in FFR. The data clearly underscore the pronounced shift in order of recall from acquisition to FFR, especially for the Non-categorized group.

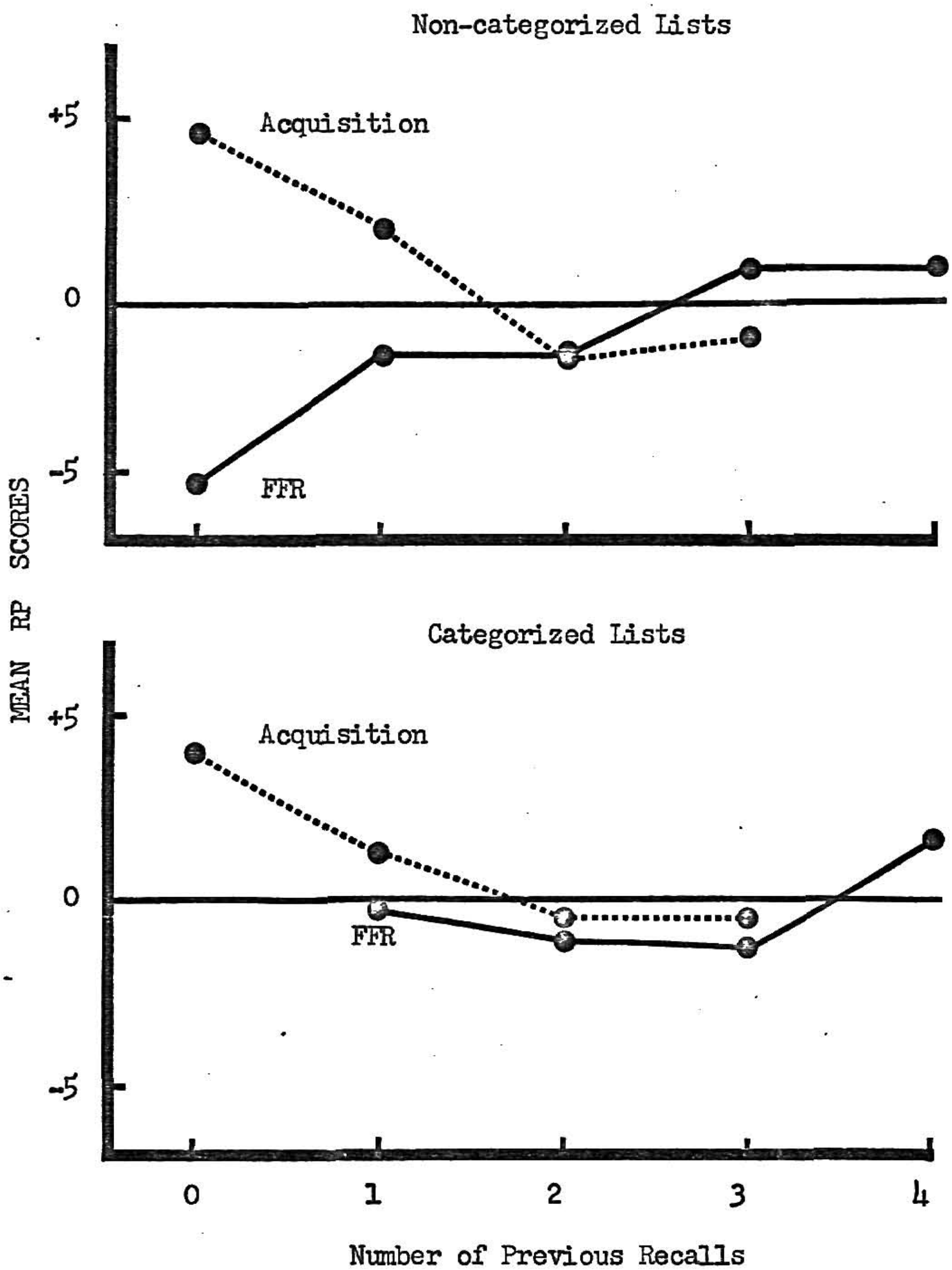


Fig. 5. Order of recall (RP) as a function of number of previous recalls, plotted separately for acquisition (Trial 4 only) and for FFR. Data are plotted separately for categorized and non-categorized lists.

# DISCUSSION

The major finding of the present investigation was the change in output order from acquisition to FFR. During acquisition, Ss emitted items in terms of increasing strength. New items (with minimal item strength) were output first followed by items that had been recalled once, twice and thrice in that order. During FFR, Ss recalled items in terms of decreasing strength. Items that had been recalled thrice and four times during acquisition were emitted first in FFR, followed by items recalled twice and once during acquisition. Items recalled for the first time in FFR were emitted last. Hence, whether items are recalled in terms of increasing or decreasing strength is clearly dependent on the task at hand. In the present experiment, Ss employed the priority strategy during the learning of four successive lists. With a change in task, the priority strategy disappeared and was replaced by an output order opposite to that during acquisition. The controversy then about how items are output from memory (see Battig et al., 1965; Mandler & Griffith, 1969; etc.) appears traceable to a failure of previous investigations to take into account procedural differences between tasks as determiners of output order.

As noted in the introduction, a major difference between acquisition and FFR is the absence in the latter task of a study (input) trial before attempted recall. Among other things, a study trial allows S to pay special attention to weak or what are perceived as difficult items (Brown & Slaybaugh, 1971), thereby possibly increasing their momentary availability or strength at the time of recall (Postman & Keppel, 1968). The Ss may therefore employ the strategy of recalling new items first and before old items to prevent their being lost or interfered with by attempted recall of old items (see Brown & Thompson, 1971). When the input trial is absent however, as in FFR, selective attention is not possible and items are output in terms of decreasing strength or in terms of their long-term availability in memory (Osgood, 1953).

The presence or absence of an input trial is not the only difference between acquisition and FFR. Another difference is expressed in Tulving's (1972) distinction between episodic and semantic memory. The main procedural (or operational) difference between these two memories in relation to recall is that the former is referenced by its perceptible autobiographical properties or attributes, (e.g., specific times, places or events that may have occurred during learning) as is conveyed by the instruction, "recall the items just presented to you". On the other hand, semantic memory is referenced by a stimulus that bears some semantic relation to the to-be-recalled material, e.g., "recall the states of the U.S.". Recall during acquisition may be largely episodic. However, during FFR the instruction, "now recall all words from all lists", may functionally be more similar to the instruction, "recall the names of foreign countries that you learned in school". This instruction given before FFR while alluding to some episodic occurrence of to-be-recalled material may also have a strong semantic component. That is, items may be recalled in terms of an organization related to item properties other than autobiographical attributes. The effect of the above as well as other differences between acquisition and FFR (e.g., the greater number of to-be-recalled items in FFR as compared to acquisition) cannot yet be specified with certainty.

The results of the present experiment also indicate that the measures of priority and clustering are not indexing conflicting strategies in learning categorized lists. The analysis of categorized lists during acquisition showed that the measures of priority and clustering did not always covary together. However, both strategies did occur together; high priority and high clustering were found in the same protocols. As mentioned in the results section, there are a number of ways Ss can achieve both high clustering and priority. It is also apparent that clustering takes longer to develop. Clustering increased as a function of lists while priority was fully manifest in first-list learning. This is not surprising. A strategy that is based on conceptual relationships among items (clustering) can be expected to develop much slower than a strategy that is based on the properties of individual items and that changes on a trial-to-trial basis (priority).

Finally, the present data indicate that the relatively poor recall of the Categorized group during FFR was largely due to a decrement in number of categories recalled and number of exemplars recalled per category. When a category was recalled the exemplars recalled from that category tended to be a sizable proportion of the set of exemplars presented during acquisition even though the number of exemplars recalled per category was less for FFR than for acquisition. This is consistent with the findings of Cohen (1966) who reported, among other things, that given one exemplar is recalled from a category the average number of exemplars recalled per category is invariant over a wide variety of conditions.

In conclusion, we may say that priority is a strategy specific to acquisition and not a long term memory phenomenon. Our own interpretation of the present results gives import to the study (input) trial and its relationship to the availability of items at the time of recall. Also, it seems that priority and clustering are not conflicting strategies in categorized lists. Rather, both may be found in the same protocol and are probably independent of each other.

# APPENDIX I

Non-categorized Lists

| List 1 | List 2 | List 3 | List 4 |
|---|---|---|---|
| check | industry | blood | interest |
| volume | lad | fellow | knife |
| heart | market | east | clouds |
| uniform | dream | key | account |
| nation | earth | opinion | nails |
| labor | knight | part | enemy |
| magazine | newspaper | hole | diamond |
| nest | finger | devil | wool |
| park | king | dish | yesterday |
| grave | root | journey | view |
| hair | bone | boat | eggs |
| wagon | secretary | wine | milk |
| kiss | pan | meal | Paul |
| library | office | shell | neck |
| face | university | hotel | horn |
| oak | lady | iron | steel |
| dollar | bowl | sale | Boston |
| example | history | land | shoes |
| tent | cheek | noise | tree |
| inch | family | nature | order |
| border | major | cattle | doctor |
| judge | noble | range | jump |
| deed | Jack | master | cents |
| ocean | attitude | language | train |
| paper | witness | cap | palace |
| railway | ship | question | rose |
| chicken | habit | author | lake |
| artist | chief | parent | goat |

Categorized Lists

| List 1 | List 2 | List 3 | List 4 |
|---|---|---|---|
| Arthur | breast | indigo | kidnapping |
| Pete | liver | mauve | stealing |
| Fred | abdomen | tan | fraud |
| Edward | hand | lavender | cheating |
| Frank | chest | brown | suicide |
| Roger | intestine | olive | murder |
| John | tooth | gray | speeding |
| | | | |
| polo | gun | teacher | horse |
| pool | arrow | professor | moose |
| wrestling | mortar | psychologist | deer |
| golf | sword | judge | camel |
| surfing | spear | mechanic | buffalo |
| fencing | cannon | policeman | beaver |
| track | bayonet | salesman | elk |
| | | | |
| volcano | lime | heat | nails |
| lake | lemon | blizzard | file |
| valley | coconut | wind | pliers |
| ridge | orange | snow | hatchet |
| glacier | fig | lightning | lathe |
| desert | tomato | shower | level |
| crater | prunes | rainbow | wrench |
| | | | |
| belt | quarters | velvet | Britain |
| dress | coins | flannel | Chile |
| hose | yen | satin | Germany |
| scarf | shilling | lace | Israel |
| overcoat | guinea | linen | Japan |
| stocking | bonds | calico | India |
| vest | pence | silk | Brazil |

# REFERENCES

Baddeley, A. D. Prior recall of newly learned items and the recency effect in free recall. Canadian Journal of Psychology, 1968, 22, 157-163.

Battig, W. F. Further evidence that strongest free-recalled items are not recalled first. Psychological Reports, 1965, 17, 745-746.

Battig, W. F., Allen, M., and Jensen, A. R. Priority of free recall of newly learned items. Journal of Verbal Learning and Verbal Behavior, 1965, 4, 175-179.

Battig, W. F., and Montague, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. Journal of Experimental Psychology, 1969, 80, (Monograph No. 3, Part 2).

Bousefield, W. A., and Sedgewick, C. H. An analysis of sequences of restructured associative responses. Journal of General Psychology, 1944, 30, 149-165.

Bousefield, W. A., Whitmarsh, G. A., and Esterton, J. Serial position effects and the "Marbe effect" in the free recall of meaningful words. Journal of General Psychology, 1958, 59, 255-262.

Brown, S. C., and Slaybaugh, G. D. Relationship between output order and confidence of recall. Paper read at Psychonomic Society meetings, 1971.

Brown, S. C., and Thompson, C. P. Relationship between item strength and order of free recall. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 444-448.

Cohen, B. H. Some-or-none characteristics of coding behavior. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 182-187.

Deese, J. Serial organization in the recall of disconnected items. Psychological Reports, 1957, 3, 577-582.

Deese, J., and Kaufman, R. A. Serial effects in recall of unorganized and sequentially organized verbal material. Journal of Experimental Psychology, 1957, 54, 180-187.

Mandler, G., and Griffith, M. Acquisition and organization of new items in the free recall of random lists. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 545-551.

Murdock, B. B., Jr. The serial position effect in free recall. Journal of Experimental Psychology, 1962, 64, 482-488.

Osgood, C. E. Method and Theory in Experimental Psychology. New York: Oxford, 1953.

Postman, L., and Keppel, G. Conditions determining the priority of new items in free recall. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 260-263.

Roberts, W. A. The priority of recall of new items in transfer from part-list learning to whole-list learning. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 645-652.

Roenker, D. L., Thompson, C. P., and Brown, S. C. A comparison of measures for the estimation of clustering in free recall. Psychological Bulletin, 1971, 76, 45-48.

Sheull, T. J. Clustering and organization in free recall. Psychological Bulletin, 1969, 72, 353-374.

Shuell, T. J., and Keppel, G. Item priority in free recall. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 969-971.

Smith, A. D., D'Agostino, P. R., and Reid, L. S. Output interference in long-term memory. Canadian Journal of Psychology, 1970, 24(2), 85-89.

Thompson, C. P. Organization in memory: Multi-trial free recall of categorized lists. In R. F. Thompson and J. Voss (Eds.), Topics in Learning and Performance. New York: Academic Press, 1972.

Thorndike, E. L., and Lorge, I. The Teachers Word Book of 30,000 Words. New York: Teachers College, Columbia University Bureau of Publications, 1944.

Tulving, E. Theoretical issues in free recall. In T. R. Dixon and D. L. Horton (Eds.), Verbal Behavior and General Behavior Theory. Englewood Cliffs, N.J.: Prentice Hall, 1968.

Tulving, E. Episodic and semantic memory. In E. Tulving and W. Donaldson (Eds.), Organization and Memory. New York: Academic Press, 1972.

Tulving, E., and Arbuckle, T. Y. Sources of intratrial interference in immediate recall of paired associates. Journal of Verbal Learning and Verbal Behavior, 1963, 1, 321-334.

Underwood, B. J. Degree of learning and the measurement of forgetting. *Journal of Verbal Learning and Verbal Behavior*, 1964, 3, 112-129.

Underwood, B. J., and Schulz, R. W. *Meaningfulness and Verbal Learning*. New York: J. B. Lippincott Company, 1960.

## Footnotes

[1]Several investigators (e.g., Postman & Keppel, 1968; Baddeley, 1968) have argued that PRNI is an artifact of recency as evidenced by the fact that new items tend to rotate to favored serial positions (i.e., end positions) as a result of randomization and by an elimination of PRNI when recency is eliminated by some interfering interpolated task. On the other hand, Mandler and Griffith (1969), Battig (1965), Roberts (1969), and Brown and Thompson (1971) have shown that PRNI also occurs when only items from the middle of the list are considered, thereby discrediting the notion that priority is an artifact of recency.

[2]The effect of serial position was largely restricted to new items. During early trials there was a slight tendency for old end items to be emitted before old middle items. However, the difference was not present during latter trials.

[3]This formula is used to underscore the similarity between SRR and the z-score. The expression may easily be made positive by multiplying out the negative sign, i.e., $(Mdn_R - R_i)/\sigma_R$.

[4]The same biases apply to maximum negative priority but in the opposite direction because maximum positive and negative priority are symmetric to each other.

[5]The present objections to O-E apply when any segment of the recall protocol is used as the unit of analysis (e.g., fourths, halves, or eighths of recall).

[6]It may be noted that RP is related to RPI by the following relationship:

$$RP = \frac{RPI_i}{Max(RPI)} \text{ , where} \tag{5}$$

$RPI_i$ = the sum of the RPI scores of the i items of interest, (e.g., new items falling in recency positions, items recalled once, etc.) and,

Max(RPI) = sum of RPI scores if all items of interest were emitted first.

[7]One tailed t-tests under the null hypothesis that RP scores (averaged across lists) were not greater than zero (i.e., no priority) all proved to be significant. For the Non-categorized group these means were .17, .38 and .46 for Trials 2, 3 and 4, respectively. For the Categorized group they were .20, .25 and .38, respectively.

[8]On Trial 1 priority and clustering cannot be compared because priority is undefined on the first trial.

# COMPARISON OF OUTPUT ORDER IN IMMEDIATE AND DELAYED FREE RECALL

by

LUIS MORALES FLORES JR.

B.A., University of the Philippines, 1966

---

AN ABSTRACT OF A MASTER'S THESIS

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Psychology

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1973

ABSTRACT

During recent years a considerable amount of research in free recall has been directed toward the specification of the order in which subjects (Ss) recall items from memory. In this regard, some studies have reported that items are recalled in terms of decreasing strength (i.e., the best learned items recalled first), while others have found that items are recalled in terms of increasing strength (i.e., weak or poorly learned items recalled first). The major purpose of the present experiment was to resolve the discrepancy. Another purpose was to examine clustering (the tendency of Ss to recall conceptually related words together) and the relationship between clustering and order of item recall.

A total of 41 Ss learned four successive lists of either categorized or non-categorized words. Each list was twenty-eight words in length and was practiced for four study-test trials. Immediately after the learning of the lists (acquisition), all Ss were given a final free recall (FFR) test wherein they were asked to recall all words from all lists. To facilitate analysis a measure of output order was developed (relative priority) which was shown to be superior to previous measures because of its invariance to factors unrelated to output order (e.g., list length and different numbers of newly-recalled items). The results showed that order of recall during acquisition was in terms of increasing

item strength (priority). On Trial 4, for example, words recalled for the first time were emitted first followed by items that were recalled correctly once, twice and thrice on previous trials, in that order. During FFR Ss recalled items in terms of decreasing strength. Items correctly recalled thrice and four times during acquisition were recalled first in FFR. These were followed by items recalled once and twice during acquisition. Items recalled for the first time in FFR were recalled last. The dramatic shift in the manner items were recalled from acquisition to FFR underscored the importance of characteristics of the task as a determiner of output order. The results were interpreted in terms of interference theory and stressed the role of the input (study) trial to account for the shift in order of recall.

Analysis of clustering and output order in acquisition showed that clustering increased across trials on List 1 and 2 and thereafter remained relatively stable on Lists 3 and 4. The tendency for weak items to be recalled first (priority), however, increased across trials for all lists. The lack of a systematic covariance between clustering and priority suggested the independence of the two phenomena.