Reprinted from JOURNAL OF EXPERIMENTAL PSYCHOLOGY, Vol. 68, No. 4, October 1964
Printed in U. S. A.

SIZE OF REHEARSAL GROUP AND SHORT-TERM MEMORY [1]

WAYNE A. WICKELGREN
Massachusetts Institute of Technology

132 undergraduates were given sequences of 6 to 10 digits presented at 1 digit per second with ordered recall instructions. Ss were instructed to rehearse silently in nonoverlapping groups of 1, 2, 3, 4, or 5 digits. Rehearsing in 3's was optimal, being superior to 2's in ordered recall ($p < .01$), item recall ($p < .01$), and position recall ($p < .05$), insignificantly superior to 4's in ordered and position recall, but not item recall, and significantly superior to 5's by ordered and position recall ($p < .01$), but not by item recall. Errors in positioning digits tended to the same position in different groups for groups rehearsing by 2's and 3's ($p < .01$) and to other positions in the same group for groups rehearsing by 4's and 5's ($p < .05$). The results support the hypothesis that only 3 serial-position concepts (beginning, middle, and end) are important cues in short-term memory.

In memory-span experiments *S*s have frequently indicated that they group items in 2's, 3's, 4's, 5's, etc. Oberly (1928) reported that when introspectively trained *S*s indicated the sequences on which they grouped and the sequences on which they did not group, the median memory span for ungrouped sequences was 4.1; the memory span for both grouped and ungrouped sequences was 8.7. A study of improvement in memory span with practice by Martin and Fernberger (1929) reported that marked improvement occurred only after the two *S*s in the experiment tried various methods of grouping. Improvement consisted in mastering increasingly larger size groups to a maximum at five items per group. When *S*s attempted grouping in 6's, their performance declined, but the experiment was not continued long enough to determine if they could ever learn to group in 6's more effectively than in 5's.

Fraisse (1945) made a nonintrospective attempt to determine the number of groups into which *S*s divide a sequence of 10 digits. A group was defined as a cluster of correct items separated by one or more errors. The most frequent number of groups was two (42.3%) and the next most frequent number of groups was three (32.9%). This determination of grouping method assumes that *S*s rarely forget the items they have grouped and virtually never remember two groups in succession without making an error between the groups. These assumptions seem unjustified.

More recently, Pollack, Johnson, and Knaff (1959), in a study of terminal running memory span, attempted to manipulate grouping method by temporal grouping of the items presented in 1's, 2's, 3's, 4's, and 6's. For both known and unknown length sequences the optimum size group was four, but for known length sequences there were no significant differences between the different temporal grouping procedures.

Any attempt to manipulate grouping method is ambiguous unless sufficient significant differences are obtained to indicate that the manipulation was effective. For example,

[1] This investigation was supported by Grant MH 07726-01 from the National Institute of Mental Health, Public Health Service.

the absence of any significant differences between different size groups in the known length condition of Pollack et al. could mean that when the length of the sequence is known, any of these grouping methods is equally good, or that when the length of the sequence is known, *S*s use the same method of grouping irrespective of temporal grouping in the presentation of items. The latter is quite possible since it appears that *S*s were not instructed to rehearse only the items presented in the last group.

Severin and Rigby (1963) studied the ability to dial from memory a 7-digit telephone number after a 4-sec. study period. The telephone number was presented in one of four different patterns of spatial grouping, for example, 924-1758, 9-241-758, 92-41-758, 9-24-17-58. The conventional 3-4 grouping was superior to the 1-3-3, 2-2-3, and 1-2-2-2 groupings, which were all equivalent. Severin and Rigby interpret their results in terms of positive transfer from previous dialing experience.

Assumptions concerning grouping have been made to account for other effects in short-term memory. Waugh (1960) suggests that there is a primacy group and a recency group in all memory-span experiments and that these two groups are responsible for the serial-position curve. Miller (1956) assumes that under some circumstances *S*s can group items into "chunks" and, thereby, increase their memory spans because short-term memory is limited by the number of "chunks," not by the information in each chunk.

In order to investigate further the effects on short-term memory of different methods of grouping, it is necessary to have a relatively precise definition of what a method of grouping is. When someone says he grouped in 3's, what does he mean? Presumably there is a difference in the manner of remembering an ungrouped sequence of 10 digits, five groups of 2 digits, three groups of 3 digits and one group of 1 digit, two groups of 4 digits and one group of 2 digits, two groups of 5 digits, etc. For the purposes of the present investigation it is assumed that, whatever else a grouping method is, it is a method of rehearsal. Grouping in 2's means rehearsing in 2's; grouping in 3's means rehearsing in 3's; etc.

The present experiment is designed to investigate the consequences for short-term memory of instructing *S*s to rehearse a sequence of digits in groups of one, two, three, four, and five during presentation. Previous studies have investigated the effects of grouping on only one dependent variable, ordered recall. Ordered recall is sensitive to forgetting of items and to forgetting of the correct position of an item in the sequence. This study will attempt to determine whether grouping affects recall of items and recall of position in a comparable manner. In addition, the attempt to control rehearsal methods in the presentation period may provide evidence on the role of such rehearsal in other studies of short-term memory.

Method

Using a Wollensak (Model T-1515-4) tape recorder, 132 *S*s were given sequences of digits at 1 digit per second. The *S*s had a 20-sec. interval between the end of one sequence and the beginning of the next sequence in which to record in order the sequence just heard, leaving blanks to indicate positions for which they could not recall the correct item and did not wish to guess. The *S*s were not allowed to write down more items than the sequence contained.

There were six silent rehearsal conditions: (R1) *S* was instructed to rehearse only the preceding item. (R2) *S* was instructed to rehearse the first item in a pair after hearing the first item in the pair and to rehearse the pair of items after hearing the second item of the pair. The pairs were to be nonoverlapping, and *S* was instructed not to go back to a previous pair. (R3) *S* was instructed to rehearse in nonoverlapping groups of three items. Rehearsal after each of the first two items of any group of three was the same as in R2; after the third item all three items were to be rehearsed. (R4) *S* was instructed to rehearse in nonoverlapping groups of four items. Rehearsal after each of the first three items in any group of four was the same as in R3; after the fourth item the entire group of four items was to be rehearsed. (R5) *S* was instructed to rehearse in nonoverlapping groups of five items. Rehearsal after the first four items in any group was the same as in R4; after the fifth item *S* was to rehearse the entire group of five items. (R3V) *S* was instructed as in R3 with the added instruction to attempt to visualize each group of three items.

132 *S*s were randomly assigned to one of six groups, 22 *S*s per group. Each group received two of the six conditions with half of the *S*s in each group receiving one condition first and half receiving the other condition first. Group R1-2 received Cond. R1 and R2. Group R2-3 received Cond. R2 and R3. Group R3-4 received Cond. R3 and R4. Group R4-5 received Cond. R4 and R5. Group R3-5 received Cond. R3 and R5. Group R3-3V received Cond. R3 and R3V.

There were two sets of 35 sequences each consisting of 7 sequences at each length from 6 to 10, inclusive. Each group received the first set of sequences under one condition and the second set of sequences under the other condition. The 7 sequences in each set at each length were of the following seven types:

(All different) No digit repeated in the sequence (921475).
(ii_B) A Length 2 run at the beginning of the sequence (773519).
(ii_M) A Length 2 run at the middle of the sequence (698834).
(ii_E) A Length 2 run at the end of the sequence (498355).
(iii_B) A Length 3 run at the beginning of the sequence (777351).
(iii_M) A Length 3 run at the middle of the sequence (688834).
(iii_E) A Length 3 run at the end of the sequence (498555).

Results

The data were analyzed for "ordered recall," "item recall," and "position recall" of both entire sequences and individual items. The *S*'s report of a sequence is correct by an *ordered-recall* criterion and by a *position-recall* criterion if and only if all items are recalled in the correct order. The *S*'s report of a sequence is correct by an *item-recall* criterion if and only if all items are recalled correctly, irrespective of order. The *ordered-recall* error rate for a condition is the number of sequences incorrect by an ordered-recall criterion divided by the total number of sequences in the condition. The *item-recall* error rate is the number of sequences incorrect by an item-recall criterion divided by the total number of sequences. The *position-recall* error rate is the number of sequences correct by an item-recall criterion, but incorrect by an ordered-recall criterion, divided by the number of sequences correct by an item-recall criterion. The item-recall error rate and the position-recall error rate are statistically independent and the ordered-recall error rate combines these two independent factors. Analysis of ordered recall, item recall, and position recall for individual items is entirely analogous to the same analysis for sequence recall. Only the analysis for sequence recall, summed over all sequence types, will be reported. The results obtained by analysis of item recall and the results obtained by separate analysis of each sequence type were virtually identical to the results reported below. Since the 10 different sequences of each type in each condition are evenly distributed over five lengths, no definite conclusion is possible regarding the existence of complex interactions between sequence type and length in their combined effects on the efficiency of different rehearsal conditions.

Table 1 reports the error rates in ordered recall, item recall, and posi-

TABLE 1

SEQUENCE RECALL ERROR RATES (%)

Group	Condition	Ordered Recall	Item Recall	Position Recall	Condition	Ordered Recall	Item Recall	Position Recall
R1-2	R1	42.7	35.4	11.4	R2	46.4	40.1	10.5
R2-3	R2	44.8**	38.9**	9.8*	R3	34.5	30.6	5.7
R3-4	R3	38.1	33.2	7.3	R4	39.5	32.1	10.8
R3-5	R3	33.3**	29.9	4.9**	R5	46.1	34.4	18.0
R4-5	R4	30.0*	23.0	10.0*	R5	35.0	25.1	13.2
R3-3V	R3	30.2	25.7	6.2	R3V	28.0	24.1	5.1

Note.—An error rate in the left condition that is significantly different from the corresponding error rate in the right condition by the Wilcoxon matched-pairs signed-ranks test is indicated next to the appropriate error rate in the left condition by asterisks.

* $p < .05$.

** $p < .01$.

tion recall for each condition in each group. The error rates for each condition averaged over all groups in which that condition occurred are presented in Fig. 1. Differences in error rates between the two conditions in each group were tested for statistical significance by the Wilcoxon matched-pairs signed-ranks test with $N = 22$ in each comparison.

The significance levels from the comparisons of ordered recall in Groups R1-2, R2-3, R3-4, R3-5, and R4-5 were combined using the method of Fisher (1938) into an overall significance level for the effectiveness of the manipulation of rehearsal group size. Only one measure of recall can be used because the significance levels being combined must be independent; ordered recall was chosen because it is the only one of the three measures of recall that includes all the data. Fisher's method yielded χ^2 (10) $= 19.70$, $p < .05$, indicating that the instructional manipulation of silent rehearsal was effective.

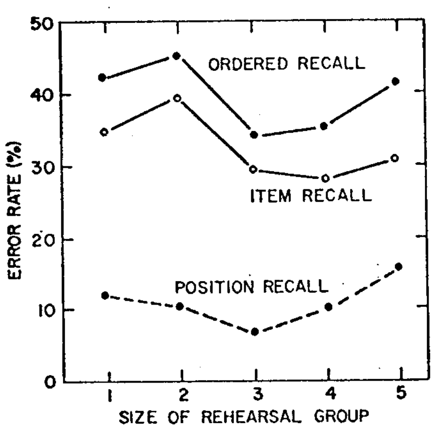

FIG. 1. Effect of size of rehearsal group on sequence recall.

Rehearsing in 3's was optimal, but rehearsing in 4's was almost as effective. Rehearsing in 3's was insignificantly superior to 4's in ordered recall and position recall and insignificantly inferior to 4's in item recall. Rehearsing in 3's was significantly superior to 2's in every respect. Rehearsing in 3's was significantly superior to 5's in ordered and position recall, but not in item recall. Similarly, rehearsing in 4's was significantly superior to 5's in ordered and position recall, but not in item recall. There were no significant differences between R1 and R2, nor were there any between R3 and R3V. The insignificant superiority of R1 to R2 in ordered and item recall coincides with results obtained by Pollack et al. (1959), suggesting that perhaps no grouping is somewhat superior to the poorest method of grouping.

As illustrated in Fig. 1, item-recall errors reach a minimum at R3, R4, and R5; position-recall errors, however, rise sharply from R3 to R4 to R5. The decline in ordered recall from R3 to R4 to R5 is the result of poorer recall of position, not poorer recall of items.

To obtain more information about the curvilinear relationship between size of rehearsal group and position recall, every item that was recalled correctly by an item-recall criterion, but recalled in the wrong position, was classified into one of three position-error categories. *Within-group* errors refer to items recalled in the correct group, but at the wrong position within the group. *Within-position* errors refer to items recalled in the wrong group, but in the correct position within the group. *Other* errors refer to items recalled in the wrong group and wrong position within the group.

To obtain an independent measure of the relative frequency of within-group and within-position errors for Cond. R2, R3, R4, and R5, the frequency of each type of error was compared to the frequency of all *other* errors. In order to determine if the relative frequency of within-group or within-position errors was affected by rehearsal group size, it was necessary to compare the relative frequency of each type of error in any rehearsal condition to the same relative frequency computed for a standard "ungrouped" condition, in this case R1. Computation of the comparable relative frequency in R1 involves breaking R1 into the same groups as in the condition with which it is being compared. Chi-square tests were used to test for significant differences between relative frequencies.

Table 2 presents the relative frequency of within-group and within-position errors for the all-different sequences in each condition and the comparable relative frequency for the all-different sequences in R1. Only the all-different sequences were analyzed in this way to eliminate any variance resulting from the runs of identical items in the other sequences. Compared to R1, the relative frequency of within-position errors was significantly greater for R2 and R3 and the relative frequency of within-group errors was significantly greater for R4 and R5.

TABLE 2

ERRORS IN POSITIONING ITEMS

Cond.	Error Ratios in R2 to R5		Comparable Error Ratios in R1	
	Within-Group / Other	Within-Position / Other	Within-Group / Other	Within-Position / Other
R2	.61	2.75**	.66	1.11
R3	2.15	1.19**	1.36	.33
R4	3.33*	.22	1.64	.30
R5	4.72***	.00	1.43	.00

Note.—Error ratios in R2 to R5 that are significantly greater (χ^2 test) than the comparable error ratios in R1 are indicated by asterisks.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

Discussion

Let us assume that rehearsal of two items in close succession strengthens the association between them. It is the rehearsal of an item that takes time. Therefore the more associations that are strengthened for each item rehearsed, the better. When one rehearses a group of n items he strengthens $n - 1$ direct forward associations (and perhaps $[n(n-1)/2] - (n-1)$ remote forward associations). If no other factor were operative, this would imply that the larger the rehearsal group the better, because the ratio of strengthened direct (and remote) associations to rehearsed items becomes progressively more favorable with increasing size of rehearsal group. This explains why grouping in 3's is superior to grouping in 2's or 1's. It does not explain why grouping in 3's is insignificantly superior to 4's and significantly superior to 5's. Another factor must be operative. This factor must explain: (*a*) why there is a decrement in recall of position from 3's to 4's to 5's, (*b*) why there is no decrement in recall of items from 3's to 4's to 5's, and (*c*) why grouping method systematically affects the positioning of items recalled in *incorrect* positions.

One obvious possibility is that *S*s have difficulty in rehearsing five items in the 1-sec. interval between items. Three sets of facts argue against this. First, *E* determined the rate of presentation so that he had no such difficulty, and no *S* reported any such difficulty. Second, the largest group of five items is rehearsed only once in each sequence and only after the first four items have been made relatively familiar by immediately prior rehearsal. Sperling (1963) reports a maximum rate of rehearsal for highly familiar sequences of 10 syllables per second. Third, this explanation is not consistent with the absence of any decrement in item recall from R3 to R4 to R5, and it provides no explanation of the systematic position errors.

A second possibility is that *S*s rehearse larger groups at faster rates. If we assume that direct associations are formed in short-term memory between adjacent items and that remote associations are formed between nonadjacent items, then a faster rate of rehearsal implies less difference between direct and remote associations in strength of association. This accounts for the decrement in position recall from R3 to R4 to R5, and it is certainly consistent with the absence of any decrement in item recall from R3 to R4 to R5 since there are as many or more strong associations between items in the larger rehearsal group conditions. The large number of strong remote associations would tend to keep item recall constant in the face of a decline in position recall. The explanation in terms of remote associations also explains the greater frequency of within-group errors in R4 and R5 compared to groups of identical size in R1. However, it is totally incapable of explaining: (*a*) the *absence* of a significantly greater frequency of within-group errors in R2 and R3 than in groups of identical size in R1, (*b*) the greater frequency of within-position errors in R2 and R3 compared to identical positions in R1, and (*c*) the *absence* of a significantly greater frequency of within-position errors in R4 and R5 compared to identical positions in R1.

The elevated frequency of within-position errors in R2 and R3 and the abrupt shift that occurs between R3 and R4 in the dominant position error tendency suggest the third hypothesis that rehearsing in groups introduces two sets of *serial-ordering cues*—the first set corresponding to groups and the second set corresponding to positions within a group. The data for both within-group errors and within-position errors suggest that only three different serial-order cues (beginning, middle, and end) are used in each set. So long as the mapping from *position* serial-order cues to items within a group is one-to-one, there is a relatively low frequency of *within-group* errors. When there are more than three items in a rehearsal group, then the mapping is not one-to-one and the frequency of within-group errors increases. So long as the mapping from *group* serial-order cues to groups of items is one-to-one, there is a relatively low frequency of *within-position* errors (errors across groups that maintain the correct within-group position). When there are more than three groups of items, then the mapping is not one-to-one and the frequency of within-position errors increases. Thus, the hypothesis about serial-order cues accounts for the systematic position errors and the decrement in recall of position from R3 to R4 to R5. It also accounts for the lack of a decline in recall of items from R3 to R4 to R5, since there are associations from serial-ordering cues to every item and one-to-many mappings lose order information, not item information.

Remember that the ratio of strengthened associations to rehearsed items accounts for why R3 is superior to R2 and R1 in both item and position recall.

The hypothesis about serial-order cues accounts for all the remaining findings and is compatible with the first factor. The hypothesis about serial-order cues is also compatible with the hypothesis about remote associations. Therefore, the present findings do not deny the importance of remote associations in short-term memory. The findings are only evidence *for* the importance of serial-order cues in short-term memory. Serial-order cues may be another mechanism, in addition to the greater strength of direct associations, for discriminating between direct and remote associations in recall.

REFERENCES

FISHER, R. A. *Statistical methods for research workers.* London: Oliver & Boyd, 1938.

FRAISSE, P. Etudes sur la mémoire immédiate: III. L'influence de la vitesse de presentation et de la place des éléments. La nature du present psychologique. *Annee psychol.*, 1945, **45**, 29–42.

MARTIN, P. R., & FERNBERGER, S. W. Improvement in memory span. *Amer. J. Psychol.*, 1929, **41**, 91–94.

MILLER, G. A. The magical number seven. *Psychol. Rev.*, 1956, **63**, 81–97.

OBERLY, H. S. A comparison of the spans of "attention" and memory. *Amer. J. Psychol.*, 1928, **40**, 295–302.

POLLACK, I., JOHNSON, L. B., & KNAFF, P. R. Running memory span. *J. exp. Psychol.*, 1959, **57**, 137–146.

SEVERIN, F. T., & RIGBY, M. K. Influence of digit grouping on memory for telephone numbers. *J. appl. Psychol.*, 1963, **47**, 117–119.

SPERLING, G. A model for visual memory tasks. *Hum. Factors*, 1963, **5**, 19–31.

WAUGH, N. C. Serial position and the memory span. *Amer. J. Psychol.*, 1960, **73**, 68–79.

(Received September 25, 1963)