Initial-recall grouping in free-recall learning*

JOHN A. ROBINSON

University of Louisville, Louisville, Ky. 40208

In the course of a demonstration free-recall learning (FRL) experiment, it was noted that Ss differed markedly in the extent to which they grouped initially recalled words together in subsequent recall periods. Further examination indicated that the degree of such grouping was inversely related to both rate and asymptote of learning. These results were interpreted as the consequence of persisting in the use of an inappropriate encoding and retrieval scheme.

In the course of a demonstration free-recall learning (FRL) experiment, it was noted that Ss differed markedly in the extent to which words recalled on the first trial were grouped together in subsequent recall periods. The question of interest was whether the amount or consistency of such initial recall grouping (IRG) was systematically related to level of performance.

SUBJECTS

Twenty-one enlisted men assigned to duty at the U.S. Army Armor Center, Fort Knox, Kentucky, served as Ss in this experiment.

MATERIALS

The list to be learned consisted of 10 common two-syllable nouns selected from a wide range of Thorndike-Lorge frequencies of occurrence. All words had different initial letters. Five randomizations of the list were prepared in such a way as to minimize second-order sequential redundancies in list presentation.

PROCEDURE

All Ss were given standard FRL instructions at the beginning of the experiment. The list was presented on a Stowe memory drum at a rate of 1.5 sec per word with a 1.0-sec interword interval. A red line appeared in the aperture of the memory drum to signal the end of the list, at which time S began his recall. There were 10 alternating study-test trials. Ruled sheets of paper were provided for written recall. The retention interval was 1 min.

RESULTS

Birnbaum's (1968) P_{0-0} measure was used to index the degree of IRG. Adapted for our purposes and renoted as P_{11} , this measure is the ratio of the number of adjacent pairs of initially recalled words observed in any particular instance to the maximum possible number of such pairs. The latter value is equal to the number of

initially recalled words present in the designated instance minus one. The index is then a proportion, and high values indicate that initially recalled words were produced as a block on subsequent trials rather than interspersed among newly recalled words. It should be noted that the value of P_{11} is independent of the total number of words recalled on any trial and of the specific number of initially recalled words that recur on any later trial.

For each S, P_{11} was calculated for each of Recall Trials 2 through 10. The Ss were then rank ordered from highest to lowest on the basis of average P_{11} for all trials. These averages ranged from .22 to .82, with a mean of .51 and an SD of .18. For comparative purposes, the distribution of ranked scores was subdivided into three groups comprising Ss who exhibited, respectively, a high (H), medium (M), or low (L) average degree of IRG. The individual-trial and total-trial P_{11} means for these groups are given in Table 1. It can be seen that Group L progressively abandoned IRG, but that Groups M and H maintained such organization, though to different degrees, throughout practice.

The mean number of words recalled on successive trials is shown in Fig. 1 for each group. It is apparent that the three groups differ both in rate and asymptote of learning and, further, that recall performance is inversely related to degree of IRG. Particularly remarkable is the fact that the average performance of Group H was nearly invariant from Trial 2 through Trial 9, whereas recall increased rather

Table 1
Mean P_{11} for Subgroups on Successive Recall Trials

Recall Trial	Group		
	High	Medium	Low
2	.73	.45	.57
3	.82	.54	.48
4	.61	.47	.30
5	.63	.59	.28
6	.77	.47	.05
7	.80	.63	.26
8	.74	.38	.21
9	.73	.38	.50
10	.56	.42	.09

*Preparation of this report was supported by Grant 1008-68 from the Air Force Office of Scientific Research, Office of Aerospace Research, United States Air Force. The data were collected by R. A. Kulp and codified for analysis by Barry J. Rabin.

steadily—though at different rates—for Groups M and L during the same period. Stated another way, the average gain in recall by Trial 10, relative to Trial 1, was only 1.71 words for Group H but was 4.43 and 5.29 words for Groups M and L, respectively. This result is all the more interesting considering the fact that the direction of differences was exactly the opposite on the first trial. It is to be noted, however, that the inversion of rank occurred rapidly, viz, by Trial 3, and was sustained thereafter.

The recall data were evaluated in several ways. First, a Group by Trials analysis of variance was performed, the results of which indicated that both main effects were significant: for groups, $F(2,18) = 14.67$, $p < .01$, for trials, $F(9,162) = 23.23$, $p < .01$. The Groups by Trials interaction was also significant, $F(18,162) = 2.05$, $p < .01$. Jonckheere's (1954) S statistic was used to evaluate the additional question of whether or not the obtained order of group means was nonrandom. This statistic is based on rank differences of individual scores in all possible pairs of between-group comparisons. The individual scores in this instance were the total numbers of words recalled over all trials. An S of 95 was obtained, $Z = 4.00$, $p < .001$. Thus, the hypothesis that group differences were randomly ordered was rejected. Finally, a rank-order correlation of $-.65$ ($p < .01$) obtained by conversion from S indicated that the average degree of IRG was systematically and negatively related to overall performance.

Inasmuch as the value of P_{11} is independent of the number of "old" (i.e., initially recalled) words present in any instance, one cannot infer the character of

Fig. 1. Mean number of words recalled as a function of average degree of IRG.

recall directly from the value of the grouping index. That is, a high value cannot be construed as indicating that recall consisted mainly of old words, or a low value as indicating that "new" words (i.e., words not recalled on the first trial) predominated in recall. In fact, as can be judged from Table 2, there were consistent differences between groups in both the proportion of old words retained and the proportion of new words added in subsequent recall periods. Thus, the disruptive effect of IRG was not confined to the acquisition of new words, but interfered with retrieval of old words as well.

Battig, Allen, & Jensen (1965) have described a form of grouping in FRL which they termed "the priority of recall of newly learned items" (PRNI). It is the tendency for items that were not recalled on the first trial to occur in early output positions when they are recalled for the first time on some later trial. Thus, IRG and PRNI are complementary phenomena and should covary in magnitude. The standard recall-rank score (SRR) described by Battig et al (1965) was computed for each new item at the point of its first occurrence. The scores were then averaged over items and Ss within groups. Positive scores indicate grouping of new items in premedian recall positions, negative scores indicate grouping into postmedian ranks, and values at or near zero indicate dispersion of new words over the entire range of recall positions. The obtained mean SRR scores were: for Group H, +.33; for Group M, -.08; and for Group L, -.04. Thus, the magnitude of the PRNI effect did generally conform to the degree of IRG. These scores are, however, partly confounded by differences between groups in level of initial recall. That is, the probability that some new words would obtain negative rank scores, even when completely segregated, is greater when there are more new words to be recalled. Nevertheless, inspection of the protocols confirmed that new words were more widely dispersed in the recall of L Ss and more obviously grouped in the recall of H Ss. While segregation of new and old words persisted in the recall of H Ss, there was no discernible stability to the order in which the groups as groups were recalled on later trials.

DISCUSSION

These results indicate that IRG was a

Table 2
Mean Proportion of Old and New Words
Produced on Successive Trials as a Function
of Average Degree of IRG

Recall Trial	Group					
	High		Medium		Low	
	Old	New	Old	New	Old	New
2	.81	.52	.75	.60	.92	.55
3	.76	.59	.79	.65	.80	.70
4	.75	.69	.82	.68	.80	.84
5	.76	.59	.79	.79	.80	.80
6	.71	.66	.75	.68	.82	.90
7	.71	.69	.79	.75	.82	.90
8	.79	.62	.79	.73	.85	.88
9	.76	.66	.65	.84	.89	.90
10	.79	.74	.86	.85	.90	.93

nonoptimal form of organization in FRL. IRG is, de facto, grouping by state of learning, i.e., a simple binary categorization of words as old or new relative to the first recall trial. It is, then, an organizational scheme that is based on a relatively transient aspect of S's performance rather than on inherent, and hence more stable, attributes of the words as words.

It could be argued that by adopting IRG as a recall strategy, S tags each item as new or old during list presentation and that these tags function both as storage locations or bins and as retrieval cues, i.e., as labels for the bins. By this hypothesis, recall consists of simply emptying each bin in succession. But this hypothesis cannot account for the consistently depressed recall curve of H Ss. In particular, it provides no explanation of why only some of the words in each bin were recalled on each trial. It would be more likely by this view that one or the other group of words would be unavailable, i.e., that only new or only old words would be recalled. Perhaps the bin hypothesis is correct, but Ss were unable to adequately discriminate state of learning after several practice trials. This suggestion is reasonable, given the highly transient basis of the new-old distinction. However, if Ss were unable to discriminate state of learning, they should not have been able to group words in that manner in recall.

An alternative explanation is that the detrimental effects of IRG were the result of a perseverative organizational set. We have already noted that IRG was based on performance rather than on stimulus attributes. It seems likely, however, that as a result of the continually changing presentation order, Ss would have become

aware of other ways of organizing the list, e.g., in terms of associative or mnemonic relations among the words. Furthermore, these alternatives should have become more salient with each repetition of the list. As a consequence, Ss who had adopted IRG would have been faced with a conflict during presentation, viz, to persist in the use of an established though less salient scheme or to abandon it (in whole or in part) in favor of emerging alternatives. Retrieval is presumably guided by the nature of the encoding that has taken place, and recent evidence clearly indicates that input must be coded in a consistent fashion from trial to trial if increments in recall probability are to occur (Bower & Winzenz, 1969). Thus, if as a result of the conditions just described, some items were encoded in other than new-old terms, they would not have been active members of the bins at that point and could not have been retrieved by means of the still dominant IRG schema.

This perseveration-cum-conflict hypothesis accounts for the highly segregated nature of H S's recall as well as for the some-or-none character of that recall. It is also consistent with other reports of the effects of inappropriate or conflicting organization on FRL (e.g., Tulving, 1966; Bower & Lesgold, 1969). The results of our observations confirm the view, then, that it is not organization per se but *appropriate* organization which facilitates FRL.

REFERENCES

- BIRNBAUM, I. M. Free-recall learning as a function of prior-list organization. *Journal of Verbal Learning & Verbal Behavior*, 1968, 7, 1037-1042.
- BATTIG, W. F., ALLEN, M., & JENSEN, A. R. Priority of newly learned items. *Journal of Verbal Learning & Verbal Behavior*, 1965, 4, 174-179.
- BOWER, G. H., & LESGOLD, A. M. Organization as determinant of part-to-whole transfer in free recall. *Journal of Verbal Learning & Verbal Behavior*, 1969, 8, 501-506.
- BOWER, G. H., & WINZENZ, D. Group structure, coding, and memory for digit series. *Journal of Experimental Psychology*, 1969, 80 (2, Pt. 2).
- JONCKHEERE, A. R. A distribution-free k-sample test against ordered alternatives. *Biometrika*, 1954, 41, 133-145.
- TULVING, E. Subjective organization and effects of repetition in multi-trial free-recall learning. *Journal of Verbal Learning & Verbal Behavior*, 1966, 5, 193-197.