JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR 4, 175-179 (1965)

Priority of Free Recall of Newly Learned Items[1]

WILLIAM F. BATTIG

University of Maryland, College Park, Maryland

MAX ALLEN

University of Connecticut, Storrs, Connecticut

AND

ARTHUR R. JENSEN

University of California, Berkeley, California

Data are presented from three independently conducted free-recall experiments, representing a variety of procedures and learning materials, which demonstrate that newly-learned items tend to be recalled prior to those items which have been recalled correctly on previous trials, in sharp contrast with the widespread assumption that order of free recall directly reflects item strength.

As a corollary to Marbe's law, which states that overall frequency or strength of free word associations by the group is inversely related to latency of emission by individual *S*s, it has been generally assumed that the order in which items are given in free recall directly reflects item strength, so that stronger items tend to be recalled prior to weaker items. However, both Marbe's law and the experimental evidence offered in its support (Osgood, 1953) have been based largely on studies of word association. Research pertaining to actual free recall of individual items appears to be limited to studies demonstrating that (a) frequently recalled English words tend to be recalled earlier than less frequent words (Bousfield, Cohen, and Silva, 1956); and (b) with lists consisting of at least 10 items, the first and last items presented tend to be recalled both more frequently and earlier than items from the middle of the list (Bousfield, Whitmarsh, and Esterson, 1958; Deese and Kaufman, 1957). These studies are interpreted as supporting Marbe's law, but neither measured recall order for more than a single trial. None of the variety of reported studies of free-recall learning over repeated trials has attempted to verify this relationship of recall order to item strength within a single list, where strength is indexed by the number of previous trials on which the item has been recalled correctly.

During the analysis of three separate and otherwise unrelated free-recall experiments concerned with quite different problems, each of the present authors has independently found a marked tendency for new or previ-

[1] This report was prepared during the authors' joint tenure at the Institute of Human Learning, University of California, Berkeley, and was supported in part by Public Health Service research grants M-05769-02 and HD-00929-03 to the first author, by a Public Health Service predoctoral fellowship to the second author, and by a National Science Foundation grant to the Institute of Human Learning.

ously incorrect items to be recalled earlier than previously correct items on a given trial, and for recall rank to decrease (become later in the sequence) with increasing number of previously correct recalls, in direct opposition to Marbe's law. The details of procedure and results of these experiments are reported fully elsewhere (Allen, 1964; Battig, Merikle, and Schild, 1965; Jensen, in press).

Method

Experiment I

The materials for Exp. I (Allen, 1964) were CVC trigrams of medium (47–53%) Glaze association value. Six different lists were constructed, representing all possible combinations of three list lengths (10, 15, or 20 items) with two levels of intralist similarity. Low-similarity lists involved minimal duplication of letters within and across trigrams. In the high-similarity lists, the same consonant was repeated 3–4 times as either the first or third letter, but no letter appeared in both positions. Each list was learned by a different group of 12 *S*s, all 72 of whom were volunteers from undergraduate psychology courses at the University of California. Each trial began with the successive presentation of all items by automatic slide projector at a 5-sec rate, after which *S* was given a blank sheet of paper and allowed 2 min to write down as many of the items as possible. All *S*s received 10 trials, and each item was presented in an unsystematically different serial position on each trial.

Experiment II

In Exp. II (Battig *et al.*, 1965), the materials were also lists of trigrams. However, each trigram represented a letter-order rearrangement (anagram) of a common three-letter English word. Each of 60 paid summer-session students or affiliates of the University of Virginia learned one of 20 lists to a criterion of one errorless trial or a maximum of 15 trials. The same 15 three-letter combinations, no two of which contained more than one letter in common, were used in all lists. The 20 lists differed only with respect to the particular combination of letter-order rearrangements included therein, each of the five non-word rearrangements of each trigram being used for an equal number of *S*s. Each trial began with the presentation of the 15 trigrams by automatic slide projector at a 3-sec rate, with serial order being varied unsystematically from trial to trial. After each presentation series, *S* was allowed 1 min to write down as many of the trigrams as he could remember. Half of the *S*s were required to reproduce the letter-order of each trigram as presented. The other 30 *S*s were permitted to give the three letters in any order, and most of these responses (73.1%) represented the word solution.

Experiment III

The materials in Exp. III (Jensen, in press) were specially chosen to obtain lists of words of maximal familiarity. The 40-item recall lists were made up from a pool of 100 common nouns having the following characteristics: (a) no fewer than three nor more than six letters; (b) Thorndike-Lorge (1944) frequencies of at least 100 per million; and (c) correct spelling by at least 90% of eighth graders, according to the Iowa Spelling Scale.[2]

The experimental treatment was administered to 123 undergraduate students in an introductory course in educational psychology at the University of California, in four groups of approximately 30 *S*s each. Different but partially overlapping lists of 40 words drawn from the pool described above were used for each group. Since the procedure was otherwise identical for all groups, and no appreciable differences between groups were found in any of the recall measures, the data of all groups were combined for the present analysis.

On each trial, the 40 words were presented successively by an automatic projector at a 2-sec rate; each word appeared for 1 sec and was followed by a 1-sec interval before the appearance of the next word. Immediately after exposure of the last word, the room lights were turned up, and the *S*s were asked to write as many of the words as they could remember, regardless of order. Four min were allowed for the written recall, after which *S*s were told to place their answer sheets in a Manila envelope, to prevent any extraneous cues on subsequent trials. The serial order of the items was unsystematically varied on each of the six trials. The first three trials consisted of the same 40 items, but on Trial 4, half (20) of the words were discarded from the list and 20 "new" words were added (from the pool of 100 words), unsystematically intermixed with the "old" words. This revised list was then presented on Trials 4–6, each time with the items in an unsystematically different order.

Results

Recall Ranks of Newly Learned Items

All three experiments revealed a consistent tendency for new or previously incorrect items to be recalled prior to previously correct items.

[2] Green, H. A. *The new Iowa spelling scale.* Iowa City, Iowa: State Univer. of Iowa, Bureau of Educational Research and Service.

In Exp. I, this comparison was based upon a *standard recall-rank* score, obtained for each item on each trial by determining the difference of its recall rank from the median (assigning positive values to items recalled before the median, and negative values to items recalled thereafter), and dividing this difference by the standard deviation of the total number of recall ranks on that trial in order to obtain comparable measures for the various list lengths. The standard recall ranks summed over the 10 trials for all newly recalled items were found to be positive for 58 of the 72 *S*s, and the mean of this measure based on all *S*s (2.89) was significantly above zero, $F(1, 66) = 60.67$, $p < .001$.

The recall rank measure in Exp. II differed from that in Exp. I only in that the difference between the item and median recall rank was not divided by the standard deviation, since all lists were of equal length. This measure revealed that items not recalled correctly on the previous trial were recalled on the average earlier than the median item for 52 of the 60 *S*s, and that the mean recall rank per item (1.06) was significantly higher than the median (zero) rank, $F(1, 54) = 60.58$, $p < .001$. Separate analyses of items given correctly for the first time (0.99) and of previously correct items that had been missed on the immediately preceding trial (1.15) revealed the priority of free recall to be comparable in magnitude for both types of items. Even though all but one of the *S*s failing to show recall priority for previously incorrect items were in the group allowed to reproduce the items in any letter order, the mean recall ranks of these items were identical for the two groups.

For each *S* in Exp. III, the mean absolute recall rank within each trial was computed separately for items newly recalled for the first time, and for items that had been recalled correctly on one or more previous trials. Once again the new items were found to be recalled consistently earlier than previously correct items, and the difference averaged over trials (2.32) was highly significant, $F(1, 122) = 93.49$, $p < .001$.

Effect of Number of Previous Correct Recalls

Although the most striking differences in recall rank were between items that had and had not been recalled correctly on one or more previous trials, there were indications in all three experiments of a continued although lesser trend toward later recall as number of previously correct recalls increased beyond one. Shown in Fig. 1, separately for the high- and low-similarity conditions in Exp. I, are the mean recall ranks of the first, second, and third correct recalls for all items that were recalled correctly on at least three

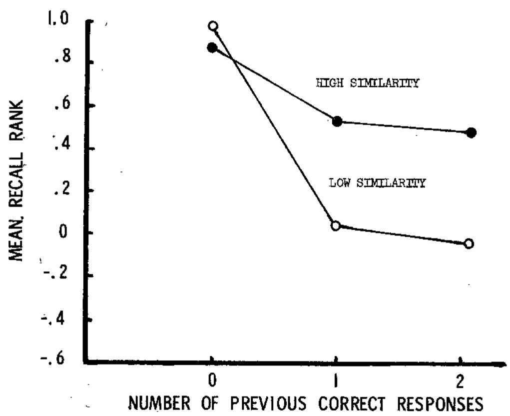

FIG. 1. Mean recall rank (difference of item from median recall rank) as a function of number of previous correct responses, for high- and low-similarity conditions of Exp. I.

trials (based on 95% of all items). The overall decrease toward the median with increasing number of previous recalls was highly significant, $F(2, 2036) = 11.71$, $p < .001$. The effect was more pronounced for low-similarity lists as indicated by a significant difference between slopes of the high- and low-similarity curves in Fig. 1, $F(1, 1018) = 4.30$, $p < .05$. There was also a marginally significant overall difference between the two similarity conditions, $F(1, 1018) = 3.67$, where 3.85 is required for significance at the .05 level. This similarity difference can probably be attributed to the greater difficulty *S*s had in identifying and distinguishing between previously correct and incorrect items when these are highly similar to one another, and a consequent attenuation of the priority of recall of the latter. The results in Fig. 1, however, were not significantly affected by list length.

The tendency for recall ranks to either remain relatively stable or to decrease with increasing number of correct recalls is depicted more clearly in Fig. 2, which presents for Exp. II the mean recall rank for all items over the

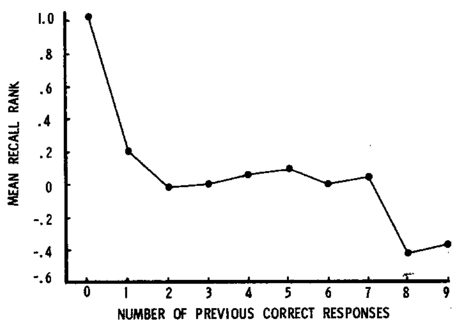

FIG. 2. Mean recall rank (difference of item from median recall rank) as a function of number of previous correct responses, for all items of Exp. II.

range of 0–9 previous correct responses. As in Fig. 1, there is a consistent decrease in recall rank from 0–2 previous recalls, along with an additional secondary drop for 8–9 recalls.

In Exp. III, mean differences between new and previously recalled items were tabulated separately for each trial, rather than as a function of number of previously correct recalls. Since the latter must inevitably increase systematically on later trials, the finding of a consistent and significant increase across trials in magnitude of mean differences between recall ranks of new and previously correct items, which yielded $F(4, 488)$ of 10.68 ($p < .001$), is consistent with the trends shown in Fig. 1 and 2. These mean differences, for Trials 2–6 respectively, were 0.17, 2.22, 2.54, 2.78, and 3.88.

DISCUSSION

In view of the consistency of the results over the wide range of materials and conditions employed in the present experiments, the tendency for new or previously incorrect items to be recalled prior to previously correct items would appear to be a phenomenon of wide generality. The direction of recall priority is diametrically opposed to that expected on the assumption that order of recall provides a direct index of item strength (e.g., Bousfield *et al.*, 1956, 1958; Osgood, 1953), and would seem to demonstrate conclusively the paramount importance in free-recall performance of factors other than response strength. Moreover, theories based upon reinforcement as the primary determinant of performance appear to be totally irreconcilable with the present results, inasmuch as the first responses to be made tend to be those that have not occurred previously and therefore could not have been previously reinforced.

The continued decline in order of recall with increasing number of previous correct recalls has probably been underestimated somewhat both in Exps. I and II. The demonstration in Exp. II of recall priority of previously correct items that were incorrect on the immediately preceding trial, which was at least as great as that observed with items given correctly for the first time, indicates that the inclusion of such items in Figs. 1 and 2 must have substantially attenuated the differences displayed therein. Since most of these items would fall within the range of 2–7 previous correct recalls where no apparent change is shown in Fig. 2, it is likely that this constancy can be at least partially accounted for by the recall priority of such items. Moreover, the magnitude of the decline from positive ranks toward zero, as displayed in Figs. 1 and 2, would necessarily have been enhanced had the ranks been corrected for the systematic increase on later trials in total number of recalled items, and consequent inflation of the possible range of rank scores, as had been done in Exp. I through the computation of standard recall ranks to adjust for similar effects produced by list length. In Exp. III, the increase in mean differences over trials is probably at least partially attributable to the aforementioned source of bias.

Since none of the present experiments had been designed to investigate the recall-priority phenomenon, its elucidation and explanation will clearly require further research. However, certain of the present findings are suggestive that *S*s develop a "strategy" in free-recall learning situations whereby items not yet given correctly receive special emphasis or attention by *S*, both during item presentation and attempted recall. Alternatively or in addiaddition, *S*s may attempt to give these previously incorrect items first, before these can be lost or forgotten, and prior to previously correct items, which have been sufficiently well learned to resist the interference and consequent forgetting produced by interpolated attempted recall of other items (e.g., Tulving and Arbuckle, 1963).[3] The effect may also reflect a general tendency for *S*s to group or categorize items together on the basis of their state of learning or level of learning difficulty (e.g., Fallon and Battig, 1964).

Previous studies have demonstrated recall order to depend on serial order of presentation, such that the first items to be recalled tend to come primarily from the end and secondarily from the beginning of the list with sequentially unstructured materials (Bousfield *et al.*, 1958; Deese and Kaufman, 1957). Both in Exps. I and III, the method of varying serial order from trial to trial insured that no item could be presented in a given serial position on more than one trial, thereby producing a possible systematic bias favoring the appearance of previously incorrect items in favored serial positions. To evaluate the role of this factor in the present results, the percent of newly recalled items presented in each serial position was determined for Exp. I, and found to be relatively consistent across serial positions, except for a slightly greater frequency of such items in the initial and final positions. That serial-position differences can be ruled out as a major determinant of the recall priority effect in the present experiments is further indicated by the closely similar results of Exp. II, wherein the unsystematic procedure for determining serial order of presentation precluded any bias against repetition of items in the same serial position.

[3] Underwood (personal communication, 1964) has referred to the recall-priority effect as "frosting-on-the-cake," suggesting an analogy between eating the cake first while saving the frosting until last, and recalling the more troublesome difficult items first while saving the easy ones for last.

References

Allen, M. The effect of formal similarity on three measures of retention. Unpublished doctoral dissertation, Univer. of California, 1964.

Battig, W. F., Merikle, P. M., and Schild, M. E. Anagram free-recall and recognition learning, and paired-associate transfer. *J. verb. Learn. verb. Behav.*, 1965, **4**, 44-52.

Bousfield, W. A., Cohen, B. H., and Silva, J. G. The extension of Marbe's Law to the recall of stimulus-words. *Amer. J. Psychol.*, 1956, **69**, 429-433.

Bousfield, W. A., Whitmarsh, G. A., and Esterson, J. Serial position effects and the "Marbe effect" in the free recall of meaningful words. *J. gen. Psychol.*, 1958, **59**, 255-262.

Deese, J., and Kaufman, R. A. Serial effects in recall of unorganized and sequentially organized verbal material. *J. exp. Psychol.*, 1957, **54**, 180-187.

Fallon, D., and Battig, W. F. Role of difficulty in rote and concept learning. *J. exp. Psychol.*, 1964, **67**, 85-88.

Jensen, A. R. An "adjacency" effect in free recall. *Quart. J. exp. Psychol.*, in press.

Osgood, C. E. *Method and theory in experimental psychology*. New York: Oxford, 1953.

Thorndike, E. L., and Lorge, I. *The teacher's word book of 30,000 words*. New York: Columbia University, 1944.

Tulving, E., and Arbuckle, T. Y. Sources of intratrial interference in immediate recall of paired-associates. *J. verb. Learn. verb. Behav.*, 1963, **1**, 321-334.

(Received March 26, 1964)