METHODS & DESIGNS

The statistical evidence for negative transfer in part-whole free recall*

MICHAEL S. HUMPHREYS† and ROBERT M. SCHWARTZ
University of British Columbia, Vancouver, British Columbia, Canada

A series of studies on part-whole free recall led to the conclusion that learning part of a list before learning the entire list produces negative transfer late in learning. The statistical evidence for this conclusion is shown to depend upon assumptions about (1) the asymptotic level reached and (2) the relative magnitude of the variance between conditions as compared to the variance of Ss within conditions. Evidence concerning these assumptions is reviewed, and it is argued that there was insufficient evidence to support a conclusion of negative transfer in part-whole free recall.

In part-whole free recall, two groups of Ss learn two lists. Generally, both groups learn an identical second list which contains twice as many words as the first list. For the experimental group, all of the first-list words are contained in the second list; for the control group, none of the first-list words is contained in the second list. Typical second-list learning curves for experimental and control groups are simulated in the upper part of Fig. 1. The curves show a superiority for the experimental condition on the first few trials. However, the learning curve for the control condition, as in much of the published data (e.g., Tulving, 1966; Bower & Lesgold, 1969; Novinski, 1969), actually surpasses, but is not significantly greater than, that of the experimental condition on the later trails.

Observation of second-list learning curves such as those shown in the upper part of Fig. 1 has led to the conclusion that part-whole free recall is a negative transfer paradigm. For instance, Tulving (1966) stated that "learning of part of a list prior to the learning of the whole list retards the acquisition of the whole list [p. 196]." The statistical tests which have been used to support a conclusion of negative transfer involve showing a higher rate of learning by control Ss and failing to show an overall difference between experimental and control Ss. Another aspect of the data which supports a conclusion of negative transfer is the superior performance by the control condition on the later trials.

The usual interpretation of negative transfer in part-whole free recall has involved the assumption that part-list organization may be inappropriate for organization of the whole list (see, e.g., Tulving, 1966). Recent studies (Slamecka, Moore, & Carey, 1972; Schwartz & Humphreys, in press), however, have questioned the relevance of these findings to the role of organization in free recall. In this paper, a further question involving part-whole free recall is raised. That is, do the part-whole free recall data support a conclusion that the experimental condition shows negative transfer?

There are two problems in determining whether the experimental condition shows negative transfer. The first problem involves the tendency for the control condition to perform slightly better on the later trials of second-list learning. That is, can one distinguish learning curves which start slightly apart and then cross over from learning curves which start slightly

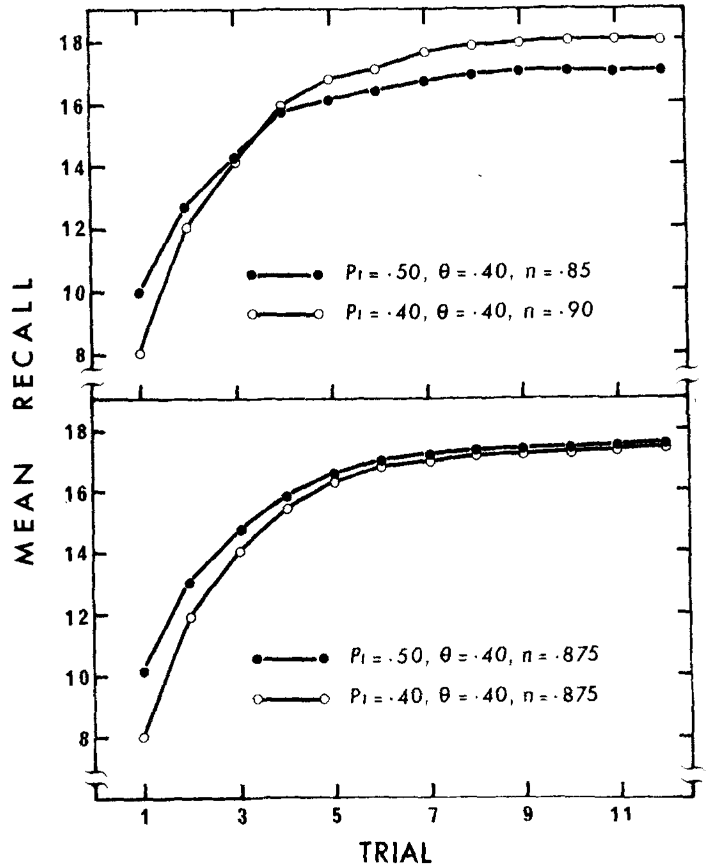

Fig. 1. Theoretical curves from statistical learning theory, simulating possible part-whole effects.

*This research was supported by Grant APA 337 from the National Research Council of Canada. We thank G. J. Johnson for helpful criticisms and suggestions and Clyde Curry for his assistance with the simulations.

†Requests for reprints should be sent to M. S. Humphreys, Department of Psychology, Northwestern University, Evanston, Illinois 60201.

apart and then approach a common asymptote (see the upper and lower sets of curves in Fig. 1)? The second problem involves the tendency for the control group to learn the second list at a faster rate than the experimental group. That is, can one distinguish faster learning rate (i.e., a faster approach to asymptote) for the curve which starts at the lower level from the tendency for the amount learned to be proportionate to the amount remaining to be learned (i.e., the typical negative acceleration of learning curves)?

As both problems involve performance at asymptote, it is important to determine what the asymptotic level is in free recall learning. Most authors have ignored the question of whether there is a constant learning rate over the trials of the free recall experiment or indeed whether performance will ever become perfect. In the present paper, however, it is argued that the asymptote is less than perfect, i.e., the probability correct at asymptote is less than 1.0. It is also argued that individual differences are not negligible over the trials when group learning curves appear to be at asymptote. It is then shown that, under these conditions, it is difficult to discriminate between negative transfer and an absence of strong positive transfer. Also, it is shown that these conditions affect estimates of learning rates.

ASYMPTOTIC LEVEL IN FREE RECALL

In this section, evidence is presented for two arguments. First, it is argued that late in learning, free recall curves appear to stabilize at less than perfect performance. Second, it is argued that late in learning, the relative magnitude of individual differences within conditions as compared to differences between conditions shows a marked increase. Two points which relate to these arguments should be clarified. First, there is no contention that the apparent stability of the free recall curves late in learning could be shown to meet statistically acceptable criteria of an absolute asymptote. Second, the idea that free recall learning curves stabilize at less than perfect performance is not meant to have profound implications about the learning process. It may simply be the case that processes such as boredom and fatigue late in learning counteract the effects of practice.

The published data on the part-whole effect clearly indicate that the average learning curve has not, under the experimental conditions employed, approached an asymptote of 1.0. The curves may be rising at the end of the experiment, but they are certainly rising at a reduced rate. For example, in Tulving's (1966) first part-whole experiment, Ss were given eight trials on a 36-item second list. Learning curves increased over the eight trials but did not exceed a mean recall of 27 items. In Tulving's (1966) second part-whole experiment, Ss were given 12 trials on an 18-item list. The experimental group reached a level of nearly 14 correct on the fifth trial but showed very little improvement for the next seven trials. Tulving (1966) suggested that Ss in the experimental condition "might have had some real trouble ever reaching perfect performance [p. 196]." In the same experiment, the learning curve for the control condition showed little increase late in learning and barely exceeded 16 correct on the 12th trial. In fact, there was a decline in mean recall on Trials 9-11, with improvement on Trial 12.

In their replication of Tulving (1966), Bower and Lesgold (1969) used a 32-item second list which was presented for six trials. The learning curves for both experimental and control groups showed very slow growth over the last three trials, and mean recall never exceeded 24 items. Similarly, in the replications by Novinski (1969), Wood and Clark (1969), and Ornstein (1970), second-list learning curves were characterized by little or no improvement over the last few trials, with performance on these trials being substantially less than perfect. Thus, under the conditions used (18- to 36-item lists, 6-12 trials, 1- to 2-sec presentation rates), group learning curves have not reached an asymptote of 1.0.

There are also data which suggest that there may be substantial individual differences in the asymptotic level reached after 8-12 trials on lists of lengths similar to those used in part-whole experiments. For example, Shapiro and Bell (1970) presented Ss a 20-item list for 12 learning trials. Then they divided Ss on the basis of organization scores (see, e.g., Bousfield & Bousfield, 1966) into groups of high, medium, and low organizers. Learning curves for all three groups appeared to have stabilized late in learning; however, asymptotic level for the three groups was different. High organizers reached an asymptote of approximately 18 words, while low organizers reached an asymptote of only 13-14 words; asymptotic level for medium organizers was between that of high and low organizers. Also, the differences between high, medium, an;low organizers increased from the 1st to the 12th trial.

A study conducted in the authors' laboratory presents further evidence for individual differences late in learning. Ss were given 12 trials on one list of 16 items and then given 12 trials on an unrelated list of 16 items. Presentation time was 1.5 sec per item; recall was written, and Ss were allowed 60 sec in which to record their recall. The rank order correlation between the number correct on first and second lists was significant, rho = .87, $p < .01$. This significant correlation suggests the existence of stable individual differences. In Fig. 2, mean second-list learning curves for the three fastest and three slowest first-list learners are shown. As Fig. 2 indicates, there has been very little reduction in the difference between the two groups over the 12 trials. Thus, this study, along with Shapiro and Bell's (1970) study, suggests individual differences in asymptotic level after 8-12 free recall trials.

THE IMPORTANCE OF THE CROSSOVER

The failure to find a main effect of conditions, even when the experimental group has an initial advantage, does not necessitate the conclusion that the experimental group shows negative transfer late in learning. First, the argument will assume that experimental and control groups approach the same asymptote. Then, the reliability of the crossover will be examined.

Along with the main effect of conditions, the components of variance in part-whole free recall include the effect of trials and the Conditions by Trials interaction. The variance of Ss within conditions serves as the error term for the conditions effect, and the variance of Trials by Ss within Conditions serves as the error term for the effect of trials and the Conditions by Trials interaction. If it is assumed that both groups approach a common asymptote which is less than 1.0 and that there are stable individual differences in asymptote, then the effect of additional trials will reduce the likelihood of observing a significant main effect of conditions. That is, each additional trial will contribute to the sum of squares of Ss within conditions but not to the sum of squares of conditions. The addition of trials does not change the degrees of freedom for the Ss within conditions error term, so the error term will grow without corresponding increases in the conditions sum of squares.

To make the preceding argument more clear, a part-whole experiment, in which both groups approach a common asymptote, was simulated. The exponential learning curve of statistical learning theory was used to provide trial-by-trial performance for

20 control and 20 experimental Ss learning a 20-item second list. For all Ss, the learning parameter, θ, was set equal to .50. To create individual differences at asymptote, π, half of the Ss in each group were given asymptotes of .75 and the other half of the Ss in each group were given asymptotes of .95. For the experimental Ss, those with an asymptote of .95 were given an initial value, p_1, of .60 and those with an asymptote of .75 were given an initial value of .40. Ss in the control condition with an asymptote of .95 were given initial values of .40, while those with asymptotes of .75 were given an initial value of .20. That is, the experimental Ss should be getting an average of 8-12 items on the first trial, while the control Ss should be getting an average of 4-8 items. Average performance at asymptote, for both conditions, should be 17 items, with half of the Ss receiving 15 items and half 19 items.

Learning curves from a typical simulation, graphed for all four types of Ss and averaged for both groups, are shown in Fig. 3. In Table 1, the results for analyses of variance computed on the first 3, 6, 9, and 12 trials are shown. The significant main effect of conditions, present when three trials are analyzed, is no longer present when nine or more trials are analyzed. The inclusion of additional trials beyond six had only slight effects on the significance of the trials variable and the Trials by Conditions interaction. Thus, although the experimental group had a large initial advantage and learned at the same rate to the same asymptote as the control condition, no main effect was found when there were nine or more second-list trials.

In the preceding paragraph, it was shown that if both groups approached a common asymptote, a failure to observe overall positive transfer does not support the conclusion that learning in the experimental condition was retarded in comparison to learning in the control condition. However, the finding that the control group learning curve surpasses that of the experimental group may indicate that the two groups do not approach a common asymptote. This possibility may be questioned for two reasons: First, to the authors' knowledge, no study has reported superiority of the control group on later trials to be statistically significant, probably because the differences are slight and statistical analysis would require a post hoc decision as to which trials to analyze. Second, such superiority might be observed because of the possibility of individiaul differences in the asymptotic level reached. If there are individual differences in asymptote, one would expect, by chance, small differences late in learning favoring one of the conditions. Observing no differences between groups on part-list learning would not be an adequate control, because it is possible that a difference in asymptotic level observed during the learning of the longer second list would not be observed during the learning of the shorter part list. Again,

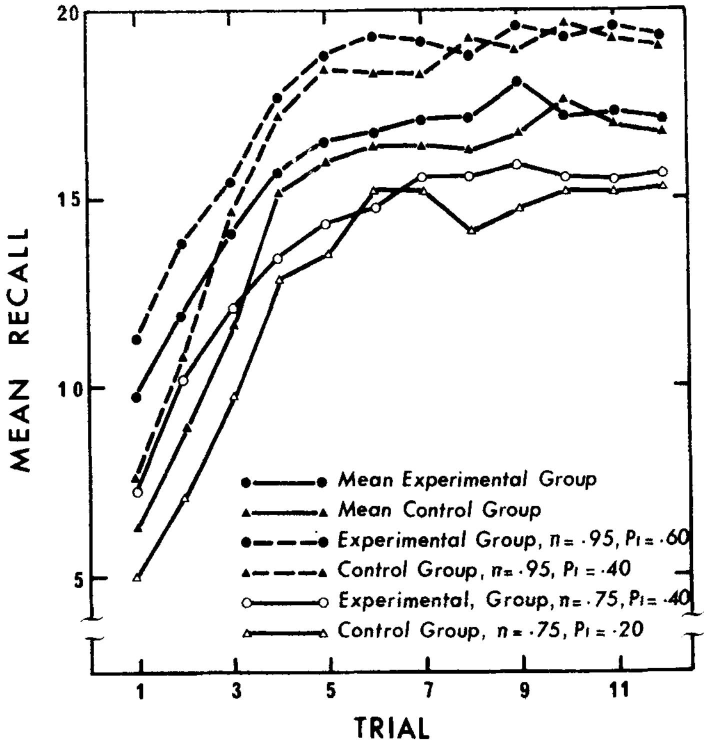

Fig. 3. Simulated part-whole effect for two types of Ss.

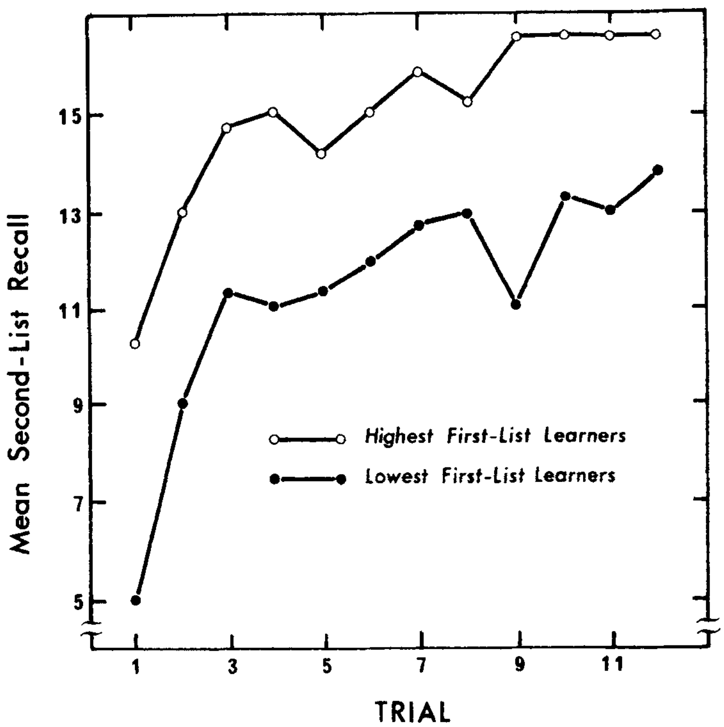

Fig. 2. Second-list performance for the three fastest and three slowest first-list learners.

Table 1
Analyses of Variance of a Simulated Part-Whole Experiment as a Function of the Number of Trials Analyzed

Trials	Source	Sum of Squares	Mean Square	df	F
1-3	Groups	252.3	252.3	1	13.7
	Trials	646.3	323.2	2	82.7
	Groups by Trials	2.1	1.1	2	.3
	Ss/Groups	699.0	18.4	38	—
	Trials by Ss/Groups	296.8	3.9	76	—
1-6	Groups	145.7	145.7	1	4.6
	Trials	2554.9	511.0	5	151.2
	Groups by Trials	111.8	31.7	5	6.6
	Ss/Groups	1202.9	22.4	38	—
	Trials by Ss/Groups	642.0	3.4	190	—
1-9	Groups	141.9	141.9	1	3.3
	Trials	3599.1	449.9	8	150.9
	Groups by Trials	130.2	16.3	8	5.5
	Ss/Groups	1653.2	43.5	38	—
	Trials by Ss/Groups	906.6	2.9	304	—
1-12	Groups	103.6	103.6	1	1.9
	Trials	4186.5	380.6	11	143.0
	Groups by Trials	169.1	15.4	11	5.8
	Ss/Groups	2083.6	54.8	38	—
	Trials by Ss/Groups	1112.1	2.7	418	—

lack of appropriate tests makes the possibility of individual differences in asymptote difficult to statistically eliminate.

In summary, this section has shown that the failure to observe overall superiority of the experimental group did not imply that the experimental group showed negative transfer late in learning. Rather, it is the crossover of experimental and control curves which provides the greatest evidence for negative transfer effects, although the reliability of this crossover was never established.

ASSESSING THE SLOPE OF LEARNING CURVES

Attempts to show faster rates of second-list learning for control Ss are based on unproven and possibly false assumptions. First, Tulving's (1966) use of slope values probably reflects initial values more than learning rates. Second, Bower and Lesgold's (1969) use of estimates of the learning rate parameter, θ, creates distortions if the asymptote is less than 1.0.

Tulving's (1966) principal reason for concluding that part-list learning retards whole-list learning was that the experimental group learned at a slower rate than the control group. To measure learning rate, Tulving determined the slope of the best-fitting line for each experimental and control S and showed that the control group Ss had higher slopes than the experimental group Ss. Slope values appear to have been used by Tulving in preference to the Conditions by Trials interaction in an effort to reduce the effect of differences between conditions on the initial trial.

However, if the two learning curves approach a common asymptote, Tulving's procedure may, to a large extent, reflect the contribution of the initial values. That is, there would have to be no ceiling effect for this procedure to be justified. Performance in the studies reported was less than 1.0; however, there still may be a ceiling effect as the asymptote approached may be less than 1.0. Even if there were no ceiling effect, observing a higher slope for the curve which started at the lower value may simply imply that the amount learned is proportionate to the amount remaining to be learned, i.e., that learning curves tend to be negatively accelerated.

To clarify the preceding argument, refer to the simulated learning curves in Fig. 1. For both sets of curves, the control group's learning curve has a higher slope than that of the experimental group. However, these curves are the exponential curves of statistical learning theory, where the amount learned is proportionate to the amount remaining to be learned. The general expression for such learning curves is:

$$p_n = \pi - (\pi - p_1)(1 - \theta)^{n-1}, \quad (1)$$

where p_n is the proportion recalled on Trial n, p_1 is the initial value, θ is the learning rate, and π is the asymptote. The difference in slope is the most pronounced for the curves in the upper part of Fig. 1, where there is a slight crossover as well as a difference in the initial value, but no difference in the learning rate parameter. There is still a marked difference in slope for the two curves in the lower part of the figure, and here the only difference is in the initial value. Thus, the negative acceleration of these learning curves and the initial superiority of the experimental condition leads to the observation that the control condition has a higher slope.

Bower and Lesgold (1969) recognized that a difference in slope did not necessarily reflect a difference in learning rate. To adjust for starting value, they fitted the exponential curve of statistical learning theory to their data. This procedure probably would be justified if the asymptotic level were 1.0. However, as the following proof shows, the value of θ is underestimated when the asymptote is less than 1.0, and the magnitude of the underestimation is directly related to the initial value.

Assume that the common asymptote, π, in experimental and control conditions is strictly less than 1.0. Also assume that θ is the same for both conditions and that the initial value for the control condition, $P_{C,1}$, is strictly less than the initial value for the experimental condition, $P_{E,1}$. The estimate of the learning rate parameter for the control condition, $\hat{\theta}_C$, and the experimental condition, $\hat{\theta}_E$, based on one learning trial and assuming an asymptote of 1.0 are:

$$\hat{\theta}_E = \frac{P_{E,n} - P_{E,n-1}}{1 - P_{E,n-1}}$$

and

$$\hat{\theta}_C = \frac{P_{C,n} - P_{C,n-1}}{1 - P_{C,n-1}} \quad (2)$$

The proportions correct on Trial n for the control condition, $P_{C,n}$, and the experimental condition, $P_{E,n}$, are:

$$P_{E,n} = (\pi - P_{E,n-1})\theta + P_{E,n-1}$$

and

$$P_{C,n} = (\pi - P_{C,n-1})\theta + P_{C,n-1} \quad (3)$$

Substituting Eq. 3 in Eq. 2 and simplifying:

$$\hat{\theta}_C = \frac{(\pi - P_{C,n-1})\theta}{1 - P_{C,n-1}}$$

and

$$\hat{\theta}_E = \frac{(\pi - P_{E,n-1})\theta}{1 - P_{E,n-1}} \quad (4)$$

To prove that $\hat{\theta}_E < \hat{\theta}_C$ note that:

$$P_{C,n-1} < P_{E,n-1} \quad (5)$$

$$-(1 - \pi)P_{C,n-1} > -(1 - \pi)P_{E,n-1}$$

$$(\pi + P_{C,n-1}P_{E,n-1}) - (1-\pi)P_{C,n-1}$$

$$> (\pi + P_{C,n-1}P_{E,n-1}) - (1-\pi)P_{E,n-1}$$

$$(\pi - P_{C,n-1})(1 - P_{E,n-1})$$

$$> (\pi - P_{E,n-1})(1 - P_{C,n-1})$$

$$\frac{(\pi - P_{C,n-1})}{1 - P_{C,n-1}} > \frac{(\pi - P_{E,n-1})}{1 - P_{E,n-1}}$$

$$\frac{(\pi - P_{C,n-1})\theta}{1 - P_{C,n-1}} > \frac{(\pi - P_{E,n-1})\theta}{1 - P_{E,n-1}} \quad (6)$$

Substituting Eq. 4 in Eq. 6:

$$\hat{\theta}_C > \hat{\theta}_E$$

Thus, the preceding proof has shown that the estimate of θ based on the amount of improvement from Trial n − 1 to Trial n is smaller for the condition with the highest initial value. Since this holds true for all n, it would also hold for an estimate which is based on the average amount of improvement over trials in a learning experiment.

To illustrate the magnitude of this effect, assume an experimental and control condition with equal asymptotes, π = .75, and equal learning rates, θ = .50. Also assume that the experimental condition starts with a higher initial value, $P_{E,1}$ = .50, than the control condition, $P_{C,1}$ = .25. The expected estimate of learning rate, based on improvement from Trial 1 to Trial 2 for the experimental condition is $\hat{\theta}_E$ = .25, while that for the control condition is $\hat{\theta}_C$ = .33. Thus, the incorrect assumption that π = 1.0 produced a difference in the estimates of the learning rates.

To summarize this section, it was shown that the measures of learning rate which have been used to support the contention of negative transfer for the experimental group in part-whole free recall are possibly inadequate. First, using slope for an estimate of learning rate does not eliminate the effect of different starting values. Second, the use of the learning rate parameter from statistical learning theory depended on the assumption that the asymptote was 1.0.

CONCLUSION

The present critique of the statistics used to demonstrate the part-whole effect does not disprove Tulving's (1966) hypothesis of inappropriate organization, nor is it contradictory to other lines of evidence concerning the role of organization in free recall. It does, on the other hand, indicate that the current evidence cannot eliminate the possibility that in part-whole free recall, the two learning curves start apart and then approach a common asymptote of less than 1.0. In addition, a conclusion of negative transfer based on slope values or estimates of learning rate without taking into consideration the asymptote is shown to be unjustified. A conservative conclusion would be that there is a lack of strong positive transfer in part-whole free recall.

While the present paper has concentrated on part-whole free recall, it is likely that the problems of inference which occur when negatively accelerated learning curves approach a common asymptote of less than 1.0 are present in other transfer situations; that is, in a transfer situation, Ss typically learn an identical second list and differ only in the conditions of first-list learning. When differences between individuals do not converge as rapidly as differences between conditions, transfer might be observed only on the first few trials.

REFERENCES

BOUSFIELD, A. K., & BOUSFIELD, W. A. Measurement of clustering and of sequential constancies in repeated free recall. Psychological Reports, 1966, 19, 935-942.

BOWER, G. H., & LESGOLD, A. M. Organization as a determinant of part-to-whole transfer in free recall. Journal of Verbal Learning & Verbal Behavior, 1969, 8, 501-506.

NOVINSKI, L. Part-whole and whole-part free recall learning. Journal of Verbal Learning & Verbal Behavior, 1969, 8, 152-154.

ORNSTEIN, P. A. Role of prior-list organization in a free recall transfer task. Journal of Experimental Psychology, 1970, 86, 32-37.

SCHWARTZ, R. M., & HUMPHREYS, M. S. List-differentiation in part-whole free recall. American Journal of Psychology, in press.

SHAPIRO, S. I., & BELL, J. A. Subjective organization and free recall: Performance of high, moderate, and low organizers. Psychonomic Science, 1970, 21, 71-73.

SLAMECKA, N. J., MOORE, T., & CAREY, S. Part-to-whole transfer and its relation to organization theory. Journal of Verbal Learning & Verbal Behavior, 1972, 11, 73-82.

TULVING, E. Subjective organization and effects of repetition in multi-trial free recall learning. Journal of Verbal Learning & Verbal Behavior, 1966, 5, 193-197.

WOOD, G., & CLARK, D. Instructions, ordering, and previous practice in free-recall learning. Psychonomic Science, 1969, 14, 187-188.