TRANSFER EFFECTS IN WHOLE/PART FREE-RECALL LEARNING

ENDEL TULVING AND SHIRLEY OSLER
University of Toronto

Ss learned an 18-word list and then a 9-word list under the conditions of free recall. For experimental Ss the words in List 2 were all taken from List 1; for control Ss the words in List 2 were different from those in List 1. Trial-by-trial recall and organization of recall of List 2 were higher for control Ss than for experimental Ss, indicating negative transfer from whole-list to part-list free-recall learning. This evidence, at variance with the hypothesis that individual list items are learned independently from one another in the free-recall task, is consistent with the view that free-recall learning is dependent upon organization of items into higher-order memory units.

TWO RADICALLY DIFFERENT HYPOTHESES have been put forth by theorists concerned with free-recall learning (FRL). One, the hypothesis of independence of list items in FRL, holds that recall of a given list item is not influenced by recall or non-recall of other items in the list. Trial-to-trial increments in recall, according to the various conceptualizations of FRL subscribing to this hypothesis, are attributable to the increased strength of traces of items, to the higher degree of availability of items, or the stronger associative bonds between each item and the general experimental context in which it is learned. The independence hypothesis is one of the fundamental assumptions made by authors of mathematical models of FRL (e.g., Miller & McGill, 1952; Waugh & Smith, 1962); it has been explicitly advanced by Asch and Ebenholtz (1962) in their theoretical treatment of FRL, and it is implicit in recent analyses of processes involved in paired-associate learning. Underwood, Ekstrand, and Keppel, for instance, claim that FRL provides "the best technique for assessing factors influencing response learning independent of associative learning" (1965, p. 448).

The alternative hypothesis—the interdependence hypothesis—has been most widely accepted as a basic assumption by those psychologists who have used free recall as an analytical tool for the study of organizational processes in memory. The relevant theoretical positions have been reviewed at some length elsewhere (Tulving, 1967). One version of the theory (Tulving, 1962, 1964, 1966) holds that in the course of

1 This research was supported by the National Science Foundation under Grant No. GB-3716, and by the National Research Council of Canada under Grant No. APT-36.

2 Present address: Department of Psychology, University of Toronto, Toronto 5, Ont.
memorization of a list of words, the words are organized into higher-order memory units (S units). Recall of a list word is greatly influenced by the recall or non-recall of other words within the same S unit, although it may be independent of recall or non-recall of specific words outside the S unit of which it is a member. Trial-to-trial increments in recall are a consequence of the development of higher-order S units, reflecting the increased size of these units rather than the increase in the number of units that can be retrieved.

The independence and interdependence hypotheses are mutually incompatible. If one is correct, the other one must be wrong: either all items are recalled independently of one another or the recall of some items depends on the recall of some other items. Since the success of theoretical analyses of FRL cannot help but be influenced by the tenability of the initial assumptions underlying these analyses, the question of independence versus interdependence of items in FRL is obviously of considerable importance. While the existing evidence seems to favour the interdependence hypothesis (Tulving, 1967), the issue does not yet appear to have been settled. Further relevant evidence is therefore needed.

The present paper reports a simple experiment in whole/part FRL whose outcome has some bearing on the issue. The experiment represents the "mirror image" of a part/whole FRL experiment reported earlier (Tulving, 1966). In that experiment, subjects learned a "whole" list of 18 words following the learning of a "part" of the list (9 words). In the present experiment, they learned the 9-word part list following the learning of the 18-word whole list. The question of interest had to do with the nature and extent of transfer from such whole-list learning to part-list learning.

According to the independence hypothesis, transfer ought to be positive and quite sizable, and it ought to increase with the amount of prior learning of list items. Traces of items, their availability, or their associations with the general experimental extralist context would be strengthened during memorization of items in the first list, and the learning of the same items in the second list should be facilitated. The only thing that changes from the first list to the second is the intralist context: some of the items present in the earlier list are no longer present and need not be recalled. The removal of some of the items from the list, however, should not affect the retrievability of the remaining items, if items are retrieved independently of one another.

According to the interdependence hypothesis, on the other hand, there are no compelling reasons to expect positive transfer from whole-list learning to part-list learning. If a list to be memorized exceeds the memory
span, its mastery requires organization of individual items into higher-order S units. A higher-order S unit functions as a unit in recall in the sense that either all or none of its constituent items are retrieved by a subject. S units developed during whole-list learning are therefore inappropriate to part-list learning; their retrieval during part-list learning, under the conditions of the present experiment, would, on the average, generate as many correct part-list responses as incorrect ones. To learn the part list the subject must either modify the units formed during whole-list learning or else form completely new units. The amount and direction of transfer from the whole list to the part list would depend on the extent to which existing whole-list units are appropriate or inappropriate to part-list learning. With randomly selected words for the part list one might expect, on the basis of the results of the earlier part/whole FRL experiment (Tulving, 1966), that, if anything, the formation of new S units in the part list might be hindered by the prior organization of the words into different S units in the whole list, with a consequent difficulty of recalling the words and possible negative transfer.

Thus, while it appears that any version of the independence hypothesis would have to predict positive transfer in whole/part FRL, the amount of such transfer being directly related to the amount of whole-list learning, the interdependence hypothesis would not be inconsistent with the absence of positive transfer or even with negative transfer.

METHOD

Design

The design can be described as a $3 \times 2$ factorial transfer design. Each of six groups of Ss learned two lists, the first containing twice as many words as the second. One factor was the amount of practice—henceforth referred to as practice—with the first list: 6, 12, or 24 trials. The other factor was the nature of the first list in relation to the second—henceforth referred to as groups. For Ss in three experimental (E) groups, the first list included all the words in the second list; for the Ss in the three control (C) groups there was no overlap in the composition of the first and second lists.

Materials

Two second lists (A and B) were used. Half of the Ss in each of the six groups learned List A as their second list and the other half learned List B.

Lists A and B consisted of 9 words each, common English nouns or words that can be used as nouns, randomly selected from among those having Thorndike-Lorge (1944) C count frequencies ranging from 14 to AA. Words in List A were: agreement, conquest, costume, cup, demand, enthusiasm, ornament, pudding, settle. Words in List B were: bait, chimney, cliff, hall, name, peace, opening, spoon, survey.

The first list for all three E groups consisted of the 18 words from Lists A and B.
The first list for all three C groups consisted of an independent set of 18 nouns randomly selected from the same pool from which Lists A and B were selected.

**Subjects**

Ss in this experiment were 72 young women, first-year students at the University of Toronto. They were assigned to the three E groups and the three C groups, each consisting of 12 Ss, in the order of their appearance in the laboratory.

**Procedure**

Learning of both the first and the second list took place under the standard FRL procedure. Ss were tested individually. At the beginning of the session S was informed she would have to learn two lists of words, one after the other, and standard FRL instructions were given. Words were presented by means of a Stowe memory drum at the rate of one sec./word. In both lists, the input order of words was systematically varied both across Ss and across trials for each S.

Ss had been instructed to read each word aloud as it appeared. Recall was oral, E recording all the recalled words in writing and checking the record against the tape-recorded version of the S's recall wherever necessary. The amount of time provided to Ss for recall on every trial was 36 sec. for first lists containing 18 words, and 18 sec. for second lists containing 9 words. The interval between the last trial of the first list and the first trial of the second list was approximately two min. No information was provided to Ss in experimental groups about the relation of the second list to the first list. After the prescribed number of trials had been completed with the first list, Ss were told that they would learn a second list consisting of nine words, the procedure being very much the same as before.

**RESULTS**

**First List Learning**

The control subjects learned a different first list from that learned by the experimental ones—a shortcoming in the design of the experiment. The comparison of the performance of different groups on the first list, therefore, cannot fully serve as a basis for verifying the equality of the learning ability of the groups. Since subjects were assigned to different groups without any known systematic bias, however, such a verification is not essential. The performance of subjects in learning the first list is nevertheless of some interest in that the data obtained from the second list can be evaluated more meaningfully in the light of known facts about first list performance.

The first analysis of the data was done on the number of words correctly recalled. An analysis of variance on arcsine transformed recall scores was carried out with groups (E and C), practice (6, 12, and 24 trials), and trials (1 to 6) as treatment variables. Data from Trials 1 to 6 only were used because they were available for all six groups. The only significant F was associated with trials: $F = 147.8$, 5 and 360 df, $p < .01$. The next largest F was for practice $\times$ trials: $F = 1.08$, 10 and 360 df, $p > .25$. 
Because groups and practice were irrelevant in determining the recall scores over the first six trials, the data were pooled over these variables. The mean numbers of words correctly recalled by the 72 subjects on Trials 1 to 6 were 7.44, 10.72, 11.25, 12.32, 12.81, and 12.88, respectively. The 48 subjects who received at least 12 trials of practice on the first list had a mean recall score of 13.62 on Trial 12; and the 24 subjects who received 24 trials of practice had a mean recall score of 15.38 on Trial 24.

Extralist intrusions were few. Over the first six trials, 36 subjects in the three E groups gave a total of 34 intrusions, while the 36 subjects in the three C groups gave a total of 27 intrusions. Almost half (26 out of 61) of these intrusions occurred on the very first trial. The number of repetitions of correctly recalled words within a trial, however, was considerable. Over the first six trials, the mean number of repetitions per subject was 8.75 in the three E groups, and 9.27 in the three C groups. The mean numbers of repetitions made by all 72 subjects on Trials 1 to 6 were 0.43, 0.93, 1.61, 1.85, 2.17, and 2.02, respectively. On Trial 12 the 48 subjects made 2.33 repetitions on the average, and on Trial 24, the mean number of repetitions for 24 subjects was 1.62.

Subjective organization of recall was evaluated in terms of the ITR measure described by Bousfield and Bousfield (1966) and by Bousfield and Abramczyk (1966). This measure represents the difference between the number of obtained intertrial repetitions of ordered pairs of items (O-ITR) and the number of such repetitions expected by chance (E-ITR). Repetitions of correctly recalled words within a trial were excluded from the data in calculating the ITR measures.

For each subject a mean ITR measure was found by summing his ITR scores over five pairs of successive trials (Trials 1 and 2, 2 and 3, ... 5 and 6) and then dividing the sum by the number of pairs of trials, i.e. by five. Analysis of variance performed on the 72 ITR scores thus obtained, with groups (E and C) and practice (6, 12, and 24 trials) as treatment variables, yielded no F ratios that even approached significance. The largest F was associated with practice: F = 1.23, 2 and 66 df, p > .25. The mean ITR score for the 72 subjects was 0.43 (S.D. = 0.21).

Second List Recall

As the reader will remember, the data of primary interest in this experiment had to do with the learning of the second list. For E groups the second list constituted a part list whose learning followed the learning of a whole list, while for the C groups, the second list was simply a new list whose learning followed the learning of an unrelated longer list.

An analysis of variance on arcsine transformed recall scores, with groups, practice, lists (A and B), and trials (1 to 12) as treatment
variables, showed that significant Fs were associated with groups (F = 5.82, 1 and 60 df, p < .02), trials (F = 27.23, 11 and 660 df, p < .01), and groups × trials (F = 2.53, 11 and 660 df, p < .01). None of the other main effects or interactions approached significance at the .05 level.

The effects of groups, trials, and their interaction are shown graphically in the form of two learning curves in Figure 1. The C group in Figure 1 refers to the three combined control groups of the original design, and the E group represents the three original experimental groups. As Figure 1 shows, the E group recalled more words on the first trial than did the C group, but beginning with the second trial the recall of the C group consistently exceeded that of the E group. The highest level of recall (8.11 words) reached by the E group on the last (twelfth) trial was reached by the C group on the fifth trial.

The data depicted in Figure 1, together with the results of the analyses of variance, thus lead to the conclusion that under the conditions of the present experiment transfer from whole-list learning to part-list learning is negative. The amount of practice on the first list does not seem to influence the amount of such transfer. Subjects receiving 6 or 24 trials of practice with the whole list showed the same amount of negative transfer.

The number of extralist intrusions was again rather small. Over the 12 trials, 36 subjects in the E group gave a total of 77 intrusions. Among these, 65 represented words that had appeared in the first list. The 36 subjects in the C group gave a total of 35 intrusions, including 13 words...
from the first list. Approximately 25 per cent of all the intrusions were
given on the first trial.

Repetitions of correctly recalled words within a trial were less numer-
ous than in the first list, probably because of the reduced length of the
list. Over the first six trials, the mean number of repetitions per subject
was 2.47 in the E group, and 3.37 in the C group; over Trials 7 to 12
these figures were 3.07 and 3.70, respectively. Thus, there were, on the
average, 0.53 repetitions per subject per trial.

To assess subjective organization of recall on the second list mean ITR
scores were calculated over the 11 overlapping consecutive pairs of trials
for each subject in the same manner as was done for the data from the
first list. The resulting 72 ITR scores were subjected to an analysis of
variance, with groups (E and C), practice (6, 12, and 24 trials), and lists
(A and B) as treatment variables. The analysis yielded significant Fs for
groups \(F = 5.92, 1 \text{ and } 60 \text{ df}, p < .02\), practice \(F = 6.80, 2 \text{ and } 60 \text{ df},
p < .01\), groups × lists \(F = 5.22, 1 \text{ and } 60 \text{ df}, p < .05\), and groups ×
practice × lists \(F = 4.88, 2 \text{ and } 60 \text{ df}, p < .02\). The mean ITR score
for the E group was .37 and for the C group .72. The means for the three
levels of first-list practice (6, 12, and 24 trials) were .44, .29, and .91,
respectively. The interaction between groups and lists was attributable
to practically identical scores for the two groups on List A (means of .55
and .57 for the E and C groups, respectively) accompanied by a sizable
difference on List B (means of .19 and .72, respectively). Finally, the
three-way interaction among groups, practice, and lists was attributable
primarily to an exceptionally high ITR score on List B for the C group
subjects who had had 24 trials of practice on the first list (mean of 1.77
for that group as compared with the over-all mean of .48 for all other
groups).

Since the relations between ITR scores on the second list and the
treatment variables used in the analysis are quite complex, these data
must be interpreted with caution. The fact remains, however, that the
E group did not appear to organize the recall on the second list to the
same extent as did the C group.

DISCUSSION

The major finding of this experiment was that of negative transfer from
whole-list to part-list free-recall learning: subjects could not memorize
a list of words as efficiently after they had attempted to memorize the
same words as a part of a longer list as they could after the learning of
an unrelated but otherwise comparable list. The experiment failed to
yield evidence on the relevance of the amount of first-list practice to the
observed transfer effects. The negative transfer effects were not reliably different for 6, 12, or 24 trials of practice on the first list.

The boundary conditions under which the negative transfer effects in FRL can be demonstrated remain to be determined by future research. Length of lists, rate of presentation, amount of time provided for recall, form of recall, the nature of the material, and other factors may be found to be critical. In this connection it should be noted, however, that Garner and Whitman (1965) reported some whole/part FRL data for lists composed of four-letter nonsense words. They found no positive transfer from learning a total set to learning a subset. Because their control group did not benefit from possible warm-up and learning-to-learn effects as did the whole/part group, their conclusion is probably too conservative. With a more appropriate control group, they also might have found reliable negative transfer as we did in this experiment.

The finding of negative transfer is interesting in itself, but it is important from the point of view of the theoretical issue raised at the beginning of the paper only in so far as it clearly implies absence of positive transfer. It thus raises doubts about the adequacy of various versions of the independence hypothesis that have been more or less explicitly proposed in the psychological literature. Especially as used by mathematical modellers (e.g., Miller & McGill, 1952; Kintsch & Morris, 1965; Waugh & Smith, 1962), the independence hypothesis clearly implies that the intralist verbal context is irrelevant in determining the rate at which individual list items are learned. Changing the context, therefore, should make no difference to the ability to recall an item. The results of the present experiment clearly show that it does. Why?

Many different interpretations of the absence of positive transfer in whole/part FRL may be possible. In our opinion, however, all of these would have to be based on assumptions about the interdependence of items in FRL.

We prefer to discuss the findings of this experiment in terms of organizational processes as outlined in the introduction to this paper for two reasons. First, the concept of organization of recall already has been shown to have some heuristic value in integrating a variety of free-recall phenomena and in suggesting new experiments (Tulving, 1967). Second, there was some evidence in the present experiment that the lower recall scores of the E group on the second list were accompanied by lower organization scores when recall and organization were compared with corresponding measures of the C group. This evidence appears to be consistent with the interpretation that in learning the part list the experimental subjects must learn to reorganize the words constituting the second list. It is difficult to see how any version of the independence hypothesis
would be able to account for the differences between the two groups in subjective organization, particularly since the ITR measure is mathematically independent of the number of items recalled.

Before concluding, we would like to raise a point about one of the details of the outcome of the experiment. On the first trial of the second list, mean recall of the E group was slightly higher than that of the C group (6.56 words vs. 5.95). This finding might be construed by some as evidence for positive transfer or at least as evidence against negative transfer. Closer reflection and analysis of the data make clear, however, that such an argument would be wrong.

There is some reason to believe that the superiority of experimental subjects over control subjects on the first trial reflects nothing more than "unidentified intrusions" from the first list. The argument is as follows. A subject may recall a word from the first list while recalling words from the second list. If this word did not occur in the second list, it would be identified by the experimenter as an interlist intrusion; if it did occur in the second list, it would be given full credit as a word remembered from the second list, although in fact it is an "unidentifiable intrusion." The number of unidentifiable intrusions can be estimated from the number of identifiable intrusions. The number of both kinds of intrusion would be expected to be the same, since the number of first list words occurring in the second list in this experiment was equal to the number of first list words not occurring in the second list, and because the number of first list words the subject "knows" to have occurred in the second is probably no different from the number of first list words the subject "knows" did not occur in the second list. On the first trial, the mean number of identifiable intrusions, and hence the estimated number of unidentifiable intrusions, was 0.58, which almost exactly matches the difference between the mean number of words recalled by the two groups on that trial.

Taking into consideration the unidentifiable intrusions from the first list, therefore, we may conclude that even on the very first trial of the second list there was no evidence of positive transfer. Organizational processes are assumed to play a relatively small role on the first learning trial, although they are responsible for trial-to-trial increments in recall. Equality of recall scores of the two groups on the first trial of the second list, therefore, is not surprising.

A simple experiment such as the present one cannot be expected to settle an issue as broad and as basic as that of independence versus interdependence of items in FRL. But apart from having shed some light on the issue, the present experiment has once again demonstrated the important role played by intralist verbal context in FRL. This role can be easily overlooked in typical experiments in which the context remains
constant over trials. When it is manipulated experimentally, its effects become obvious. No theory of FRL can afford to ignore these effects.

RéSUMÉ

Apprentissage d’une liste de 18 mots, puis d’une liste de 9 mots, dans une situation de libre rappel. Pour le groupe expérimental, les mots de la seconde liste sont tous empruntés à la première; mais, pour le groupe contrôle, tous les mots sont différents. L’analyse item par item montre que le nombre d’item retenus et l’organisation du rappel sont supérieurs pour le groupe contrôle que pour le groupe expérimental, ce qui traduit un phénomène de transfert négatif de l’apprentissage de la liste globale à celui de la liste partielle. Ces données sont contraires à l’hypothèse voulant que les item individuels d’une liste s’apprennent indépendamment les uns des autres dans une tâche de libre rappel; mais elle est compatible avec l’idée que l’apprentissage par libre rappel dépend de l’organisation des item en unités mémorielles d’un ordre supérieur.

REFERENCES


(First received November 24, 1966)