JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR 13, 181–193 (1974)

On Negative Transfer: Effects of Testing One List on the Recall of Another[1]

ENDEL TULVING

University of Toronto and Yale University

AND

MICHAEL J. WATKINS

Yale University

Two experiments were designed to clarify the priority effect in the A–B, A–C paradigm, higher first-list than second-list recall in the MMFR test following equal amounts of study of both lists. The results of the experiments suggested that the priority effect is a consequence of severe impairment of learning A–C pairs when A–B pairs are not tested prior to the presentation of A–C pairs. This impairment of learning of A–C pairs, negative transfer, is independent of specific A–B pairs tested, but its overall magnitude appears to be an inverse function of the proportion of pairs in the A–B set that are tested. Since no theory of transfer or interference makes any provisions for the effects of testing of one list on the learning and recall of another, the testing effects observed in these experiments are puzzling.

In the summer of 1967, the senior author, with the collaboration of two of his students, Roy Patterson and Gerald Malis, conducted an experiment that had what then appeared to be an anomalous outcome. In the MMFR test of a two-list paired associate retroaction experiment, in which pairs of items in the two lists conformed to the A–B, A–C paradigm, recall of response members from the first list was considerably higher than recall of response terms from the second list. Because this outcome appeared to be at variance with the typical results from the A–B, A–C paradigm, it was thought to be worthy of further study. This study was undertaken by Gerald Malis who in 1970 completed his doctoral dissertation on what he called the "priority effect" in the A–B, A–C paradigm; higher recall of B than C terms in the A–B, A–C paired-associate task in which the amount of practice on the two sets of pairs, A–B and A–C, is nominally identical. He replicated the initial findings, and isolated one of the necessary conditions of the priority effect. Malis' work, however, left open certain questions about the effect. We have now completed two additional experiments that have helped to clarify the nature and conditions of the effect. These experiments are described in this paper.

Since the initial experiment was not published, and since Malis' (1970) dissertation is not readily accessible, we begin by summarizing these studies. We will also briefly describe some other experiments reported in the literature in which findings similar to the priority effect have been noted.

THE INITIAL EXPERIMENTS

The Tulving, Patterson, and Malis experiment was designed in an attempt to study retroactive and proactive interference under

[1] This research has been supported by the National Science Foundation Grant No. 24171X and National Research Council of Canada Grant APA 39. We are grateful to Olga Watkins for experimental assistance. The paper was written during Endel Tulving's residence as a Fellow at the Center for Advanced Study in the Behavioral Sciences, Stanford, California.

conditions somewhat different from those that had until that time been used in most conventional experiments. In the typical retroaction and proaction experiments (for a review, see Keppel, 1968), a relatively short list of pairs is learned to a relatively high criterion, followed by the learning of another short list to a relatively high criterion. Under these conditions, at the time of the subsequent retention test of the first list, the recall performance on the second list is high—or can be assumed to be high if recall of the second list is not tested—and this state of affairs makes it difficult to compare recall from one list in relation to the recall from the other, a comparison that at that time seemed to be relevant to the attempts to understand retroactive and proactive effects.

Therefore, in the Tulving, Patterson, and Malis experiment, subjects learned lists of 24 pairs on a single trial. To facilitate initial learning, the members of all pairs were familiar words weakly associated (less than 1% in the free-association norms). The basic design was 2 × 2 factorial, with two paradigms (A–B, A–C, and A–B, C–D) and two kinds of final retention tests (MMFR, and a non-cued test of all response members). In addition, there were four control groups: two standard single-list RI controls, one tested with stimulus members of pairs as cues, the other tested without cues, and two corresponding PI controls. There were 16 subjects in each of the eight groups. The 24 pairs in a list were presented visually, on a single trial, at the rate of 3 sec per pair. No immediate test was given for any list in any group. The two lists given to experimental groups were separated by a 30 sec unfilled interval. The final cued recall or non-cued recall test for one or two previously seen lists was given 30 sec after the presentation of the second list and at corresponding times in the control groups.

The summary of the data from the experiment, in the form of mean number of words recalled, is presented in Table 1. For our present purposes the finding of interest is represented by the low level of recall from the second list in the A–B, A–C paradigm. In both cued and non-cued recall this level was lower than that from the first list in the same paradigm, and also lower than recall from the second list in the A–B, C–D paradigm. We will refer to the higher A–B than A–C recall in the MMFR test in the A–B, A–C paradigm as the priority effect.

TABLE 1

MEAN PROPORTION OF RESPONSE WORDS RECALLED IN THE CUED AND NON-CUED TEST

	Cued Recall		Noncued Recall	
Paradigm	List 1	List 2	List 1	List 2
A–B, A–C	.48	.32	.25	.17
A–B, C–D	.44	.50	.13	.23
RI Control	.51	—	.23	—
PI Control	—	.68	—	.30

Data from the unpublished experiment by Tulving, Patterson, and Malis (1967).

Since the typical finding from conventional A–B, A–C paradigm is retroactive interference, characterized, among other things, by higher recall of C than B terms, and since in this experiment there was no evidence of the typical RI effect, the finding appeared rather puzzling.

There were several features of the experiment that distinguished it from previous work, notably the presentation of the material on only a single study trial in each list, and the fact that no immediate test of either the first or the second list was given prior to the MMFR test. In thinking about possible sources of the anomalous finding, we focused on these two features.

Briggs (1957) has reported, as far as we can tell, the only two-list retroaction experiment using the A–B, A–C paradigm in which very small amounts of original and interpolated learning were incorporated into the design. Although he did obtain numerically higher

recall from the first (21%) than the second list (16%) in the OL-2, IL-2 condition (two trials of original and two trials of interpolated learning), these data were yielded by the MFR method that does not fully reflect what the subject remembers. Furthermore, in relation to the single-list control group, the OL-2, IL-2 condition did result in a reasonably respectable amount of 37% retroactive inhibition. Finally, although subjects in the Tulving, Patterson, and Malis experiment were shown the to-be-remembered materials only on a single trial, they did recall close to 50% of the 24 pairs in the first list. Hence it did not appear that low levels of learning were the source of the anomalous A–B priority effect.

The possibility that the A–B priority was a consequence of the absence of immediate testing of the two lists, however, received some support from other experiments we discovered reported in the literature.

Priority in Short-Term Memory Experiments

In an experiment by Goggin (1966), subjects were given a single presentation of a pair of A–B, A–C lists. In each list there were just two CVC–word pairs. Since neither the A–B nor the A–C list was given an immediate test, the procedure in effect amounted to the presentation of a single list of four paired associates comprising two CVC's (A terms) and four words (B and C terms). After the presentation of such a list, the subjects engaged in a digit-reading task for an interval of six to 40 sec. Then the MMFR test was given. Contrary to the expectations based on earlier data and interference theory, the B terms were recalled better than the C terms at all retention intervals.

A similar advantage in the recall (and recognition) of B terms over C terms in an intralist A–B, A–C paradigm was found by DaPolito (1966). His study differed from Goggin's in a number of ways: the A terms consisted of distinctive pictures and the B and C terms were numbers, the lists were of a more conventional length, and no distractor task was interpolated between the presentation of the list and the recall test. However, DaPolito's experiment did share one feature with that of Goggin that now seems to be of critical importance: there was no initial testing of the presented A–B and A–C pairs, so that the MMFR test was the first and only test administered. In this test, the A–B pairs were recalled with the probability of .54 while the A–C pairs had a recall probability of .34.

In a second experiment, DaPolito used pairs consisting of monosyllabic words and numbers, and varied the number of A–B presentations within the list. In the MMFR test the B terms were recalled with a probability of .49, .73, and .82, for one, two, and three presentations respectively, while the C terms were recalled with a probability of about .30 irrespective of the number of A–B presentations. Greeno, James, and DaPolito (1971) interpreted these data as implicating interference with storage, and used them in support of their theoretical view of negative transfer and forgetting as manifestations of interference with storage and retrieval of pairs of items.

In yet another relevant study, Bruce and Murdock (1968, Experiment I) used the paired-associate probe procedure. A list was presented consisting of six word-pairs and followed by the representation of one of the A terms as a cue for the recall of the other member of the pair. In each list there were four pairs that consisted of eight unrelated and randomly paired words, and two pairs standing in an A–B, A′–C relation to one another, with the A and A′ terms acoustically similar. Bruce and Murdock found that, while recall of the B terms was equal to recall of the control items, recall of the C terms was impaired (except in the last two serial positions where performance equalled that for control items). Thus this study agrees with those of Goggin (1966) and DaPolito (1966) in showing higher recall of B than C terms. Bruce and Murdock did not offer any explanation for these results,

but we note again that the result was achieved in absence of initial testing of A–B and A′–C pairs.

The Role of Immediate Testing: Malis' Experiment

The role of immediate testing in both A–B, A–C, and A–B, C–D paradigms was investigated and reported in the unpublished dissertation by Malis (1970). He took as his point of departure the Tulving, Patterson and Malis experiment. We will here briefly consider the data primarily from his first experiment, since the findings of the second experiment essentially only replicated the findings from the first.

Malis used lists of 24 word pairs, presented at the rate of 4 sec/pair on a single trial. A short interval followed each list, and the MMFR test was given after the second interval. There were three experimental variables: (a) paradigm—the two lists conformed either to the A–B, A–C, or to the A–B, C–D paradigm, (b) material—although word-pairs were used throughout, the words within a pair were either low (less than 1 %) normative associates or were unrelated (the same low normative associates randomly paired), and (c) presence or absence of immediate testing—an immediate test of each list was given during the respective intervals following presentation, or the immediate test was not given and the interval was unfilled. These three variables were combined in a 2 × 2 × 2 factorial design with a different group of 16 subjects serving in each of the eight conditions.

The mean proportions of words recalled from the first and second lists in both the immediate and MMFR tests are shown in Table 2. These data refer to the random pairing condition; for the condition in which associated pairs were used, recall performance was generally higher but the pattern of results very similar.

Table 2 shows that in the absence of immediate testing the MMFR data replicated the earlier finding: In the A–B, A–C paradigm, recall of first-list words was more than twice as high as recall of the second-list words, while in the A–B, C–D paradigm recall of words from the second list was higher than those from the first list. When immediate tests of the two lists were given, the recall of words from the first list was lower than that from the second list in both the A–B, A–C and the A–B, C–D paradigms. The results of Malis' experiment thus demonstrate the critical part that initial testing of A–B and A–C pairs plays in determining the recallability of these pairs in a subsequent recall test such as the MMFR. They also make clear that immediate tests are critical in changing the pattern of results only for the A–B, A–C paradigm. In the A–B, C–D paradigm, initial testing of the first and second lists resulted in a higher level of recall in the MMFR test, but the presence or absence of immediate tests did not in any way interact with the first and second lists, as it did in the A–B, A–C paradigm. Because the presence or absence of immediate tests did not affect the outcome from the A–B, C–D paradigm, this paradigm was not included in the design of the two new experiments described in this paper.

Malis' experiments showed that immediate testing of both the first and second list

TABLE 2

MEAN PROPORTION OF WORDS RECALLED IN IMMEDIATE AND MMFR TESTS

Paradigm and Condition	Tests			
	Intermediate		MMFR	
	List 1	List 2	List 1	List 2
A–B, A–C				
Immediate tests	.55	.56	.41	.53
No immediate tests	—	—	.44	.20
A–B, C–D				
Immediate tests	.48	.61	.39	.56
No immediate tests	—	—	.28	.40

Data from the unpublished dissertation by Malis (1970).

produced recall data on the MMFR test that were quite different from those obtained under conditions where both lists were immediately tested. But they left unanswered the question as to the role of immediate testing of *each* of the two lists in the A–B, A–C paradigm. Malis' data did clearly suggest that when *both* lists were tested, no A–B priority effect occurred, while when *neither* was tested, the effect was strong. Logically it is entirely possible that the presence or absence of one of the two tests is critical, while that of the other is inconsequential. The first of the two experiments we report in this paper was, therefore, designed to examine recall (both in the immediate tests and in the MMFR test) in the A–B, A–C paradigm, under all four possible combinations of presence or absence of immediate tests: both lists, neither list, only the first, or only the second list tested.

Experiment I

Design

Thirty-two subjects, young adults of both sexes, were tested in eight groups of four. All groups were first given a practice task in which they saw a short A–B list, followed by an A–C list, and then were tested for both on the MMFR test. In addition, the subjects in the first four groups were given an immediate test of the A–B list and a drawing task after presentation of the A–C list. The second four groups of subjects performed the drawing task after the first list, and were given an immediate test of the A–C list.

For the two experimental lists that provided the data of interest, the four members of each group were assigned to different conditions, which were defined by the immediate testing procedure: In the BC condition both the A–B and A–C lists were tested; in the $\bar{B}\bar{C}$ condition neither list was tested; only the first list was tested in the $B\bar{C}$ condition; finally, in the $\bar{B}C$ condition just the A–C list was tested. Thus after each list had been presented, two of the subjects within a group were given an immediate recall test, while the remaining two performed the picture-drawing task. All subjects were given an unpaced MMFR test following the second-list recall or drawing task.

The first two groups of four subjects were treated in identical fashion. Treatment of the third and fourth groups was the same as the first two groups, except that those words which had served as B terms now became C terms, and vice versa. Thus while the same 3-word sets were used, allocation of words to B and C positions was balanced between groups. The remaining four groups comprised an exact replication of the first four groups, but with a completely new set of words.

Materials

Forty sets of three words were constructed from a pool of 120 concrete nouns. All words contained 3, 4, or 5 letters and had a Thorndike–Lorge frequency of 25 per million or higher. Words were allocated to sets on an essentially random basis, but with care taken to avoid obvious within-set semantic or acoustic relations. Twenty sets were used for the first 16 subjects, and 20 for the second 16 subjects. One word from each set was randomly chosen as the A term. Each A term was paired with each of the other two words and typed in left-right arrangement for presentation via a Kodak Carousel projector.

In addition to the experimental lists, a single pair of A–B, A–C practice lists, each containing five pairs of words, was prepared for presentation to all subjects.

A number of picture slides were used for the drawing task. These depicted two or three persons in various interacting situations.

Procedure

The subjects were tested in groups of four. They were told that the experiment concerned memory for both verbal and non-verbal material, and a general outline of the procedure was given. The five-item practice pair of A–B and A–C lists was then shown at a rate of 4 sec per word pair. Each subject was given

an immediate recall test of one of the lists, and a picture memory task after the other. All subjects were then given an MMFR test, in which they were to try to recall both the B and C terms for each A term. For both immediate and MMFR tests, a sheet containing all the A terms arranged in a new random order was presented. The subjects did not know whether an immediate recall test would be given until after list presentation, when they received either a response sheet containing the A terms for the recall test, or a blank sheet for the drawing task. A single picture was shown after each list for 15 seconds. The drawing task consisted of studying the picture, and then trying to draw it from memory.

The procedure for the experimental A–B, A–C lists was essentially the same as for the practice lists, except that list length was increased to 20 word-pairs, and more time was allowed for recall. The immediate recall interval was 2.5 minutes, and 3.5 minutes were allowed for the MMFR test.

Results

The mean proportions of words recalled in the immediate tests of the A–B and A–C terms, where such tests were given, are shown in the left-hand panel of Table 3. In the BC condition, where an immediate test of each list was given, the proportions of words recalled in the two lists were .58 and .48, respectively. The difference was not quite statistically reliable, $t(7) = 1.94$, $p < .10$. In the $B\bar{C}$ condition, the proportion of words recalled from the first list was .53, a value rather similar to the corresponding value in the BC condition. In the $\bar{B}C$ condition, the mean proportion of words recalled from the A–C list was .15. This mean was reliably different from the corresponding mean in the BC condition, $t(14) = 4.97$, $p < .01$.

TABLE 3

PROPORTIONS OF WORDS RECALLED IN IMMEDIATE TESTS OF LIST 1 AND LIST 2, EXPERIMENTS I AND II

Condition	Experiment 1		Experiment 2	
	List 1	List 2	List 1	List 1
BC	.58	.48	.55	.36
$B\bar{C}$	.53	—	.56	—
$\bar{B}C$	—	.15	—	.34
Mean	.55	.31	.55	.35

Thus, the data from immediate recall tests show that the presence or absence of the A–B test is highly critical for the learning of the A–C list, as judged by the immediate recall test: When the A–B test was given (condition BC) immediate recall of the A–C list was .48, when it was not given (condition $\bar{B}C$) immediate recall of A–C was .15.

As might be expected, the rather striking impairment in the recall of the A–C list under conditions where the immediate test of A–B was not given is carried over into the MMFR test. The data from the MMFR test are tabulated in Table 4, in the form of four (corresponding to the four experimental conditions) 2 × 2 contingency tables showing recall patterns of individual items from A–B, A–C triplets. Each entry in Table 4 is based on 160 observations (20 items × 8 *S*s).

In each 2 × 2 contingency table the recall and non-recall of B items is crossed with recall and non-recall of corresponding C items with marginal proportions showing unconditionalized probabilities of recall and non-recall of B and C items. The four individual 2 × 2 contingency tables in Table 4 are arranged in such a way that the two rows of tables represent presence and absence of the immediate test of the A–B list, and the two columns represent the presence or absence of the immediate test of the second (A–C) list. Inspection of the four 2 × 2 contingency tables suggests that the two tables within each row are quite similar to one another while those within each of the two columns are quite different.

Three observations about the data shown in Table 4 are of interest. First, unconditional-

TABLE 4

PROPORTIONS OF WORDS RECALLED AND NOT RECALLED FROM LISTS 1 AND 2 IN THE MMFR TEST, EXPERIMENT I.

Condition B C

List 1 Recall	List 2 Recall +	−	Total
+	.24	.26	.50
−	.20	.30	.50
Total	.44	.56	1.00

Condition B C̄

List 1 Recall	List 2 Recall +	−	Total
+	.23	.21	.44
−	.14	.42	.56
Total	.37	.63	1.00

Condition B̄ C

List 1 Recall	List 2 Recall +	−	Total
+	.07	.21	.28
−	.07	.65	.72
Total	.14	.86	1.00

Condition B̄ C̄

List 1 Recall	List 2 Recall +	−	Total
+	.05	.19	.24
−	.07	.69	.76
Total	.12	.88	1.00

ized proportions of recall of words from the A–C list were very low in the two conditions (B̄C and B̄C̄) in which the A–B list had not been immediately tested, in relation to the two conditions (BC and BC̄) in which it was. Proportions of words recalled from the two earlier conditions were .14 and .12, while from the two latter conditions the corresponding proportions were .44 and .37. The very low level of immediate recall of the A–C list in condition B̄C (.15) noted above makes clear that this low level of A–C recall in the MMFR test simply reflects the low level of original learning.

Second, the initial testing of the A–B list considerably facilitates B items in the MMFR test. In the BC and BC̄ conditions, where A–B lists were tested, the mean proportions of B items recalled in the MMFR tests were .50 and .44, respectively, while in the B̄C and B̄C̄ conditions in which the A–B list had not been immediately tested, the corresponding proportions were .28 and .24. Immediate testing of the A–C list also resulted in a small numerical superiority in the recall of C items in the MMFR test, in comparison with that from the two conditions in which the A–C list was not immediately tested, .44 and .14 versus .37 and .12, but the differences were not statistically reliable.

Third, the conditional probability of recall of C given that the corresponding B was recalled was higher, in all four experimental conditions, than the conditional probability of recall of C given that the corresponding B was not recalled. With the data pooled over all four experimental conditions the former value, P(C)|(B), was .41, while the latter, P(C)|(B̄), was .19.

Discussion

Two major conclusions emerge from these data. First, the presence or absence of the immediate test of the A–B list is highly critical in determining the level of the immediate recall of the A–C list. When the A–B test is absent, subjects apparently have great difficulty learning the A–C pairs, with the consequence that their immediate recall of the A–C list is very low, and therefore the recall of A–C pairs in the MMFR test is also very low.

The finding that the presence or absence of the immediate test of the A–B list, but not that of the A–C list, is critical in determining the superiority of A–B recall over A–C recall

in the MMFR test has clarified the priority effect observed in earlier experiments under conditions where neither A–B nor A–C list had been tested immediately. Furthermore, the present data also make it clear that the low level of recall of the A–C pairs in the MMFR test in the $\bar{B}\bar{C}$ condition seems to be attributable largely to the impairment of the original learning of the A–C pairs rather than manifesting some sort of proactive inhibition effects. The non-analytical nature of earlier experiments precluded the distinction between negative transfer and inhibition effects, as these two terms are traditionally defined.

The second major finding had to do with large facilitating effects of immediate recall of the A–B list on the recall of A–B pairs in the MMFR test. In the earlier experiments such effects had not been observed. In the Malis (1970) experiments, for instance, recall of the A–B pairs was approximately identical under BC and $\bar{B}\bar{C}$ conditions. The difference is probably a consequence of the interpolated activity in which the subjects in the $\bar{B}\bar{C}$ condition engaged between the presentation of the A–B and the A–C list in the present experiment, and absence of any such activity in Malis' experiments. Thus, the subjects could have "rehearsed" the A–B pairs prior to the presentation of the A–C list, or the memory traces of A–B pairs may have "consolidated" in the interval, while the interpolated activity used in the present experiment may have precluded such "rehearsal" or "consolidation".

We have no ready explanation for the dependency of recall between C and B terms associated with common stimulus items. It is possible that these represent nothing more nor less than item–selection artifacts: Subjects learned both B and C responses to "easy" stimulus items. While some theorists (for example, Martin, 1971, 1972) think that contingency relations between B and C items in the A–B, A–C transfer and retroaction paradigm are of critical importance in evaluating extant versions of interference theory, other theorists (for example, Postman & Underwood, 1973) have pointed out that the independence or dependence relations between individual B and C items are of no theoretical relevance. Therefore the present positive correlation between B and C recall may not be a particularly important finding.

The one obvious question to ask, given the findings from this experiment, concerns the function of immediate tests of A–B lists. The presence or absence of such testing seemed to be highly critical in determining learnability of the A–C list, but the data do not make clear whether it is the act of testing of the first list in general or the act of recall of certain tested A–B items in particular that reduces the difficulty of subsequent A–C learning. The conditional recall data shown in Table 4 do not help to answer the question because of possibility of item– and subject–specific selection artifacts (Hintzman, 1972). It seemed, therefore, that a more direct test of these relations would be useful. Experiment II was designed to determine whether the effect of A–B testing on A–C learning is pair–specific or list–specific.

Experiment II

The main difference between this experiment and the first was that the four testing conditions were made a within–subject rather than a between–subject variable. This was achieved by allocating one quarter of the A terms of a single A–B, A–C pair of lists, to each of the four conditions. Thus, with 24 word pairs per list, six A terms were presented after both lists (the BC condition), six after the first ($B\bar{C}$), six after the second ($\bar{B}\bar{C}$), and six were not used in either test ($\bar{B}\bar{C}$). The primary question of interest was whether the different effects of the immediate testing of the A–B and A–C terms found in Experiment I would be replicated. In particular, would absence of the immediate test for an A–B term serve to selectively impair learning of the corresponding C term as well as its subsequent recall in the MMFR test, as it did

in Experiment I? If so, the striking negative transfer effects found in the absence of the first–list test must be given a pair–specific interpretation; if not, the findings of Experiment I should be interpreted in terms of the more general effects of list testing.

Method

Thirty-two subjects, male and female Yale undergraduates serving for pay, were tested in small groups. All subjects were shown the same pair of A–B, A–C lists, but for half of the subjects corresponding B and C terms were interchanged. The lists contained 24 word-pairs, with the individual A–B, A–C words sets randomly sampled from the pool used in Experiment I.

The 24 A terms were randomly allocated to four groups of six words. Each of these word groups served in a given experimental condition (BC, B$\bar{C}$, $\bar{B}$C, or $\bar{B}\bar{C}$) for eight subjects, with conditions balanced between subjects according to a Latin Square design. Presentation order of the 24 pairs was separately randomized for each list. After the presentation of each list, each subject was tested with the 12 A terms from the appropriate two word groups. All subjects were given the same MMFR test.

Before testing began, the procedure was explained, and two eight-pair A–B, A–C practice lists were presented. These practice lists were structured and tested in the same way as the experimental lists. Thus, immediate recall tests of four items from each list were given, and there followed an MMFR test. The experimental lists were then presented and tested in the same fashion. Each list contained 24 word–pairs, half of which were tested immediately. The lists were presented at a rate of 4 seconds per word pair; 75 seconds were allowed for the immediate recall tests, and 3.5 min for the MMFR test.

Results

Proportions of recall of B and C terms in the immediate tests of the two lists are shown in the right-hand panel of Table 3. The mean proportion of B items recalled in the immediate test was .55 for the BC set of items, and .56 in the B$\bar{C}$ set. These two proportions were almost identical with corresponding proportions of A–B recall in Experiment I, also shown in Table 3. The mean proportion of C terms recalled was .36 in the BC condition and .34 in the $\bar{B}$C. Thus, although immediate recall of words from the A–C list was considerably lower than the immediate recall of words from the A–B list, $t(31) = 4.35$, $p < .01$, recall of C terms did not seem at all to depend on immediate testing of the A–B pairs with which the C terms shared the stimulus term.

Relating these data to those from the immediate tests in Experiment I, we can draw two conclusions. First, immediate recall of A–C pairs appears to be a monotonically increasing function of the proportion of A–B pairs tested from the first list. Thus, when the number of A–B pairs tested was 0 (condition $\bar{B}$C in Experiment I) immediate recall of A–C pairs was .15; when the number of tested A–B pairs was 12 (all conditions in Experiment II), immediate recall of A–C pairs was .35 (mean of .36 and .34); and when the number of A–B pairs tested was 20 (condition BC in Experiment I), recall of A–C pairs was .48. Indeed, the mean proportion of A–C pairs in Experiment I, with the data averaged over the two conditions in which the A–C list was immediately tested, was .31, and thus essentially indistinguishable from the corresponding mean (.35) in Experiment II. The second conclusion is that the reduced impairment in recall of A–C pairs induced by the incompleteness of testing of the A–B list does not represent a pair–specific or item–specific effect. The proportion of A–C pairs recalled in the immediate test of the second list was essentially the same for the intralist word sets in which all corresponding A–B pairs had been tested (.36) as it was for the word set in which none of the corresponding A–B pairs had been tested (.34).

The MMFR data from Experiment II are

TABLE 5

PROPORTIONS OF WORDS RECALLED AND NOT RECALLED FROM LISTS 1 AND 2 IN THE MMFR TEST, EXPERIMENT II.

Condition B C

A–B Recall	A–C Recall +	A–C Recall −	Total
+	.18	.23	.41
−	.15	.44	.59
Total	.33	.67	1.00

Condition B $\bar{C}$

A–B Recall	A–C Recall +	A–C Recall −	Total
+	.19	.29	.48
−	.07	.45	.52
Total	.26	.74	1.00

Condition $\bar{B}$ C

A–B Recall	A–C Recall +	A–C Recall −	Total
+	.09	.18	.27
−	.22	.51	.73
Total	.31	.69	1.00

Condition $\bar{B}$ $\bar{C}$

A–B Recall	A–C Recall +	A–C Recall −	Total
+	.13	.14	.27
−	.18	.55	.73
Total	.31	.69	1.00

summarized in Table 5 whose format is identical with Table 4. These data simply reflect what we already know. First, the recall of C terms was essentially the same in the two conditions (BC and B$\bar{C}$) in which the A–B pairs were tested (.33 and .26) as they were in the two conditions ($\bar{B}$C and $\bar{B}\bar{C}$) in which they were not (.31 and .31), thus extending the earlier conclusion about immediate recall to recall in the MMFR test: Recall of C terms is independent of testing or absence of testing of corresponding A–B terms. Second, the testing of specific A–B pairs produces an enhanced recall of those pairs in the MMFR test. The recall of B terms in the MMFR test for conditions BC and B$\bar{C}$ were .41 and .48, respectively, while in the two conditions where A–B pairs were not tested ($\bar{B}$C and $\bar{B}\bar{C}$), the corresponding two values were identical at .27. This finding replicates the similar finding in Experiment I. And, also replicating the same finding in Experiment I, immediate recall of A–C pairs did not produce a facilitative effect in the recall of C terms in the MMFR test. Third, as again found in Experiment I, there was a positive dependence between recall of C and B terms associated with the same A term. The conditional probability of recall of C given recall of B was higher in all four conditions than the conditional probability of recall of C given no recall of B. With the data pooled over all four conditions, the two probabilities were .41 and .24, respectively.

GENERAL DISCUSSION

In an earlier study, Malis (1970) observed that the anomalous A–B priority effect—higher first than second list recall in the MMFR test in the A–B, A–C paradigm—occurred when neither list was tested prior to the MMFR test. He also found that while the effect did not occur in the A–B, C–D paradigm, it occurred in the A–B, A–C paradigm not only when each list was presented for study on a single trial but also when each list was shown on three trials. Moreover, Malis found that the presence or absence of a 30–sec interpolated interval between the presentations of the first and second list did not affect the relation between the first and second list recall.

The two new experiments reported in the present paper have clarified the priority effect in three ways. First, the results of the

new experiments showed that only the testing of the first list is critical for the production of the effect, since the presence or absence of the second-list test prior to the MMFR test made little difference in the MMFR results. Second, the results showed that testing of the first list is critical because it affects the learning of the second list as inferred from the recall test of the second list given immediately after its presentation. In Experiment I, mean percentage of immediate recall of the A–C list was 48% or 15%, depending upon whether the A–B list had been tested or not. Thus the A–B priority effect in the MMFR test is a consequence of exaggerated negative transfer induced by absence of testing of the first list. The third clarification of the A–B priority effect was provided by the data of Experiment II that showed the impairment of learning of the A–C pairs in the second list to be independent of specific A–B pairs tested in the first list. In Experiment II, the mean percentage of immediate recall of the second list was .35, and this percentage was practically identical for the two halves of the list, one whose corresponding A–B pairs had and the other whose corresponding A–B pairs had not been tested prior to the presentation of the A–C list. It appears, therefore, that the effect of testing the first list on the recall of the second list constitutes a general list effect rather than a pair–specific effect. Analyses of pairwise contingencies in the MMFR tests in both experiments revealed that in eight experimental conditions—defined by different tasks given to different groups of subjects in Experiment I, and by different sets of pairs within a list common to all subjects in Experiment II—subjects were more likely to recall C terms of those pairs whose corresponding B terms they also recalled. But this apparent pair-specific facilitation of A–C recall probably represents item or subject selection artifacts (Hintzman, 1972).

The empirical explanation of the A–B priority effect as a consequence of exaggerated negative transfer produced by absence of first-list testing seems to fit well data from those experiments in which the effect has been noted. These include the unpublished experiment by Tulving, Patterson, and Malis, two experiments reported by Malis (1970), several intralist comparisons of A–B and A–C pairs mentioned in the introduction (Bruce & Murdock, 1968; DaPolito, 1966; Goggin, 1966), as well as a recent set of experiments by Bruce and Weaver (1973). In addition, Hintzman, Block, and Summers (1973, Experiment II) have reported that subjects are more accurate in recalling the first presentation modality of a word appearing in a long list twice, each time in a different modality. If we think of the word as the common stimulus term and the two modalities, auditory and visual, as two successive response terms, the Hintzman *et al.* finding is that A–B recall is higher than A–C recall. It must be noted, though, that the result was not found by Hintzman *et al.* in their Experiment I.

While the empirical picture seems to be reasonably clear, the effects of first-list tests on the second-list learning in the A–B, A–C paradigm constitute a theoretical puzzle. Why should it matter for the acquisition and immediate recall of A–C pairs whether or not the earlier presented A–B pairs were tested or not? Many otherwise plausible hypotheses are ruled out by the fact that the effect in question seems to be a list effect rather than a pair-specific effect. As far as we can tell there are no mechanisms postulated in the classical interference theories (Keppel, 1968; Postman, 1971; Postman & Underwood, 1973) that would prepare one for the observation that testing of recall of A–B pairs is an important determinant of ease of learning of A–C pairs. Nor do we find anything in the more recently advanced theories of transfer (Martin, 1971; Greeno *et al.*, 1971) that would provide much guidance for the understanding of the testing effects. One could point out that testing "strengthens" A–B pairs—there is strong evidence for it in the data—and argue that

negative transfer is inversely related to the strength of the A–B association, but such an assertion does not explain anything. It simply restates the findings. Similarly, it is possible to reason that the testing of A–B pairs facilitates recoding of A terms when they are subsequently presented as members of A–C pairs, with the consequence that forming the A–C association is easier, but again the change in the wording of the results does not create much additional insight. Besides, the absence of pair-specific effects does not seem to encourage pursuit of explanations along either of these two lines.

About all we can say at the present time is that the interpolation of an activity requiring explicit retrieval of stored information seems to insulate the A–B list from the A–C list in a way that removes the former as an interfering component in the learning of the latter. We know that insulation of subsets of to-be-learned material, organization of material into higher-order units, plays an important role in storage and retrieval of this material. The present data suggest that insulation can be induced or enhanced through recall tests. Some other evidence in the literature also suggests that testing of one subset of the material affects subjects' handling of another subset. Darley and Murdock (1971), for instance, found a high rate of intrusions from untested lists in learning other lists, in comparison with intrusions from otherwise similar but tested lists.

We know as yet very little about the nature and conditions of insulation effects in general and insulation induced by testing in particular. We do not know to what extent it constitutes one of the control processes of the memory system (Atkinson & Shiffrin, 1968) rather than an obligatory feature of the processing of mnemonic information. We do not know what sorts of other activities would produce similar insulation effects. We do not entirely know why insulation appears to be less important or at least less readily observable in the A–B, C–D paradigm. The study of these and other problems might provide experimental facts that would help us make better theoretical sense out of the observation that testing of one set of materials affects the learning and recall of another.

References

Atkinson, R. C., & Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence and J. T. Spence (Eds.), *The psychology of learning and motivation*, Vol. 2. New York: Academic Press, 1968.

Briggs, G. E. Retroactive inhibition as a function of degree of original and interpolated activity. *Journal of Experimental Psychology*, 1957, **53**, 60–67.

Bruce, D., & Murdock, B. B., Jr. Acoustic similarity effects in short-and long-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 627–631

Bruce, D., & Weaver, G. E. Retroactive facilitation in short-term retention of minimally learned paired associates. *Journal of Experimental Psychology*, 1973, **100**, 9–17.

DaPolito, F. J. Proactive effects with independent retrieval of competing responses. Unpublished doctoral dissertation, Indiana University, 1966.

Darley, C. F., & Murdock, B. B., Jr. Effects of prior free recall testing on final recall and recognition. *Journal of Experimental Psychology*, 1971, **91**, 66–73.

Goggin, J. Retroactive and proactive inhibition in the short-term retention of paired associates. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 526–535.

Greeno, J. G., James, C. T., & DaPolito, F. J. A cognitive interpretation of negative transfer and forgetting of paired associates. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 331–345.

Hintzman, D. L. On testing the independence of associations. *Psychological Review*, 1972, **79** 261–264.

Hintzman, D. L., Block, R. A., & Summers, J. J. Modality tags and memory for repetitions: Locus of the spacing effects. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 229–238.

Keppel, G. Retroactive and proactive inhibition. In T. R. Dixon, & D. L. Horton (Eds.), *Verbal behavior and general behavior theory*. Englewood Cliffs, New Jersey: Prentice-Hall, 1968.

Malis, G. The priority effect in the A–B, A–C paradigm. Unpublished Ph.D. dissertation, University of Toronto, 1970.

MARTIN, E. Verbal learning theory and independent retrieval phenomena. *Psychological Review*, 1971, **78**, 314–332.

MARTIN, E. Stimulus encoding in learning and transfer. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory*. Washington: Winston 1972.

POSTMAN, L. Transfer, interference and forgetting. In In J. W. Kling & L. A. Riggs (Eds.), *Experimental Psychology*. New York: Holt, Rinehart & Winston, 1971.

POSTMAN, L., & UNDERWOOD, B. J. Critical issues in interference theory. *Memory & Cognition*, 1973, **1**, 19–40.

(Received August 8, 1973)