

The influence of retrieval on retention

MARK CARRIER and HAROLD PASHLER
University of California, San Diego, California

Four experiments tested the hypothesis that successful retrieval of an item from memory affects retention only because the retrieval provides an additional presentation of the target item. Two methods of learning paired associates were compared. In the *pure study trial* (pure ST condition) method, both items of a pair were presented simultaneously for study. In the *test trial/study trial* (TTST condition) method, subjects attempted to retrieve the response term during a period in which only the stimulus term was present (and the response term of the pair was presented after a 5-sec delay). Final retention of target items was tested with cued-recall tests. In Experiment 1, there was a reliable advantage in final testing for nonsense-syllable/number pairs in the TTST condition over pairs in the pure ST condition. In Experiment 2, the same result was obtained with Eskimo/English word pairs. This benefit of the TTST condition was not apparently different for final retrieval after 5 min or after 24 h. Experiments 3 and 4 ruled out two artifactual explanations of the TTST advantage observed in the first two experiments. Because performing a memory retrieval (TTST condition) led to better performance than pure study (pure ST condition), the results reject the hypothesis that a successful retrieval is beneficial only to the extent that it provides another study experience.

Inserting a recall test into the learning sequence increases the likelihood that the learner will remember something during a later test. This principle has been demonstrated empirically by many researchers, including Izawa (1966, 1970), Donaldson (1971), Madigan and McCabe (1971), Young (1971), Bartlett and Tulving (1974), Modigliani (1976), and Bartlett (1977). For example, in one such study, Bartlett and Tulving (1974) had subjects learn a list of paired associates. The subjects' retention of the pairs was later tested using a free-recall test or a recognition test. Before the final test, Bartlett and Tulving gave subjects a cued-recall test on some of the paired associates. The experimenters found that retention (as measured by the final test) was better for the items that had been tested in the cued-recall test than items that had not been tested.

Several authors have concluded that the existence of such "testing effects" is evidence that retrieval operations can themselves modify the memory trace of the retrieved item (Bjork, 1975, 1988; Cooper & Monk, 1976; Izawa, 1971, 1985a, 1985b; Wenger, Thompson, & Bartling, 1980). But, despite a large number of studies concerned with these testing effects, the question of whether retrieval per se (hereafter referred to as *retrieval*) contributes anything to the effects of testing remains open. The most significant reason for this ambiguity is that a

successful retrieval of an item from memory also results in a re-presentation of the item. It is possible that any mnemonic effects of a memory test are simply due to this re-presentation of the target material and *not* to the memory retrieval itself (as suggested by Skaggs, 1920). Obviously, this criticism applies to those studies that have shown that receiving a memory test is better for later retention than simply having an equivalent amount of free time (Bartlett, 1977; Bartlett & Tulving, 1974; Darley & Murdock, 1971; Madigan & McCabe, 1971; McDaniel, Kowitz, & Dunay, 1989; Modigliani, 1976; Runquist, 1983, 1986; Young, 1971). Whether there is an active, mnemonic effect of retrieval has important theoretical implications. As we will discuss in the General Discussion, current influential models of memory retrieval make different predictions about such effects.

One obvious strategy for examining effects of retrieval is to compare an intervening test with an intervening experimenter-provided re-presentation of the material. Under these circumstances, intervening tests have been found to be comparable to re-presentations for later performance in some studies (Birnbaum & Eichner, 1971; Donaldson, 1971; Hogan & Kintsch, 1971; Landauer & Eldridge, 1967; Tulving, 1967; Whitten & Bjork, 1977) and worse than re-presentations in other studies (Bregman & Wiener, 1970; Izawa, 1966, 1967, 1970; McDaniel & Masson, 1985). As Wenger et al. (1980) pointed out, neither of these results can rule out a benefit of retrieval, however, because subjects rarely retrieve all of the items during the intervening tests. (We assume that any benefit due to retrieval will not contribute unless the subject succeeds in retrieving the item.) Indeed, the rate of retrieval failure on intervening tests was quite high in some of these studies; Hogan and Kintsch reported 75% retrieval failures.

The authors are grateful to Robert Proctor and Frank Bellezza for several useful comments and to Patrick Duffy, Larry Insel, Mona Lam, Monique Plouf , and Sheree Tsao for assistance in conducting the experiments reported herein. The first author was supported by a San Diego Fellowship. Correspondence should be addressed to H. Pashler at the Department of Psychology, 0109, University of California, San Diego, La Jolla, CA 92093. E-mail: hpashler@ucsd.edu (Internet). hpashler@ucsd (Bitnet).

There are a few studies in which this particular problem is not relevant (Allen, Mahler, & Estes, 1969; Hogan & Kintsch, 1971; Wenger et al., 1980). These researchers *did* find an advantage for administering intervening tests over administering intervening re-presentations of the target material. The incidence of retrieval failures is not a problem here, because if the subjects were successful at all retrieval attempts, then the advantage for the tested condition could only grow larger. Unfortunately, these results are not compelling either, because of various problems associated with the particular designs employed by the experimenters. We will now briefly describe the problems with each of these studies.

Allen et al. (1969) found that 5 presentations of paired associates followed by five cued-recall tests of the pairs led to better final retention than did 10 presentations. Though suggesting that retrieval is better than presentation for later retention, the interpretation of their results is clouded by a potential procedural difficulty. Because of the nature of their design, the number of items on each test in the five cued-recall tests in the test condition was approximately one-third the number of items on each of the presentation trials in the last 5 presentations of the 10-presentation condition. Subjects thus had fewer items at a time to worry about during the tests in the test condition than they did in the last 5 presentations of the 10-presentation condition, and this may have made it easier to learn the items in the former condition.

Hogan and Kintsch (1971, Experiment 2) used a free-recall task, and found that one presentation followed by three free-recall tests was better than four presentations, for performance on a free-recall test delayed 48 h. Similarly, Wenger et al. (1980, Experiments 1-3) found that one presentation followed by three free-recall tests was better than four presentations for performance on a recognition test delayed 48 h. The problem with free-recall studies, however, is that the intervening retrieval test is different from the intervening re-presentation in several critical ways. Consider the procedure used by Wenger et al. in their Experiments 2 and 3. In the retrieval test, the subjects were simply given a period of time (20 sec for five items) to retrieve all the items they could. In the presentation condition, they heard all the items spoken at one item per 4 sec. During the retrieval trials, then, they had the opportunity to rehearse and elaborate together any items that had been successfully retrieved, since free recall is essentially subject-paced. Furthermore, it seems extremely likely that in the retrieval test, the subject would have used items from the list as retrieval cues for other items on the list. If Item A succeeds in retrieving Item B in this fashion, then the subject very plausibly experiences an automatic *joint* re-presentation of A and B together, providing special benefits in this condition that are not relevant to the basic question under discussion.

For both of these reasons, then, interitem associations should be strengthened much more in the retrieval condition than in the simple presentation condition. It is known that interitem associations can contribute power-

fully to free recall (though much less so to recognition) (Mandler, 1979). The point here is not that such beneficial effects of retrieval are spurious; they are likely to be quite real. The critical point is that they are wholly consistent with the possibility that all learning takes place after items are retrieved, with the retrieval itself contributing nothing.

Finally, Wenger et al. (1980, Experiment 4) reported a final recognition advantage for a condition with one presentation followed by three cued-recall tests over a condition with four presentations. However, this final advantage could have been due to a flaw in the design that allowed subjects in the presentation-only condition to ignore later presentations of the repeated target items. In this condition, subjects were passively presented with the target items via tape recorder; in contrast, in the cued-recall condition, subjects were required on every trial to attempt to perform the cued recall of each item. Therefore, performance in the cued-recall condition may have exceeded that in the presentation-only condition merely because in the latter, subjects were always required to perform some operation with the target items.

As this brief review makes clear, there are methodological difficulties that prevent one from answering the basic question of whether memory retrieval strengthens memory traces in other ways than simply through the re-presentation of the target item. The strategy we have employed here attempts to avoid each of these problems. To avoid the possibility of extra interitem elaboration in the retrieval tests, we used paired-associate learning, as in the study by Allen et al. (1969). Unlike those authors, however, we ensured that retention time was not confounded with retrieval test versus re-presentation. To get around the problem that poor success rates on the retrieval can obscure the true benefit that occurs when retrieval succeeds, we set up a procedure whereby subjects, when they failed on the retrieval, were still provided with 2 re-presentation of the item.

EXPERIMENTS 1 AND 2

Experiments 1 and 2 compare the effectiveness of two different methods of learning paired associates. In the first method, the stimulus and response terms of a pair were presented simultaneously, for the subject to study for 10 sec. We call these *pure study trials* (pure ST condition). In the second method, the stimulus member of a pair was presented for 5 sec alone before the response term appeared along with the stimulus member. The stimulus and response pairs remained presented for an additional 5 sec. We call these trials *test trial/study trials* (TTST condition), because the presence of the stimulus member by itself provided the subject with a chance to perform a retrieval test of the response member of the pair. These two methods of presenting paired associates are depicted graphically in Figure 1. Notice that a pure ST trial provides the subject with more time during which the stimulus and response terms are simultaneously present than

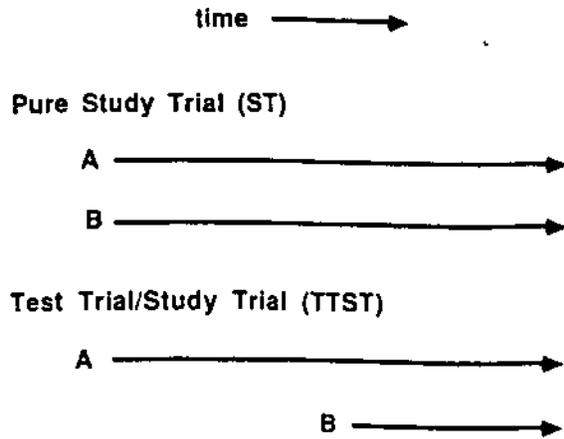


Figure 1. The pure study trial (pure ST) and test trial/study trial (TTST) methods of presenting paired associates.

does a TTST trial. Thus, according to the hypothesis that a successful retrieval of target material serves only as a re-presentation of the target items, learning in the pure ST condition should be *superior* to learning in the TTST condition. If, however, retention for pairs in the TTST condition is better than for pairs in the pure ST condition, then there must be some beneficial effect of the retrieval above and beyond the effect due to the re-presentation of the target item. Furthermore, this beneficial effect of retrieval must be greater than the benefit of an extra 5 sec of guaranteed study time in the pure ST condition.

The procedure in both experiments was to have subjects learn a set of paired associates, present the pairs in additional learning trials (half of the pairs in additional pure ST trials; half in TTST trials), then administer cued-recall tests to measure the retention of the pairs. Each paired associate was thus presented once for study, followed by either three pure ST trials or three TTST trials, and then was cued for recall after a retention period. Retention was measured shortly after learning (5 min) and after a longer delay (24 h).

Experiment 1 used nonsense-word/number pairs as stimuli. Paired-associate learning tasks are often stigmatized as being "ecologically invalid," so in Experiment 2, we kept the procedure the same but taught the subjects something useful—the English cognates of selected words from the St. Lawrence Island/Siberian Eskimo Yupik language. (Although a different language might have been more useful, we felt it important to ensure that subjects would have no previous exposure to our materials.) Because only the compositions of the paired-associate targets in the two experiments were different, the experiments are reported together.

Method

Subjects. In Experiment 1, there were 61 subjects; in Experiment 2, there were 59 subjects. All the subjects were undergraduates at the University of California, San Diego, participating as part of a course requirement. Each subject was tested individually.

Stimuli. In Experiment 1, the stimuli consisted of 40 consonant-vowel-consonant (CVC) trigram/two-digit number pairs. The CVC trigrams were chosen from the middle range of meaningfulness ratings of CVC nonsense syllables compiled by Noble (1961). The syllables had meaningfulness ratings of 2.32 (starting with KIR) to 2.37 (ending with JOS). One CVC syllable (BIC) in that range was not used because it was presumed to have greater current meaningfulness than is indicated in Noble's compilation. The 40 two-digit numbers were randomly selected from a range of 10-99. The syllables and numbers were randomly paired to form the set of target paired associates.

Two written test forms were prepared from the set of stimulus pairs; each test contained 20 cued-recall questions. For each nonsense-syllable item of a stimulus pair, the subjects were asked to write down the two-digit number with which it was paired. Half of the pairs on each test had been presented in the pure ST presentation condition and the other half had been presented in the TTST presentation condition. The particular pairs comprising each half of the test were chosen randomly from all of the pairs in that condition.

In Experiment 2, the stimuli were also 40 paired associates. The stimulus member of each pair was a noun selected from the St. Lawrence Island/Siberian Yupik Eskimo language (Badten, Kaneshiro, & Oovi, 1987). The nouns were all two syllables long (at least for naive speakers). The response member of each pair was a one-word, English semantic equivalent of its paired stimulus. The complete list of stimulus pairs is listed in the Appendix. Two cued-recall tests were created in the same manner as in Experiment 1.

Design. For both experiments, a factorial design was used with two between-subjects factors (word condition and testing order) crossed with two within-subjects factors (presentation condition and testing day). All factors had two levels. *Presentation condition* refers to the method of presentation of a given stimulus pair: half of the pairs were in the pure ST condition and the other half of the pairs were in the TTST condition. *Testing day* refers to the day of the experiment during which a test was given. All the subjects were given two tests on the stimulus pairs: one on Day 1 and the other on Day 2. *Word condition* refers to the selection of stimulus pairs that were in the presentation conditions: for half of the subjects, half of the pairs were in the TTST condition and the other half were in the pure ST condition; for the other half of the subjects, the halves assigned to the conditions were switched. *Testing order* refers to the order of presentation of the two tests over the stimulus pairs: for half of the subjects in each word condition, Test Form 1 was given on Day 1 and Test Form 2 was given on Day 2; for the other subjects, the assignment of pairs to conditions was switched. The subjects were assigned alternately to the four between-subjects conditions as they arrived at the testing facility.

Procedure. The experiments were divided into two sessions. The first session lasted 45 min and the second lasted 15 min. The subjects returned for the Day 2 test from 21 to 27 h later.

In each experiment, the subjects were told that the purpose of the experiment was to study how learning some target information affects reading comprehension. On Day 1, they were informed that they were to try to learn the target pairs, would read a prose passage, and then would take a written test on the target pairs. The subjects were told that the purpose of the written test was to verify that they had learned the target pairs. To lessen the likelihood that they would review the target pairs during the 24 h intervening between the Day 1 and Day 2 tests, they were told that they would take a comprehension quiz on the passage when they returned on Day 2. Instead, the subjects were given the second test on the target pairs on Day 2. Thus, they were not told about the test on the stimulus pairs that they would be given on Day 2, but were informed, instead, of the stimulus-pair test that they would be given on Day 1.

The subjects then watched the presentations of all of the stimulus pairs on a computer screen. Both members of the pair were entered in the display, with the stimulus member displayed just above

the response member. Each target pair was presented three times: the first time in a pure ST trial, the subsequent times in its appropriate presentation condition. Four target pairs were randomly selected at a time. Two of the pairs were from the pure ST condition and the other two were from the TTST condition. Each pair was displayed in a pure ST trial that lasted 20 sec; trials were separated by an intertrial interval of 1 sec (as were all subsequent trials). Next, the four pairs were presented in a random order in their respective presentation conditions. Pairs in the pure ST condition were displayed for 10 sec; for pairs in the TTST condition, the stimulus term appeared for 5 sec, then the response term appeared (together with the stimulus term) for an additional 5 sec. After all subsets of four pairs had been exhausted, all target pairs were again presented in their respective presentation conditions in a new random order. Again, pure ST trials lasted 10 sec and TTST trials lasted 5+5 sec.

The subjects were instructed to say aloud the response term corresponding to each stimulus term as soon as they saw the stimulus term appear on the screen, and to respond as quickly as possible. This pertained to both ST and TTST trials. The only difference, of course, was that in the ST condition, the subjects could just read the response term from the screen, whereas in the TTST condition, they had to retrieve the response term before saying it. After the termination of the pair presentations, the subjects were given 5 min to read an irrelevant prose selection from a college textbook (on Latin American politics). They were then given the first test on the stimulus pairs. The subjects were given no time limit to complete the test and were told to answer all of the questions, even if they had to guess.

When the subjects arrived for Day 2 of the experiment, they were given the test form that they had not been given on Day 1. Again, there was no time limit for this test and the subjects were asked to answer each question, even if it required guessing. After the subjects had finished with the test, they were given a written questionnaire on which they were asked to describe how they memorized the target pairs. They were also asked if they had expected to receive a second test and, if so, whether they had spent time rehearsing the target pairs during the period between the first and second sessions of the experiment. Finally, the subjects were informed of the purpose of the experiment and the reason for the deception concerning the Day 2 test.

Results

An analysis of variance was performed on the results of each experiment, with presentation condition and testing day as within-subjects factors and word condition and testing order as between-subjects factors. Neither word condition nor testing order interacted with either presentation condition or testing order, so no tests involving the former two factors will be reported.

It was decided beforehand to exclude the data from subjects who indicated in their questionnaire that they had spent time studying or testing themselves over the target pairs during the 24-h interval between the first and second sessions. None of the subjects were found to have done so, so none were excluded from the data set for rehearsing the material during this period. However, some subjects were excluded for other reasons, which are noted below.

Experiment 1. The data from 5 subjects were excluded from the analysis. Two of these subjects failed to show up on Day 2 of the experiment; 1 had already participated in a pilot version of the experiment; 1 used his hand to cover the response terms of the target pairs on pure ST

trials, and 1 was excluded due to experimenter error. Thus, data from 56 subjects were entered into the analysis. The mean numbers of pairs correctly completed on Day 1 testing were 5.8 and 6.4 for pairs in the pure ST and TTST presentation conditions, respectively. On Day 2 testing, the mean numbers of pairs completed were 3.7 and 4.1. Of the two variables of interest—presentation condition and testing day—the two main effects were reliable. More pairs from the TTST condition than the pure ST condition were correctly completed [$F(1,52) = 7.48$, $MS_e = 1.68$, $p < .01$], and more pairs were correctly completed on Day 1 than on Day 2 of testing [$F(1,52) = 109.54$, $MS_e = 2.55$, $p < .001$]. The interaction of presentation condition \times testing day was not reliable ($F < 1$).

Because of the moderate size of the cued-recall advantage for the TTST presentation condition over the pure ST condition, we performed a sign test on the number of subjects showing this advantage. On the Day 1 test, 27 subjects showed this advantage, 15 subjects showed an advantage for pure ST over TTST, and 15 subjects showed no advantage in either direction. Using the formula for the normal approximation to the binomial distribution found in Hays (1988), the pattern of results was reliable ($z = 1.70$, $p < .05$). On the Day 2 test, 29 subjects showed this advantage, 17 subjects showed an advantage for pure ST over TTST, and 11 subjects showed no advantage in either direction. This pattern of results was marginally reliable ($z = 1.62$, $.05 < p < .10$).

Experiment 2. The data from 7 subjects were excluded from the analysis. One of these subjects failed to show up on Day 2; 1 had already participated in Experiment 1; 3 subjects' data were lost because of experimenter error; and 2 of the subjects used their hands to cover the response terms of the target pairs on pure ST trials. Thus, data from the remaining 53 subjects were entered into the analysis.

The mean numbers of pairs completed on Day 1 testing were 5.7 and 6.4 for pairs in the pure ST and TTST conditions, respectively. On Day 2 testing, the mean numbers of pairs completed were 3.9 and 4.6. The effect of presentation condition was reliable [$F(1,48) = 9.08$, $MS_e = 2.74$, $p < .01$], indicating that more pairs from the TTST condition were correctly completed than pairs from the pure ST condition. The effect of testing day was also reliable [$F(1,48) = 40.89$, $MS_e = 4.24$, $p < .001$], indicating that more pairs were correctly completed on Day 1 than on Day 2. The interaction of presentation condition \times testing day was not reliable ($F < 1$).

As in the analysis of the results of Experiment 1, we performed a sign test on the number of subjects showing an advantage at cued recall for the TTST condition. On the Day 1 test, 26 subjects showed this advantage, 16 subjects showed an advantage for pure ST over TTST, and 11 subjects showed no advantage in either direction. This pattern of results was marginally reliable ($z = 1.39$, $.05 < p < .10$). On the Day 2 test, 28 subjects showed this advantage, 15 subjects showed an advantage for pure ST over TTST, and 10 subjects showed no advantage in

either direction. This pattern of results was reliable ($z = 1.83, p < .05$).

Questionnaire responses. As is noted above, questions to the subjects concerning their expectancies of the Day 2 test were used to screen them for inclusion of their data in the data set. Another question on the questionnaire concerned the manner in which subjects attempted to memorize the pairs. Most of them indicated that they used some kind of strategy for memorizing at least some of the stimuli during the presentations. In Experiment 1, 49 of the subjects (86%) reported using some kind of mnemonic strategy for memorization. In Experiment 2, 45 of the subjects (85%) reported doing so. Very few of the subjects (6 subjects in Experiment 1; 4 in Experiment 2) reported trying to remember the items by simply rehearsing them over and over.

Discussion

In both experiments, the subjects remembered more pairs from the TTST condition than from the pure ST condition. Yet, in the pure ST condition, there was more time during which *both* the stimulus member and the response member of a pair were presented to the subject. If a successful retrieval attempt is beneficial only to the extent that it provides a re-presentation of the target item, then performance in the pure ST condition should have been superior to performance in the TTST condition. Therefore, having to *retrieve* the response member of the pair in the TTST condition was more effective than simply *studying* the response member in the pure ST condition. This suggests that the retrieval processes involved in a conscious retrieval play a role in the effect of a retrieval on retention.

However, the possibility that the subjects were able to adjust their encoding based on information about the recallability of a target item cannot be completely ruled out in Experiments 1 and 2. Some researchers have suggested that a retrieval attempt provides the learner with knowledge of the recallability or degree of recallability of target items (Half, 1977; Skaggs, 1920; Thompson, Wenger, & Bartling, 1978). This knowledge can then be used to guide future encoding of the target items. In this hypothesis, advanced by Thompson et al. (1978), knowledge of recallability allows a subject to differentially allocate study time to target items based on their retrieval difficulty in a situation where the subject has multiple items available for study during a single, fixed-duration study period. If an item is difficult to recall, for instance, then the subject knows that he/she must allocate more future study time for that item than for recalling a relatively easy item. Study time is controlled in our experiments, so this cannot apply here. It might be possible, though, that subjects varied their encoding *effort* on the basis of knowledge of recallability of the items. For example, a subject may find that, when first attempting retrieval in a TTST trial, a particular item is relatively difficult to remember. This information then allows the subject to put a greater amount of encoding effort into memorization

at the next occurrence of the target item during the presentations. Reliable information about the recallability of items will not be available for items presented in the pure ST condition, since subjects are not engaged in cued retrieval of items in that condition. Therefore, in addition to the beneficial effects of retrieval processes, a strategy of varying encoding effort could contribute to the advantage found for TTST items on the final test.

Such a strategy is available to subjects only when additional presentations occur after the initial TTST presentation. Therefore, we conducted a third experiment in which there were no presentation trials following the initial TTST and pure ST trials.

EXPERIMENT 3

In Experiment 3, we attempted to assess the contributions of retrieval processes to the TTST advantage over pure ST in Experiments 1 and 2, while preventing the aforementioned encoding strategy which, as noted above, could have contributed to the TTST advantage as well. To ensure that such a strategy could not be used, we used the same general design as in the previous two experiments, but made the TTST versus pure ST manipulation on the last presentation trial only. There were three presentations of each target during the presentation period. The first two presentations were always in the pure ST condition for all items. The third presentation was either in the TTST condition or in the pure ST condition.

The extent to which the TTST advantage found in the first two experiments is due to retrieval processes versus subjects' encoding strategies can be assessed by comparing the results of Experiment 3 with those of Experiments 1 and 2. In particular, an account of the results of Experiments 1 and 2 in terms of conscious encoding strategies predicts that, in Experiment 3, there should be little or no effect of the TTST versus pure ST manipulation on final performance. Therefore, if there is still a significant advantage for TTST items over pure ST items, then we can rule out conscious encoding strategies as an explanation of the effect.

Method

Subjects. The subjects were 60 undergraduates at the University of California, San Diego, participating as part of a course requirement. Each subject was tested individually.

Stimuli. The stimuli were 30 Eskimo/English word pairs, drawn randomly from the set of 40 such pairs used in Experiment 2. A single test form was constructed for the final cued-recall test with the 30 Eskimo words listed in random order.

Design. A factorial design was used with two between-subjects factors (word condition and presentation order) and a single within-subjects factor (presentation condition). Word condition and presentation condition were the same as those for Experiments 1 and 2. *Presentation order* refers to the order of presentation of words assigned to the TTST condition and to the pure ST condition. For each of the three cycles through the list of targets during the presentation stage of the experiment, the TTST and pure ST targets were presented in the same randomly determined order for one half of the subjects. However, for the other half of the subjects, the assignment of conditions to serial positions in the presentation was

simply switched from TTST to pure ST and vice versa. This ensured that, across all subjects, neither presentation condition enjoyed a special advantage of having occurred in any particular serial positions in the presentation sequence. The subjects were assigned randomly to the four between-subjects conditions.

Procedure. The entire experiment lasted approximately 1 h. The subjects were told that the purpose of the experiment was to see how well they could learn foreign-language vocabulary items. They were informed that, after the presentations of the target items, they would be performing a short task (the distractor task) and then be taking a written test on the vocabulary items.

The items were displayed on the computer screen in the same manner as in the previous experiments. The first two presentations of each pair were made in the pure ST condition. The initial presentation of each item lasted 15 sec; the second presentation lasted 10 sec. The third presentation was made in either the TTST condition (one half of the items) or the pure ST condition (the other half of the items). During the third set of presentations, the pure ST trials lasted 10 sec and the TTST trials lasted 5 + 5 sec. Trials were separated by an intertrial interval of 1 sec.

As in Experiments 1 and 2, the subjects were instructed to say aloud the response term corresponding to each stimulus term as soon as they saw the stimulus term appear on the screen, responding as quickly as possible. This pertained to both ST and TTST trials. After the termination of the pair presentations, they were given 5 min to solve as many of 70 written arithmetic problems as they could. The subjects were then given the test on the stimulus pairs. They were given no time limit to complete the test and were told to answer all of the questions, even if they had to guess.

Results and Discussion

An analysis of variance was performed on the results with presentation condition as a within-subjects factor and word condition and presentation order as between-subjects factors. Neither word condition nor presentation order interacted with presentation condition, so no tests involving the former two factors will be reported.

The mean numbers of pairs correctly completed were 9.0 and 10.1 for pairs in the pure ST presentation condition and pairs in the TTST presentation condition, respectively. More pairs from the TTST condition than the pure ST condition were correctly completed [$F(1,59) = 11.52$, $MS_e = 2.96$, $p < .005$]. As in the previous experiments, we performed a sign test on the number of subjects showing an advantage for the TTST presentation condition. Thirty-two subjects showed this advantage, 13 subjects showed an advantage for pure ST over TTST, and 15 subjects showed no advantage in either direction. This pattern of results was reliable ($z = 1.83$, $p < .05$).

The results of Experiment 3 were extremely similar to the results of Experiments 1 and 2, suggesting that the effect of retrieval on later retention does not depend on subjects' strategies at differentially allocating encoding effort (or later study time) on the basis of the recallability of items. This strategy hypothesis, as an account of the results of Experiments 1 and 2, would predict no difference between items in the TTST condition and items in the pure ST condition in Experiment 3, since there were no subsequent study trials of items after the TTST versus pure ST manipulation.

EXPERIMENT 4

Experiment 4 was performed to rule out a relatively uninteresting explanation of the TTST advantage on cued recall that was revealed in the previous experiments. According to this explanation, subjects rehearse TTST items during list presentation at the expense of rehearsing pure ST items. If, during list presentation, the subject has free time to work on certain items (e.g., between item presentations), then he/she could conceivably opt to rehearse TTST items and not pure ST items. One reason that the subject might desire to rehearse TTST items instead of pure ST items is that TTST items may have been considered by the subject to be less well learned, due to less total time for TTST items during which both the stimulus and the response term of the target pair were presented. In this view, the TTST items are retained longer than the pure ST items, simply because they were rehearsed more than the pure ST items.

This differential rehearsal hypothesis was originally advanced by Slamecka and Katsaiti (1987) as an explanation of the *generation effect*. In the generation effect, stimuli that are produced as the result of some effortful problem-solving process are better remembered than stimuli that are given directly to the subjects (Jacoby, 1978; Slamecka & Graf, 1978). Slamecka and Katsaiti observed an advantage of "generated" items over items that were merely presented to the subjects when both kinds of items occurred on the same target list of items. However, when some subjects generated items and other subjects were presented with the items, the generation effect disappeared. Slamecka and Katsaiti suggested that subjects in the former mixed-list design rehearsed generated items at the expense of presented items. In contrast, in the latter between-subjects design, rehearsal of the generated items could not arise at the cost of not rehearsing presented items. Testing this hypothesis further in a mixed-list design, they found that requiring subjects to rehearse only the current item eliminated the previously observed generation effect.

Since the differential rehearsal explanation of the TTST advantage found in Experiments 1-3 could potentially account for the effect only in mixed-list designs, we pitted TTST items against pure ST items in a pure-list design. The subjects studied a single list of target items for two cycles through the list. On the third study cycle, however, they studied and were tested on short five-item sublists. All of the items on each smaller list were presented in this third study cycle using either the pure ST method of presentation or the TTST method. After each such study period, the subjects were given a cued-recall test on just the five previously presented items. These study-test phases alternated until all of the target items had been tested. According to the differential rehearsal hypothesis, the TTST advantage over pure ST items should be eliminated because pure ST items will not suffer from dif-

ferential rehearsal of TTST items. On the other hand, any explanation of the present TTST advantage in terms of the effects of the retrieval process itself occurring on TTST trials predicts that the TTST advantage will remain in this new design.

Method

Subjects. The subjects were 60 undergraduate students at the University of California, San Diego, participating for course credit. Each subject was tested individually.

Materials and Apparatus. The target pairs were the same 30 items used in Experiment 3; this set was divided arbitrarily into six five-item lists.

Procedure. The procedure was very similar to that in Experiment 3. As in the previous experiments, all presentations were made on personal computers. The subjects first learned the entire list of items. The set of pairs was presented twice in succession using the pure ST method. Each pair was presented the first time for 15 sec and the second time for 10 sec. The blank interval between items lasted 1 sec in all cases. For the first cycle through the list of target items, the items were always presented in the same random order for all subjects. For the second cycle, the order of items was randomized for each subject.

After the first two presentation cycles, the subject started on the short five-item study-test phases. For each five-item sublist of the entire list of target items, the subject studied all of the items in either pure ST or TTST presentation methods. The order of items within the presentation was randomized for each subject. The presentation method for each sublist alternated until all of the items had been tested (six tests). For half of the subjects, the first short list was presented using the pure ST method and for the other half of the subjects, the first list was presented using the TTST method. After the presentation of the items in the sublist was complete, the subjects engaged in a distractor task for 2 min. In the distractor task, they were given pairs of words and, for each pair, were to recite out loud a sentence that contained the two words. They were told to complete as many of these as possible. There was always a sufficient number of pairs of words so that no subject ever finished constructing sentences before the 2 min ended.

After the distractor task, the subjects were given as long as they needed to complete the cued-recall test for the just-presented five items. In the cued-recall test, they were shown all five of the stimulus members of the target pairs and asked to complete them. As in the previous experiments, they were told to guess if they did not remember an answer. The subjects typed their responses on the computer keyboard and signaled (by appropriate keypresses) when they were finished. At this point, the next study-test phase started.

Results and Discussion

An analysis of variance with presentation method as a within-subjects factor revealed that presenting items using the TTST method resulted in a higher proportion correct on cued recall (0.71) than did the pure ST method (0.66) [$F(1,59) = 7.76, p < .01$]. A sign test showed that the number of subjects showing a final advantage for the TTST method of presentation was reliable ($z = 1.75, p < .05$). Thirty subjects showed this advantage, 17 subjects showed the opposite effect, and 13 subjects showed no advantage either way.

The results of this experiment rule out differential rehearsal as an account of the TTST advantage found in the previous experiments. If the subjects had been rehearsing TTST items at the expense of pure ST items in the

previous experiments, and if this was the source of the TTST advantage obtained, then the pure-list design employed in this experiment should have resulted in no such TTST advantage.

GENERAL DISCUSSION

The results of the four experiments reported here support the hypothesis that retrieving an item from memory has beneficial effects for later retention above and beyond the effects due to merely studying the item. Experiments 1 and 2 showed that retrieving the response member of a paired associate led to better long-term cued-recall performance than did simply being presented with the item for study for an equal amount of time. Experiments 3 and 4 ruled out certain artifactual explanations of this effect by showing that the effect still obtained even when there were no further exposures to the target items after the first retrieval attempt (Experiment 3) and even when the experiment was conducted using a pure-list design (Experiment 4).

Relation to the Generation Effect

The final advantage for previously retrieved items found in the present set of experiments bears a striking resemblance to the generation effect, though, as we will discuss below, the two effects may not be related. As described above, the generation effect obtains when target items that are generated by the subject in response to some set of experimentally provided cues are better retained than items merely presented to the subjects (Jacoby, 1978; Slamecka & Graf, 1978). For example, in a generation task used by Slamecka and Graf (Experiment 1), the subjects were presented with a generation rule (i.e., synonym), a stimulus word, and the first letter of the correct response. In the control (nongenerate) task, they read both the stimulus word and the correct target word aloud. The similarity between the generation task and the TTST method of presentation employed here is, of course, that both tasks require producing target items from memory and that both tasks lead to better final retention.

The TTST advantage found in the present experiments and the generation effect could have different causes. It is possible that, while the present TTST advantage is due to the effect of retrieval processes, the generation effect is due to the extra (nonretrieval) processing usually required in generation tasks. In support of this notion, some researchers studying the generation effect have shown that subjects do not have to retrieve anything from memory in the generation task in order to produce generation effect-like results. Kinoshita (1989, Experiment 4) found that subjects who correctly copied a word originally presented with two underlined, transposed letters performed better on later recognition of the items than did subjects who simply copied correctly spelled words (though this effect did not obtain with free recall as the final test). In addition, Schmidt (1990) found that words for which subjects copied a missing letter (printed to the right of the

word) onto the correct blank space embedded in the word were later more likely to be recalled than words that were simply studied in whole form. Another, more likely, possibility is that the generation effect is often due to the combined beneficial effects of retrieval plus the extra processing of the target material demanded by the generation task.

Causes of the Retrieval Effect

The present results indicate a moderate benefit of retrieval, yet a question remains as to the nature of the benefit. We performed several post hoc analyses of the data from Experiments 1-3 in an effort to identify in more detail the basis for the advantage in the TTST condition. Errors were divided into two types: intralist errors, in which subjects produce a response term that was paired with a different stimulus term, and extralist errors, in which the response term was not part of the experiment. The results showed no consistent differences between the two conditions in the incidence of these types of errors. Furthermore, neither overall memory performance (the total number of paired associates solved correctly) nor distractor task performance (the number of math problems solved) were good predictors of the final patterns of effects for individual subjects in Experiment 3.

Though these results do not identify the basis for the TTST advantage in the present data, in the past it has been suggested that retrieval attempts may provide general practice or context at retrieval and thus boost the likelihood of correct retrieval at a future date (Landauer & Bjork, 1978; Runquist, 1983). This notion does not seem to have much merit, since the common finding is that beneficial effects of prior recall are not found for items that were not tested. Runquist (1983, Experiment 1), for example, found that initial cued-recall testing of half of the word pairs on a list did not affect the final cued-recall probability of the word pairs from the untested half of the list. Furthermore, in the present experiments, the reliable difference between the TTST condition and the pure ST condition in both experiments is evidence that general practice at retrieval does not account for testing effects. If the latter was true, then any benefit of retrieval practice in the TTST condition should also occur for the items in the pure ST condition, given the general nature of the hypothesized retrieval practice. Therefore, the general retrieval practice notion predicts no difference between the two experimental conditions at final recall.

Rather than a global retrieval effect, then, the beneficial effect of retrieval occurs at the level of individual items. There are various hypotheses about why this should be so. These include the hypotheses that the act of retrieval requires neural activity that consolidates the representation of the target item in memory (Cooper & Monk, 1976; Whitten & Bjork, 1977), that cued recall of a paired-associate item strengthens the structural, integrative information about the item (Mandler, 1979), and that the act of retrieval may either strengthen existing "retrieval routes" to the representation of the item in memory (Birnbaum & Eichner, 1971; Bjork, 1975) or require the cre-

ation of new routes (Bjork, 1975). In the latter case, it is surmised that the creation of new routes will increase the total number of retrieval routes to the representation of an item in memory and thus raise the probability of a correct recall on a later test.

However, some of these hypotheses do not predict that retrieval will be *better than study* for future retention, as has been shown in the present experiments. Presumably, consolidation (Cooper & Monk, 1976; Whitten & Bjork, 1977) and integration (Mandler, 1979) of the memory trace occur during retrieval *and* during study. It is not immediately obvious why either consolidation or integration would be more likely to occur or be more effective during retrieval than during study. On the other hand, the notions that retrieval strengthens existing retrieval routes (Birnbaum & Eichner, 1971; Bjork, 1975) or requires the creation of new routes (Bjork, 1975) can be directly extended to account for the superiority of retrieval over study in the current experiment. In the former case, retrieval routes that will be useful at later retrieval may be more likely to be strengthened during retrieval than during study. In the latter case, the creation of new retrieval routes may only be necessary during retrieval and not during study. However, the concept of a retrieval "route" is inherently vague.

Implications for Network Models

More specific models of memory may be more illuminating. Consider the class of models of memory storage in which connections are modified between units that represent different elements of patterns to be inter-associated. One such class of models uses Hebbian learning procedures (Hebb, 1949; Hopfield & Tank, 1986). In these models, connection strengths are increased between units that are simultaneously active. This learning procedure succeeds in producing networks in which presentation of the pattern representing one of the inter-associated elements will (to some degree) reinstate the pattern representing the other element. In this learning scheme, optimal associative learning can only occur when the most complete and faithful representations of both elements to be associated are present. For this reason, it is not clear why memory storage would be more effective when the subject was deprived of one of the patterns to be associated, as in the TTST condition in the present experiments. Other distributed models of associative memory, such as those of J. A. Anderson (Anderson & Hinton, 1981) and Metcalfe and Eich (Eich, 1985), would seem to have the same problems in accounting for these results.

By contrast, error-correction learning models (e.g., McClelland & Rumelhart, 1986) could provide a natural account for superior performance in the TTST condition. In this framework, learning of an association between Elements A and B can be conceptualized as the modification of connections so as to reduce the error in the version of Element B that the network itself generates in response to Element A (this version will be designated B'). According to this class of models, an attempted retrieval produces

B^* , and the difference between B^* and the canonical representation of B (available in some form of long-term memory) is computed so that connections can be modified optimally. Explicit feedback might not be necessary if the system corrects errors based on the canonical representation of whatever item best matches its output pattern. In this scheme, the system basically assumes that it is correct. Note that this sort of account is consistent with the absence of a retrieval benefit on trials where retrieval fails.

To explain why presenting B at the same time as A produces inferior learning, one need only suppose that presenting B "contaminates" the estimate B^* . Giving the subject B may directly activate the canonical representation of B, and therefore, the resulting B^* may not be the most useful for adjusting connection weights. Informally, one might say that getting B too soon prevents the network from knowing what it would have produced on its own, and thereby inhibits it from properly correcting for any error.

This view of the inferior performance in the pure ST condition is analogous to certain interpretations of associative blocking in Pavlovian conditioning. Blocking occurs when a conditioned stimulus (CS1) is paired with an unconditioned stimulus (US) while, at the same time, another CS2 is presented, which was *previously* associated with the US. What is observed is a reduction in conditioning to CS1 (compared with the conditioning that occurs when CS1 is presented alone with the US). Error-correction models like the Rescorla/Wagner model suggest that conditioning to CS1 occurs only to the degree that the US is not already expected on the basis of the stimulus context, which includes CS2 (see Sutton & Barto, 1981). That is, conditioning is proportional to the difference between the estimate of the US computed by the system (U^*) and the actual US that follows. Thus, presenting CS2 reduces the difference between U^* and US, and thereby reduces (blocks) conditioning to CS1. The error-correction model of the retrieval effect reported here would work analogously: presenting B itself may reduce the difference between B^* and B (thereby reducing learning in the pure ST condition).

While error-correction models provide an interesting account of why retrieval can be more beneficial than actual presentation of the target material, they suggest that it is not just retrieval that contributes to optimal learning. After all, in an error-correction model, learning (i.e., weight changes) takes place *after* retrieval has occurred and an estimate (B^*) of the correct response has been generated. Retrieval merely provides the system with the response estimate. Learning occurs later, while the response estimate is compared with the canonical representation of the response term. This is rather different than the most literal interpretation of the "retrieval route" idea, which might suggest that once a path has been hacked through the jungle, there is no further work to be done. The present results and analyses do not provide sufficient information to distinguish between the hypothesis that retrieval processes contribute directly to learning and the hypothesis that retrieval contributes to learning in the manner suggested above in the error-correction framework.

In summary, the superiority of TTST over pure ST presentations indicated by the present results is difficult to account for on certain well-known network models of learning (e.g., Hebbian or convolution/correlation models), but easy to account for on error-correction learning models. Given the notorious difficulty in deriving empirical predictions from these types of models, the present results suggest that further research on the role of retrieval in memory storage may provide a useful strategy for evaluating these models.

Finally, the present results suggest that the optimal method of paired-associate learning should include a large number of test trials. The practical implication of the present study for rote learning is clear: After some degree of learning has been achieved, the learner should be subjected to forced retrievals *without* having the target material present. The existence of flashcards as aids to learning suggests that implicit belief in the efficacy of test trials may be widespread, but the phenomenon is hardly mentioned in the educational literature. Careful analysis of where test trials are helpful, and comparisons of test trials with other mnemonic techniques, should have substantial potential for optimizing learning.

REFERENCES

- ALLEN, G. A., MAHLER, W. A., & ESTES, W. K. (1969). Effects of recall tests on long-term retention of paired associates. *Journal of Verbal Learning & Verbal Behavior*, **8**, 463-470.
- ANDERSON, J. A., & HINTON, G. E. (1981). Models of information processing in the brain. In G. E. Hinton & J. A. Anderson (Eds.), *Parallel models of associative memory* (pp. 9-48). Hillsdale, NJ: Erlbaum.
- BADTEN, L. W., KANESHIRO, V. O., & OOOI, M. (1987). *A dictionary of the St. Lawrence Island/Siberian Yupik Eskimo language*. Fairbanks: Alaska Native Language Center, University of Alaska.
- BARTLETT, J. C. (1977). Effects of immediate testing on delayed retrieval: Search and recovery operations with four types of cue. *Journal of Experimental Psychology: Human Learning & Memory*, **3**, 719-732.
- BARTLETT, J. C., & TULVING, E. (1974). Effects of temporal and semantic encoding in immediate recall upon subsequent retrieval. *Journal of Verbal Learning & Verbal Behavior*, **13**, 297-309.
- BIRNBAUM, I. M., & EICHNER, J. T. (1971). Study versus test trials and long-term retention in free-recall learning. *Journal of Verbal Learning & Verbal Behavior*, **10**, 516-521.
- BJORK, R. A. (1975). Retrieval as a memory modifier: An interpretation of negative recency and related phenomena. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 123-144). Hillsdale, NJ: Erlbaum.
- BJORK, R. A. (1988). Retrieval practice and the maintenance of knowledge. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues* (Vol. 1, pp. 396-401). New York: Wiley.
- BREGMAN, A. S., & WIENER, J. R. (1970). Effects of test trials in paired-associate and free-recall learning. *Journal of Verbal Learning & Verbal Behavior*, **9**, 689-698.
- COOPER, A. J. R., & MONK, A. (1976). Learning for recall and learning for recognition. In J. Brown (Ed.), *Recall and recognition* (pp. 131-156). New York: Wiley.
- DARLEY, C. F., & MURDOCK, B. B., JR. (1971). Effects of prior free recall testing on final recall and recognition. *Journal of Experimental Psychology*, **91**, 66-73.
- DONALDSON, W. (1971). Output effects in multitrial free recall. *Journal of Verbal Learning & Verbal Behavior*, **10**, 577-585.
- EICH, J. M. (1985). Levels of processing, encoding specificity, elaboration, and CHARM. *Psychological Review*, **92**, 1-38.

- HALFF, H. M. (1977). The role of opportunities for recall in learning to retrieve. *American Journal of Psychology*, **90**, 383-406.
- HAYS, W. L. (1988). *Statistics* (4th ed.). San Francisco: Holt, Rinehart & Winston, Inc.
- HEBB, D. O. (1949). *The organization of behavior*. New York: Wiley.
- HINTZMAN, D. L. (1974). Theoretical implications of the spacing effect. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola Symposium* (pp. 77-99). Hillsdale, NJ: Erlbaum.
- HOGAN, R. M., & KINTSCH, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning & Verbal Behavior*, **10**, 562-567.
- HOPFIELD, J. J., & TANK, D. W. (1986). Computing with neural circuits: A model. *Science*, **233**, 625-633.
- IZAWA, C. (1966). Reinforcement-test sequences in paired-associate learning. *Psychological Reports*, **18**, 879-919.
- IZAWA, C. (1967). Function of test trials in paired-associate learning. *Journal of Experimental Psychology*, **75**, 194-209.
- IZAWA, C. (1970). Optimal potentiating effects and forgetting-prevention effects of tests in paired-associate learning. *Journal of Experimental Psychology*, **83**, 340-344.
- IZAWA, C. (1971). The test trial potentiating model. *Journal of Mathematical Psychology*, **8**, 200-224.
- IZAWA, C. (1985a). The identity model and factors controlling the superiority of the study-test method over the anticipation method. *Journal of General Psychology*, **112**, 65-78.
- IZAWA, C. (1985b). A test of the differences between anticipation and study-test methods of paired-associate learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**, 165-184.
- JACOBY, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning & Verbal Behavior*, **17**, 649-667.
- KINOSHITA, S. (1989). Generation enhances semantic processing? The role of distinctness in the generation effect. *Memory & Cognition*, **17**, 563-571.
- LANDAUER, T. K., & BJORK, R. A. (1978). Optimum rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical Aspects of Memory* (pp. 625-632). London: Academic Press.
- LANDAUER, T. K., & ELDRIDGE, L. (1967). Effects of tests without feedback and presentation-test interval in paired-associate learning. *Journal of Experimental Psychology*, **75**, 290-298.
- MADIGAN, S. A., & McCABE, L. (1971). Perfect recall and total forgetting: A problem for models of short-term memory. *Journal of Verbal Learning & Verbal Behavior*, **10**, 101-106.
- MANDLER, G. (1979). Organization and repetition: Organizational principles with special reference to rote learning. In L.-G. Nilsson (Ed.), *Perspectives on memory research: Essays in honor of Uppsala University's 500th anniversary* (pp. 293-327). Hillsdale, NJ: Erlbaum.
- MCCLELLAND, J. L., & RUMELHART, D. E. (1986). A distributed model of human learning and memory. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 2. Psychological and biological models* (pp. 170-215). Cambridge, MA: MIT Press.
- MCDANIEL, M. A., KOWITZ, M. D., & DUNAY, P. K. (1989). Altering memory through recall: The effects of cue-guided retrieval processing. *Memory & Cognition*, **17**, 423-434.
- MCDANIEL, M. A., & MASSON, M. E. J. (1985). Altering memory representations through retrieval. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**, 371-385.
- MODIGLIANI, V. (1976). Effects on a later recall by delaying initial recall. *Journal of Experimental Psychology: Human Learning & Memory*, **5**, 609-622.
- NOBLE, C. E. (1961). Measurements of association value (a), rated associations (a'), and scaled meaningfulness (m') for the 2100 CVC combinations of the English alphabet. *Psychological Reports*, **8**, 487-521.
- RUNQUIST, W. N. (1983). Some effects of remembering on forgetting. *Memory & Cognition*, **11**, 641-650.
- RUNQUIST, W. N. (1986). The effect of testing on the forgetting of related and unrelated associates. *Canadian Journal of Psychology*, **40**, 65-76.
- SCHMIDT, S. R. (1990). A test of resource-allocation explanations of the generation effect. *Bulletin of the Psychonomic Society*, **28**, 93-96.
- SKAGGS, E. B. (1920). The relative value of grouped and interspersed recitations. *Journal of Experimental Psychology*, **3**, 424-446.
- SLAMECKA, N. J., & GRAF, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning & Memory*, **4**, 592-604.
- SLAMECKA, N. J., & KATSANT, L. T. (1987). The generation effect as an artifact of selective displaced rehearsal. *Journal of Memory & Language*, **26**, 589-607.
- SUTTON, R. S., & BARTO, A. G. (1981). Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review*, **88**, 135-170.
- THOMPSON, C. P., WENGER, S. K., & BARTLING, C. A. (1978). How recall facilitates subsequent recall: A reappraisal. *Journal of Experimental Psychology: Human Learning & Memory*, **4**, 210-221.
- TULVING, E. (1967). The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning & Verbal Behavior*, **6**, 175-184.
- WENGER, S. K., THOMPSON, C. P., & BARTLING, C. A. (1980). Recall facilitates subsequent recognition. *Journal of Experimental Psychology: Human Learning & Memory*, **6**, 590-598.
- WHITTEN, W. B., II, & BJORK, R. A. (1977). Learning from tests: Effects of spacing. *Journal of Verbal Learning & Verbal Behavior*, **16**, 465-478.
- YOUNG, J. L. (1971). Reinforcement-test intervals in paired-associate learning. *Journal of Mathematical Psychology*, **8**, 58-81.

APPENDIX

List of Paired-Associate Targets Used in Experiment 2

Eskimo words are taken from Badten, Kaneshiro, and Oovi (1987). The first column contains the stimulus member of a pair (Eskimo word); the second column contains the response member (English word).

Eskimo	English	Eskimo	English
AGLUK	JAW	NUNA	VILLAGE
AKI	MONEY	NURGU	NOOSE
ANGYAK	BOAT	PENGUQ	HILL
ARI	OTTER	QALTAQ	PAIL
ASAK	AUNT	QALU	LADLE
EGSI	DANDRUFF	QANTAK	UTERUS
ESLA	WEATHER	QUKAQ	WAIST
ESTUK	TOENAIL	SANQUN	TOOL
IGHU	LEG	SIGUN	EAR
IGLAK	VOICE	SUMEQ	IDEA
IMAO	SEA	TAMLU	CHIN
KENUK	RIDGE	TEGHIK	ANIMAL
KINGU	BACK	TEQUQ	URINE
KINGUK	MAGGOT	TEPA	ODOR
KUMLU	THUMB	TUGUN	IVORY
MALLAQ	DUST	TUMA	PATH
MEGHUN	CAN	TUYA	SHOULDER
MESUQ	JUICE	UNEQ	UNDERARM
NASAO	HOOD	UNGLUN	NEST
NULLU	RUMP	YAQUQ	WING

(Manuscript received October 10, 1989; revision accepted for publication February 4, 1992.)