JOURNAL OF MEMORY AND LANGUAGE **35,** 212–230 (1996)
ARTICLE NO. 0012

The Persistence of False Memories in List Recall

KATHLEEN B. MCDERMOTT

Rice University

Roediger and McDermott (1995) recently re-introduced a paradigm to study the creation of false memories. Subjects hear short lists of related words (e.g., thread, pin, eye, sewing, etc.), all of which are associates of a critical nonpresented word (e.g., needle); on a free recall test given immediately after list presentation, subjects often erroneously recall the critical nonpresented word. The experiments reported here explore (a) the effect of test delay on false recall and (b) whether multiple study/test opportunities reduce the proportion of critical items erroneously recalled. In Experiment 1, introduction of a 2-day delay between study and test produced probabilities of false recall that exceeded those of veridical recall. In addition, prior testing of the list enhanced false recall, much like testing enhances later recall of studied items (the testing effect). In Experiment 2, an attempt was made to reduce or eliminate the false recall effect by using a multitrial study/test procedure. Although subjects were able to reduce the proportion of critical nonpresented words erroneously recalled, they were unable to eliminate the false recall effect, even after 5 study-test trials. An interaction occurred between accurate and false recall as a function of retention interval: after a one-day delay, false recall levels rose, whereas accurate recall decreased. Results of both experiments demonstrate the persistence of this memory illusion.

Investigation into the constructive (or inferential) nature of recall was pioneered by Bartlett (1932), and most of the subsequent research on this topic has been guided by his procedural choices and theoretical assertions. Bartlett chose for his study material a complex Indian folktale containing culturally bound elements of the supernatural that were unfamiliar to his English subjects. Bartlett then administered a free recall test to his subjects and found that they tended to produce distortions in line with their cultural expectations (but see Gauld, & Stephenson 1967, for questions regarding the nature of Bartlett's test instructions). Bartlett concluded that remembering involves reconstruction of the past in light of our understanding of the world (i.e., our schemas).

Most memory researchers preceding Bartlett had chosen word lists or nonsense syllables as their study material. Bartlett argued that this approach was too restrictive, claiming that it promotes rote memorization, or reproductive memory. He argued that the interesting aspect of memory was its reconstructive nature, which is demonstrated by the influence of schemas on remembering. He speculated that word lists and nonsense syllables promote reproductive (i.e., rote) memory and that more elaborate materials (e.g., stories) are necessary to obtain reconstructive memory. Therefore, Bartlett urged researchers to choose such materials in their study of memory.

Since the time of Bartlett's (1932) book,

This research was supported by a research grant from Sigma Xi to Kathleen McDermott and by Grant F49620-92-J-0437 from the Air Force Office of Scientific Research awarded to Henry L. Roediger III. These experiments form part of a doctoral dissertation for Rice University. I thank my dissertation committee (Henry L. Roediger III, Chair; Jeffrey Bowers; Stephen L. Klineberg; Randi L. Martin; James R. Pomerantz) for their advice on this project. I also thank Ira Hyman, Doug Nelson, David Payne, and Kerry Robinson for providing comments on an earlier version of the paper and Glen Doniger, Ron Haas, Bettina Johnson, Lubna Manal, Jenny Mireles, Keith Rozendal, and Brad Thompson for their help in testing subjects and scoring data. Address reprint requests to Kathleen B. McDermott, Department of Psychology, MS 25, Rice University, 6100 Main Street, Houston, TX 77005-1892. E-mail: kmcd@rice.edu.

0749-596X/96 $18.00

there seems to have been a general acceptance of the idea that traditional list learning procedures do not allow for constructive memory processes; most researchers studying constructive memory have chosen sentences (Bransford & Franks, 1971), stories (Sulin & Dooling, 1974), slide sequences (Loftus, Miller, & Burns, 1978), or videotapes (Loftus & Palmer, 1974) as their materials (but see Underwood, 1965, for an exception). However, there has been a general departure from a different aspect of Bartlett's procedure in that recognition, not free recall, has been the test of choice. This trend is probably attributable to a general difficulty in obtaining intrusions on free recall tests; this difficulty is magnified on tests given shortly after the study episode (Cofer, 1967; Roediger & Payne, 1985; Spiro, 1980).

There is one generally overlooked exception to both of these procedural trends, in which striking memory illusions were produced in a simple, list-learning paradigm on an immediate free recall test. Deese (1959) presented subjects with lists composed of words associated to a critical, nonpresented word (e.g., *thread, pin, eye, sewing, sharp, point, prick, thimble, haystack, thorn, hurt, injection,* which are all associated to *needle*). Presentation of each list was followed immediately by a free recall test. Deese found that for some lists, subjects erroneously recalled the critical nonpresented word. Although the lists were constructed from word association norms by obtaining forward associates to the critical items (i.e., words elicited by the critical nonpresented words on a free association task), Deese hypothesized that backward associations (i.e., associations from the list words to the critical word) were crucial in producing the false recall effect. A correlational analysis of his results confirmed this prediction; the mean probability with which list items elicited the critical item on a free association test correlated .87 with the probability of obtaining an intrusion of the critical nonpresented item for that list.

Roediger and McDermott (1995) adopted Deese's paradigm as a way of studying false recall and false recognition. In one experiment (Roediger & McDermott, 1995, Experiment 2), we presented subjects with 15-word lists, which were constructed in the same manner as those used by Deese. After each of the lists, subjects were instructed either to recall the list or to do math problems. For the lists that were recalled, subjects accurately produced 62% of the words they had studied; 55% of the time they also produced the critical nonpresented item. The critical nonpresented item was recalled at a level comparable to the words in the middle serial positions (usually assumed to reflect retrieval from long-term store). This high intrusion level is very unusual given that (1) the materials used were word lists, which are generally thought to discourage constructive processes; (2) the lists were very short; (3) the test was a free recall test, which is usually resistant to intrusive errors; (4) the test was given immediately after list presentation; and (5) subjects were explicitly instructed not to guess.

In addition, Roediger and McDermott (1995) used a final recognition test to examine other aspects of false memories in this paradigm and obtained remarkably high false alarm rates for the critical lures (.81 for the lists that had been followed by recall tests and .72 for the lists presented but not followed by initial tests). These rates were virtually identical to the hit rates (.79 for the lists recalled previously and .65 for the lists not previously tested). Note too that a testing effect was obtained both for veridical recognition and false recognition: Recognition rates were higher for studied items and lures for lists that had been previously tested than for lists not previously tested. Finally, when Tulving's (1985) *remember/know* procedure was employed, subjects often claimed to remember vividly some specific aspect of the presentation of the critical items; indeed, they made this claim as often for the critical nonpresented items as they did for items actually studied. Thus, Deese's (1959) paradigm provides a powerful method by which to study illusory memories both in recall and recognition. Read (1996), Payne, Elie, Blackwell, and Neuschatz (1996), and Schacter, Verfaellie, and Pradere (1996) have recently conducted similar experiments with this paradigm.

The present experiments were designed to follow up those of Roediger and McDermott (1995) by examining some basic questions about the persistence of illusory memories in this paradigm. In Experiment 1, I examined the effect of test delay on false recall. Two types of delay were examined: (1) a short (30 s) filled delay and (2) a longer, 2-day delay. Firm predictions could be made for studied items in all conditions, with increasing delays adversely affecting performance; the primary question was how false recall would behave relative to veridical recall as a function of delay.

Experiment 2 explored the persistence of these illusory memories by employing a multitrial free recall paradigm to determine whether subjects could eventually, after multiple study/test trials, edit out their intrusions. Subjects were given 5 presentations of a list, with a recall test following each presentation. A final free recall test followed 1 day later. Kay (1955; see also Howe, 1970) asked a similar question; he presented subjects with prose passages and then tested them after a short (5 min) delay. Following the test, he re-presented the material, waited 1 week, tested subjects again, and re-presented the material. This sequence was carried out until six recall tests had occurred. Kay found that subjects had great difficulty editing out intrusions, despite the fact that they had encountered the original, correct material more recently than their erroneous recall.

In summary, the goal of the present experiments was to examine constructive processes in recall. Specifically, I was interested in examining the persistence of false recall in comparison to accurate recall over delays and over multiple trials.

EXPERIMENT 1

For studied items, introduction of a filled delay as short as 30 s eliminates the recency effect in single trial free recall, thereby attenuating overall levels of recall (Glanzer & Cunitz, 1966; Postman & Phillips, 1965). Experiment 1 addressed the question of how such a delay might affect recall of the critical nonpresented word. Subjects were presented with 24 lists; after each of 8 lists they received an immediate free recall test, after each of 8 lists they received a free recall test delayed by 30 s, and after each of 8 lists they took no initial test.

A final free recall test covering all 24 lists was given 2 days later to examine false recall after a substantially longer delay. There is some evidence in the prose memory literature to suggest that false recall might actually increase over long delays; however, the paradigm used here is quite different from the prose recognition and recall paradigms used previously. Sulin and Dooling (1974) found false recognition of schema-consistent prose passages to increase over an interval of 5 min to 1 week. The authors interpreted this result as being consistent with Bartlett's (1932) conception of remembering, which holds that memory for specific details of an event decays over time, whereas schema-consistent information remains relatively intact (see too Posner & Keele, 1970). Spiro (1980) used prose passages to examine the effect of delay on false recall and found that in certain conditions false recall increased from 2 days to 3 weeks.

This 2-day delay condition also allowed for an assessment of whether any differences found between the initial test conditions (an immediate test or a test given after a short, filled delay) might be carried over onto a final free recall test given much later. In addition, the final free recall test permitted an assessment of whether the act of recall on Day 1 would influence recall levels 2 days later. In other words, would recall of items from the lists that had been tested on Day 1 differ from recall of those not initially tested? Evidence from the literature on the testing effect suggests a positive effect of testing would occur for studied items (e.g., Wheeler & Roediger, 1992; see Roediger & Guynn, in press, for a review). The interesting question was whether such an effect might also occur in recall of critical items. Roediger and McDermott (1995) obtained a testing effect for critical lures (e.g., *needle*) on a recognition test given immediately after study of all the lists, but Payne et al. (1996) and Schacter et al. (1996) failed to replicate this outcome. The question here was whether such a testing effect would be found on a final free recall test given after a 2-day delay.

Method

Subjects. Forty-five Rice University undergraduates and summer students participated for pay, at $5 per hour.

Design. The nature of the initial test was manipulated in three within-subjects conditions (Immediate, Delayed, and No Test). Twenty-four study lists were divided arbitrarily (with the stipulation that intrusion rates obtained in previous experiments were approximately equivalent across sets) into three sets of eight lists for purposes of counterbalancing. Across subjects, each set (and therefore each list) was assigned to each condition equally often.

Materials. Twenty-four lists of 15 items each were used for this experiment; they were the same lists as those reported in the Appendix of Roediger and McDermott's (1995) article. There was no overlap in words across lists, and no critical items appeared in any list. Ordering of words within lists was held constant, and the words most strongly associated with the critical items generally occurred toward the beginning of the lists.

Lists were presented on a tape recorder in a male voice at a rate of approximately 1.5 s. Because all subjects heard all lists, only one tape was necessary. Initial test condition occurred in a mixed fashion.

Subjects were provided with two stacks of 4″ × 11″ sheets of paper; one stack contained lined sheets for recall, and one contained sheets with math problems.

Procedure. Subjects were tested individually or in groups of four or fewer. They were seated at tables with the two stacks of paper in front of them. They were told that the purpose of the experiment was to examine the relation between their memory and math abilities. It was explained that they would hear a series of lists (24 in all) presented via the tape player and that they should pay close attention to the lists because after some lists their memory would be tested.

Following presentation of each list, subjects either (1) took an immediate test (i.e., spent 90 s recalling the immediately preceding list and then did 30 s of math problems), (2) took a delayed test (i.e., did math problems for 30 s and then were asked to recall for 90 s), or (3) took no test (i.e., did math problems for 30 s and then were asked to do an additional 90 s of math problems). There were always 2 min between list presentations, and subjects never knew until after a list had been presented whether they would first recall it or do math problems. Furthermore, when a list was followed initially by math problems, subjects did not know until the end of the math period whether they would then be asked to do more math problems or to recall the list.

Instructions given to subjects were designed to disguise the fact that the math problems were included only to instantiate three test conditions (Immediate, Delayed, and None) in hopes that subjects would take the math problems seriously (and thereby not rehearse the list in the delayed condition) and that they would not realize that a final test would be given when they returned 2 days later.

Subjects were informed that they would perform two tasks after presentation of each list. They were told that immediately after list presentation, the experimenter would say ''recall'' or ''math'' to indicate what should be done. When the experimenter said ''recall,'' subjects were to take one of the lined sheets in front of them and write down as many words as they could remember from the immediately preceding list. The experimenter stressed that although the object was to remember as many words as possible, subjects should be confident that everything they wrote down had occurred in the list. They were told that they would be given 1.5 min in which to recall the list, and the stopwatch would beep when this time had passed. When the stopwatch beeped, they were to turn the sheet over. Subjects were further instructed that when the experimenter said ''math,'' they should take the top sheet from the math stack and work problems as quickly and accurately as possible until the stopwatch beeped, at which time they were to turn the sheet over. They were informed that sometimes they would recall and then do math, or do math and then recall; furthermore, they were told that sometimes they would be asked to do math problems twice in a row or to recall twice (this latter condition never actually occurred). It was stressed to the subjects that the experimenter would always tell them what they should be doing and when to do it. Questions were solicited and answered.

Upon completion of the last list and its corresponding test and/or math periods, subjects were dismissed for the day and reminded to return 2 days later for a further, unspecified experiment.

When they returned, subjects were given lined sheets of paper (with 200 lines per subject) and asked to recall all the words they could remember from the first session. They were asked to recall as many words as possible, regardless of whether the list (or the individual word) had been recalled on the first day. As in the first session, they were asked to be certain that every word they wrote down had occurred on the tape player. They were given 15 min to perform the task; after each minute, a tone sounded on a computer, signaling them to draw a line under the last item recalled and to continue attempting to recall until the session was over.

TABLE 1

PROPORTIONS OF STUDIED ITEMS AND CRITICAL NONSTUDIED ITEMS RECALLED ON THE INITIAL TESTS AND THE FINAL FREE RECALL TEST AS A FUNCTION OF INITIALTEST CONDITION

	Type of recall test			
	Initial tests		Final free recall	
Initial test condition	Studied	Critical NS	Studied	Critical NS
Immediate	.58	.44	.18	.23
Delayed	.50	.46	.15	.24
No Test	—	—	.04	.12

Results and Discussion

Initial tests. Mean proportions of studied and critical nonstudied items recalled on the initial tests are shown in the left half of Table 1. Several observations can be made from examining the table. First, although a short, filled delay attenuated the level of veridical recall (compared to the immediate test condition), the delay had no effect on the recall level of the critical items. For studied items, the difference in recall proportions in the immediate condition (.58) and delayed condition (.50) was reliable, $t(44) = 7.71$, $SEM = .009$. As shown in the serial position curve of Fig. 1, this difference was localized in the last few items presented. That is, the 30-s filled delay eliminated the recency effect for studied items, replicating prior research (Glanzer & Cunitz, 1966; Postman & Phillips, 1965).

Although robust levels of false recall were obtained in both conditions, there was no reliable difference between the conditions; the critical items were produced 44% of the time on the immediate test and 46% on the delayed test, $t(44) < 1$. The mean proportion of critical items recalled (collapsing across conditions) is represented in Fig. 1 as the dotted line. (Note that the serial position designation is meaningless for these items because they were never presented.) As apparent in Fig. 1, subjects recalled the critical nonstudied words with a probability comparable to recall of items presented in the middle portion of the list, which is usually thought to represent recall from long-term memory (e.g., Atkinson & Shiffrin, 1968; Glanzer & Cunitz, 1966).

There was a significant interaction between

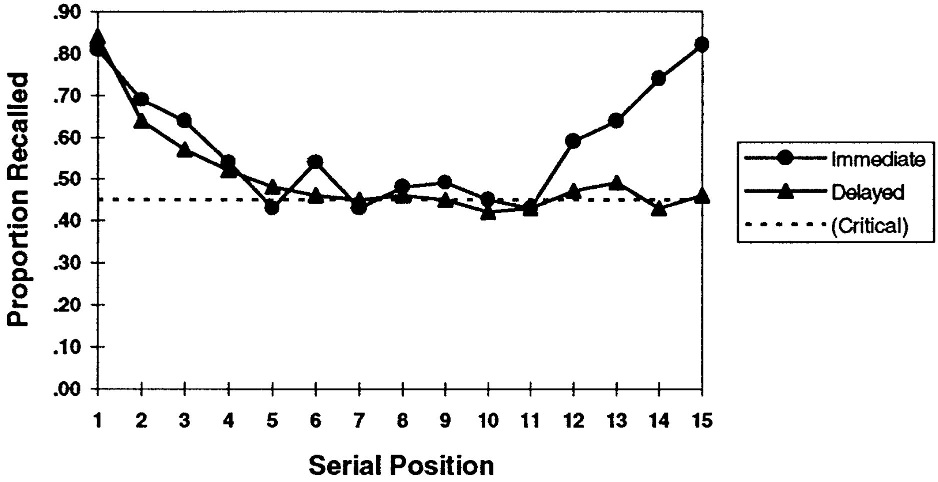

FIG. 1. Solid lines show the probability of recall of studied items on the initial recall tests in Experiment 1 as a function of serial position of the presented items. The dashed line indicates the mean probability of recall of the critical nonstudied items (collapsed across test condition). (There are no serial position data for the critical items because they were not presented.)

the test conditions and item type, $F(1,44) = 5.63$, $MSE = .02$. As discussed above, delaying the initial test affected production levels of list items but not critical intrusions. Another way of looking at the interaction is that on the immediate test, accurate recall exceeded false recall, $t(44) = 3.26$, $SEM = .04$, but after a short delay, the probabilities of false and accurate recall were not significantly different, $t(44) = 1.07$, $SEM = .045$.

Roediger and McDermott (1995) reported that on an immediate test, when the critical items were produced, they tended to occur toward the end of the subjects' output for the list: 63% of the time they occurred in the last fifth of subjects' output. An identical analysis for the immediate condition in the present experiment showed similar, but less dramatic results: 38% of the critical items occurred in the last fifth of subjects' recall protocols. For the lists on which the critical item was recalled, the mean recall position was 6.2 (of 9.6 items written down), or 65% of the way through their output. Figure 2 shows cumulative proportions recalled as a function of output quintiles for the immediate and delayed conditions of this experiment. The basic finding emerging from these data is that when the critical nonpresented item is produced on an immediate free recall test, it tends to appear toward the end of subjects' output.

An examination of the output position data for the 30-s delayed test shows a different pattern of results. When the cumulative proportion recalled is plotted against output quintile, the resulting function is almost perfectly linear, as shown in Fig. 2. This finding indicates that critical items did not occur disproportionately in any particular section of the output protocols. The mean output position was 4.9 (of 8.6 items produced on these tests), or 57% of the way through the subjects' output.

The discrepancy between immediate and delayed tests makes sense if one assumes that a major factor contributing to the late occurrence of critical intrusions found on the immediate test is that subjects begin their recall by producing the items in primary memory, those presented at the end of the list. This recall strategy allows little chance for the critical item to appear in the first part of the recall protocol. When a 30-s filled delay is introduced, and primary memory is removed from subjects' recall, more freedom

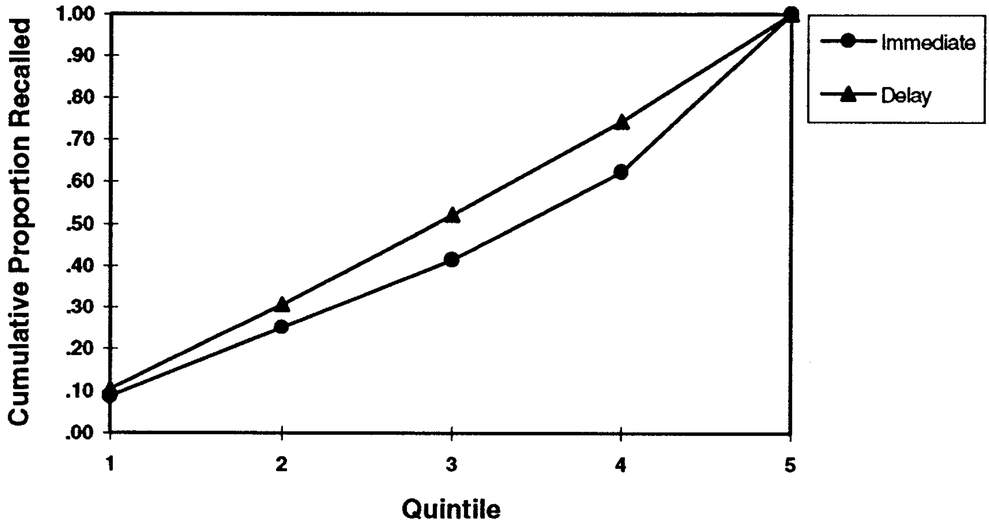

FIG. 2. Cumulative probability of recall of the critical nonpresented item as a function of output position in recall for the immediate and delayed tests of Experiment 1. Quintiles refer to the first 20% of the subject's responses, the second 20%, etc.

is permitted in the output position of the critical nonpresented item.

An analysis of subjects' errors showed that the critical nonpresented items were not the only items falsely recalled by subjects. However, of all possible words, subjects intruded an average of only .22 items per list on the immediate tests and .32 items per list on the delayed test. That is, subjects produced an item that was neither a studied word nor the critical nonpresented word on about every fifth list on the immediate tests and on every third list on the delayed test. This difference was reliable, $t(44) = 2.8$, $SEM = .036$. In general, the intrusions were semantic associates of the presented words. Occasionally, however, subjects wrote a word phonetically similar to a presented item.

In summary, results on the initial tests showed robust levels of false recall, replicating Roediger and McDermott (1995). Furthermore, there was no effect of a short, filled delay on levels of false recall. However, the standard finding that such a delay attenuates accurate recall by eliminating the recency effect was obtained. An analysis of where the critical items occurred when they were produced showed that they generally occurred toward the end of the output for the immediate recall condition (replicating Roediger & McDermott, 1995) but appeared equally distributed throughout the recall protocols for the delayed condition. Finally, intrusions other than the critical nonpresented items occurred infrequently, indicating that subjects were not simply guessing or free associating during the recall period, thereby following their instructions not to guess.

Final test. Results from the final free recall test (given 2 days later) are reported on the right side of Table 1. The overall proportions of studied items and critical items recalled on this test were lower than on the initial tests. However, caution should be used when making this comparison because the tests differed not only in delay but also in that in the first session, an initial test was given after presentation of each of the lists, whereas 2 days later, the test was a final free recall test covering all previously presented lists. Therefore, no direct comparisons are made between the data for the two test sessions.

Perhaps the most remarkable aspect of the final free recall data is that the overall proportion of critical items recalled (M = .20) exceeded the proportion of studied items recalled (M = .12). A 3 (Initial Test) × 2 (Item type) within-subjects ANOVA showed this main effect to be significant, $F(1,44) = 21.28$, $MSE = .02$. The test condition of Day 1 also exerted a significant effect on later recall, $F(2,88) = 42.56$, $MSE = .01$. There was no reliable interaction between test condition and item type, $F(2,88) < 1$, indicating that studied items and critical items behaved similarly as a function of prior testing history: Regardless of whether the lists had been tested immediately after presentation, after a 30-s delay, or not at all, the proportion of list items recalled 2 days later was lower than the proportion of critical intrusions.

Planned comparisons showed that for studied items, subjects recalled more items when they had previously taken an immediate test than when the initial test had been delayed, $t(44) = 2.06$, $SEM = .012$. However, for critical items, there was no difference in recall between the immediate and delayed conditions, $t(44) < 1$. In addition, the immediate and delayed conditions combined produced greater recall probabilities later than did the no test condition for both studied items, $t(44) = 10.59$, $SEM = .01$, and critical items, $t(44) = 5.18$, $SEM = .02$. Thus, the standard benefit from prior testing—the testing effect—was obtained for studied items; more interestingly, the same effect occurred for critical nonpresented items.

It is informative to examine the serial position functions in final free recall as a function of the three initial test conditions, as shown in Fig. 3. Because overall recall levels were low (and therefore noisy across serial positions), this curve has been smoothed by averaging three successive points to obtain each point (except the first and the last). For instance, the mean of the second, third, and fourth positions is represented in the figure as the third position. The first and last positions in the graph, however, are based on raw data. It is clear from this figure that false recall proportions exceeded veridical recall proportions in all conditions for almost all serial positions. Recall that Fig. 1 showed that the probability of false recall approximated the probability of recall of items that had been presented in the middle portion of the lists. Figure 3 shows that by 2 days later, on a final free recall test, critical items no longer behave as if they had been presented in the middle portion of the list; instead, they are recalled at levels comparable to the primacy portion of the curve. It is unclear from this experiment how much of this change in relative recall levels is attributable to the test delay and how much to the difference in the tests (final free recall covering all the lists, instead of recall of each individual list). Nevertheless, recall of the critical nonpresented items differs from recall of items in the middle of the lists.

Depicted in Fig. 4 are the cumulative recall curves for both studied items and critical nonpresented items. This graph shows the cumulative probability of recalling the items plotted for the end of each minute of the final free recall test. The first thing to note from this figure is that the critical items were recalled throughout the recall period, not just toward the end. In addition, the proportion of critical nonpresented items recalled is increasing more rapidly than the proportion of studied items recalled, $F(14,616) = 13.32$, $MSE = .01$. Because one could argue that the interaction might simply be caused by the main effect of item type combined with a floor effect in the first couple minutes of the recall period, an analysis of the last 8 min of the recall period was performed. The item type × minute interaction for this section of the curve was also reliable, $F(7,308) = 4.70$, $MSE = .00$, providing further evidence that the recall curve for critical items is diverging from that for studied items. A final observation derived from this figure is that recall of neither type of item had reached asymptote by the end of the 15-min recall period, indicating that if subjects had been allowed more recall time, they would have recalled more of both types of items.

Finally, an examination of noncritical ex-

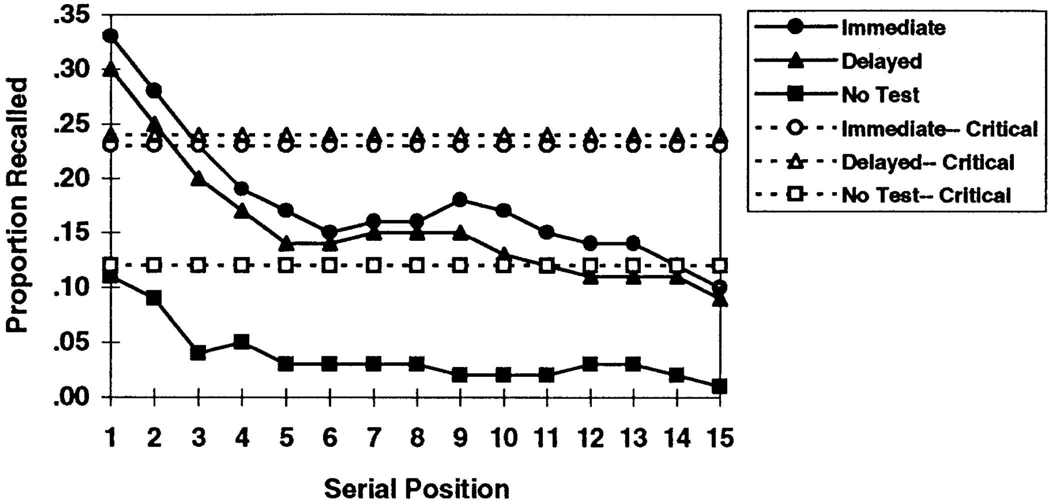

FIG. 3. Solid lines show the probability of recall of studied items on the final free recall test in Experiment 1 as a function of serial position and initial test condition. Dashed lines indicate probability of recall of the critical nonstudied items as a function of initial test condition. (There are no serial position data for the critical items because they were not presented.)

tralist intrusions revealed that on average, subjects produced 3.5 such items on their final free recall test. Therefore, after studying 24 lists, 2 days later they correctly recalled an average of 44.4 (of 360 possible) words, produced 4.7 (of 24 possible) critical intrusions, and produced 3.5 other intrusions (out of all other English words).

In sum, results on the final free recall test of Experiment 1 demonstrate the robustness of false recall in this paradigm. Indeed, after 2 days, the proportion of critical nonpresented items recalled exceeded the proportion of studied items recalled. In addition, the experiment confirms the critical role of testing in determining later false recall.

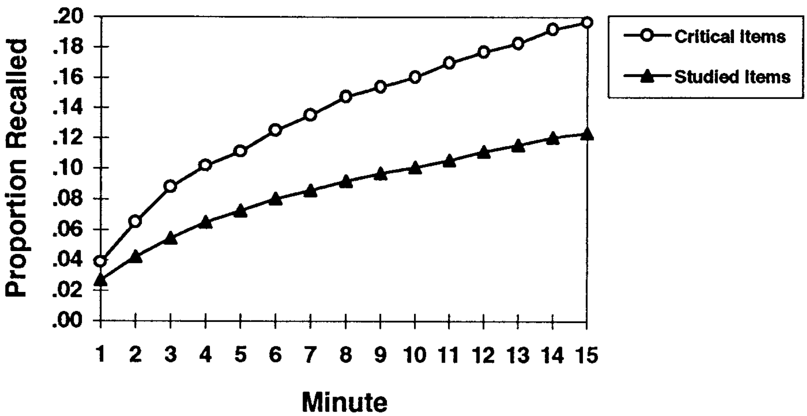

FIG. 4. Cumulative proportions of critical nonpresented items and studied items recalled for each minute of the final free recall test in Experiment 1.

Experiment 2

Experiment 2 was designed to determine whether this sort of false recall could be attenuated or eliminated across multiple study/test trials. Because correct recall would have approached ceiling quickly using only 15 words per list, three 15-word associative sets were presented in each study list. That is, subjects studied one long list of 45 words, which consisted of three groups of 15 associates (e.g., *hot, snow, warm* . . . *bed, rest, awake* . . . *thread, pin, eye* . . .). Items occurred blocked by associative set for half the subjects and in a random order for the other half.

Assuming that these lists function like categorized lists, previous research would indicate that for the studied items, there should be a main effect for ordering, with blocked lists producing higher levels of recall than randomly ordered lists (Dallett, 1964). One question addressed in this experiment was whether there would also be an effect of the blocked/random variable for critical nonpresented items. No firm predictions could be made for this variable because a plausible rationale exists for predicting an effect in either direction. It could be that in trying to remember the list items (during either encoding or retrieval or both), subjects would try to categorize and organize the lists according to common themes within the lists. Obviously, the ''common themes'' may be represented by the critical nonpresented items. Because items from the same ''theme'' are scattered throughout the list in the case of random presentation, subjects might be more confused about which specific items occurred on the random lists than on the blocked lists. If so, they might be expected to intrude the critical nonpresented words more often in random than in blocked lists. Conversely, one could argue that blocked presentation should produce higher false recall levels than random presentation because blocking words might encourage more relational processing of the words and therefore be more likely to elicit the critical word as an implicit associative response (Underwood, 1965) or to induce subjects to assume that the word had been presented. In this analysis, the theme is easier to perceive in the blocked case. Therefore, as stated above, this variable was an exploratory one. An additional question was whether any blocked/random effect (or lack thereof) for critical items would change across trials.

The primary question of the experiment was whether the proportion of critical items recalled would decrease or disappear across trials. A plausible hypothesis was that a decrease would occur because as subjects continue to learn the items present in the list, they may also learn the items not present (or conspicuous in their absence). For example, subjects may produce a critical nonpresented word on a first test but be unsure whether it had actually occurred. They could check for this word during the second presentation. Another interesting possibility was that if subjects recalled a critical item once, they might continue to recall it, carrying over their error despite repeated study opportunities. Recall that Kay (1955) reported this pattern in recall of prose passages. (His general finding was later replicated by Howe (1970) and extended to false recognition (Howe, 1972).)

Kay (1955) used long (i.e., 1 week) delays between successive re-presentations of the passage and the test phases. He assumed long delays were necessary to minimize the influence of the intervening study episodes, thereby allowing intrusions to persist across test trials despite the opportunity to reencounter the study material. Because the false recall effect on immediate tests is very robust with the paradigm used here, I decided to give the tests immediately following list re-presentation to determine whether the intrusions might persist even under conditions that seem likely to promote their disappearance.

Method

Subjects. Forty Rice University students participated in this experiment in exchange for course credit or pay.

Design. A 2 (Ordering: Blocked, Random) × 6 (Test Session: Trials 1–5, Day 2) × 2 (Item Type: Studied, Critical) mixed design was used, with ordering of lists occurring between subjects and the other two variables instantiated within subjects.

Materials. Six of the 15-word lists used in Roediger and McDermott's (1995, Experiment 2) report were selected on the basis of their producing relatively high levels of recall of the critical items. This lists were divided into two 45-word lists of three 15-word associative sets per list. One list contained associates of the 3 critical items *cold, sleep,* and *needle,* and the other list contained associates of *fruit, chair,* and *mountain.* Each subject received one 45-word list. Items within lists occurred in a fixed order, and the same order was used on all five study presentations. For lists presented in a blocked fashion, sets of associates were arbitrarily assigned to positions within the list. For lists presented randomly, items were randomly assigned to 45 positions within the list with the stipulation that no more than three items from one associative set could occur in succession.

For example, in the blocked condition, one list contained the 15 associates to *cold,* followed by the 15 associates to *sleep,* followed by the 15 associates to *needle.* For random presentation, items from the three associative sets occurred randomly (in a fixed order) within the list. There were no demarcations between associative sets within lists, and subjects were not informed about the relations of words within the lists in either condition.

Lists were recorded digitally in stereo using a 22,050-Hz sampling rate and 16-bit resolution. Words occurred at the approximate rate of 1 word per second, and they were presented via LabTec CS 1400 stereo speakers. Subjects were provided with $8\frac{1}{2}'' \times 11''$ sheets of paper on which to record their responses.

Procedure. Subjects were tested in groups of four or fewer. They were told that they were participating in an experiment designed to examine their ability to learn lists of words. It was explained that they would hear a list of 45 words presented on computer speakers, the list would be presented five times, and each presentation would be followed by a recall test. They were told that they would hear the same list every time and that the ordering of words within the list would remain constant. They were told to recall the list by writing down as many words as they could remember with the stipulation that they were to be confident that every word they wrote had been present in the list. They were given 4 min in which to recall the list, and the end of the recall period was indicated by a tone on the computer. At this time, subjects turned over their test sheet and waited for the next presentation. Questions were answered, and list presentation was begun.

Upon returning 1 day later, subjects were instructed to recall as many items as possible from the list they had learned in the previous session. As on the initial tests, they were explicitly warned against guessing. Subjects were given 12 min in which to recall, and a tone sounded on the computer every minute, signaling subjects to draw a line under the last item recalled and continue to try to remember more items. Questions were answered, and the recall period began. Note that there was no presentation of the list on the second day.

Results and Discussion

Recall probabilities (both accurate and erroneous) are portrayed in Fig. 5. Results from the first day of the experiment (the multitrial learning session) are discussed first, followed by an examination of the final recall test given a day later. Finally, conditional recall probabilities are considered.

Day 1. An overall ANOVA indicated that there was no reliable three-way interaction among item type (studied, critical), study order (blocked, random), and test number, $F(4,152) < 1$. Therefore, separate two-way analyses were performed for studied items and critical items to examine the effects of study order and test number.

STUDIED ITEMS. Initial observations of the top two curves in the figure suggest that the predicted pattern of effects occurred for studied items: Recall increased across trials, demonstrating learning of the list, and perfor-

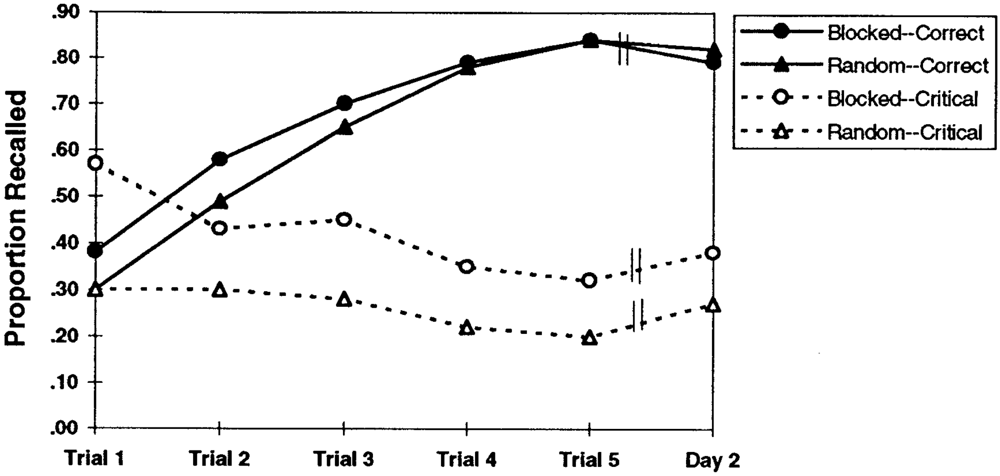

FIG. 5. Probability of recall of studied items (solid lines) and critical nonpresented items (dashed lines) as a function of test trial and presentation order.

mance was generally better in the blocked condition than in the random condition. However, ceiling effects may have masked the blocked/random effect on the last couple of trials.

A 2 (ordering) × 5 (test session) mixed ANOVA on the proportions correct was consistent with these observations. A main effect for test session was found, $F(4,152) = 826.87$, $MSE = .00$, demonstrating learning across trials. The blocked/random variable had no reliable effect, $F(1,38) = 1.87$, $MSE = .05$, but the interaction between test session and list order was reliable, $F(4,152) = 7.17$, $MSE = .002$, consistent with the claim that a ceiling effect obscured the ordering effect in the last few trials.

CRITICAL ITEMS. There were two main questions with respect to critical items in this experiment: (1) To what extent would production levels of critical items decrease across trials? and (2) Would there be a blocked/random effect? An examination of Fig. 5 suggests that the probability of false recall was greater in the blocked condition than in the random condition. False recall also diminished across trials, with proportion of intrusions dropping between Trials 1 and 5 from .57 to .32 for the blocked condition and from .30 to .20 for the random condition. Nevertheless, recall of the critical items was clearly not eliminated, as subjects produced substantial levels of false recall, even after hearing five presentations of the list.

A 2 (ordering) × 5 (trial) mixed ANOVA on the proportions generally confirmed these observations. There was a significant main effect of trial number, $F(4,152) = 4.31$, $MSE = .05$. However, the effect of study order did not reach (but did approach) the criterion for significance, $F(1,38) = 3.54$, $MSE = .38$, $p = .068$. There was no significant interaction $F(4,152) < 1$. Thus, false recall in both the blocked and the random condition decreased at a similar rate across trials. Whether list ordering exerts an effect on false recall is inconclusive if one considers this experiment alone. However, Toglia and his colleagues have recently reported a reliable effect of this variable in the same paradigm, both on an immediate test (Toglia, Neuschatz, Goodwin, & Lyon, 1995a) and after a 1-week delay (Toglia, Neuschatz, Goodwin, & Lyon, 1995b). Therefore, given the findings of Toglia et al. (1995a, 1995b), combined with the marginally reliable effect obtained here, it seems safe to conclude that greater false recall occurs after blocked than after random conditions.

Finally, it is interesting to note that on the first trial in the blocked condition, the proportion of critical nonpresented items recalled (.57) greatly exceeded the proportion of studied items recalled (.38), $t(19) = 2.39$, $SEM = .079$. By Trial 3, however, this pattern had reversed, showing a reliable advantage for studied items, $t(19) = 2.8$, $SEM = .089$.

Day 2. The data points at the far right side of Fig. 5 represent performance on the final free recall test, which occurred 1 day later in the absence of an intervening study opportunity. These data suggest that recall of studied items decreased after 1 day (i.e., items were forgotten), whereas intrusion of critical nonpresented items increased after 1 day. Paired *t* tests confirmed these observations. For studied items, the drop in recall levels between Trial 5 of Day 1 (.84) and Day 2 (.80) was reliable, $t(39) = 2.81$, $SEM = .012$. Conversely, the level of false recall rose between Trials 5 (.26) and Day 2 (.33) after 1 day, $t(39) = 2.08$, $SEM = .032$.

An examination of noncritical intrusions showed that subjects rarely made such errors. Collapsing across trial number and study order, the mean number of noncritical intrusions was .49 items per subject. Thus, subjects were not simply guessing when instructed to recall the lists.

Conditionalized recall probabilities. The data for the critical items in this experiment are interesting in part because subjects had difficulty correcting their errors despite repeated opportunities to do so. However, given the data reported thus far, an alternative interpretation exists; that is, perhaps subjects successfully edited out their initial intrusions on the first few trials, but new errors arose later in the session. Recall that each subject received a list centered around three critical items. Suppose subjects generally recalled two of these items on Trial 1 and (say) by Trial 3 had learned not to include these two but then began to (erroneously) recall the third critical item. This scenario (or a similar one) would produce the obtained pattern of results.

To determine whether such a pattern might underlie these data, I examined new intrusions on each test trial. Shown in Fig. 6 are the overall false recall levels for the blocked and random conditions (repeated from Fig. 5), along with the accompanying proportions of new intrusions recalled on each trial. ‘‘New’’ intrusions were defined as those that had not appeared on any previous test trial. It is evident from this graph that with the exception of the second trial, subjects rarely introduced new intrusions into their protocols; the learning curves are not being artificially inflated by new intrusions occurring in later trials. Therefore, the conclusion that subjects have difficulty correcting their errors seems firmly grounded.

In summary, results of Experiment 2 demonstrated that when given repeated opportunities to hear the study list, subjects could reduce the proportion of critical nonpresented items recalled. However, they did not eliminate these errors. An analysis of conditionalized recall probabilities showed that this latter finding was primarily attributable to subjects' carrying over their initial errors onto later tests, and not to the introduction of new intrusions.

General Discussion

Several noteworthy results were obtained in the two experiments reported here. First, I obtained high levels of false recall following presentation of lists of associated words, thereby replicating the findings of Deese (1959) and Roediger and McDermott (1995) (see also Payne et al., 1996; Read, 1996; and Schacter et al., 1996). The intrusions were obtained despite an explicit warning to subjects against guessing. Second, and more interestingly, the high levels of false recall were maintained over a short delay (30 s) and remained strong 2 days later. Indeed, after 2 days, the probability of false recall of critical nonpresented items exceeded the probability of recall of list items in Experiment 1. Third, both accurate recall and false recall were increased on a final free recall test by prior testing of the list 2 days earlier. Fourth, in a

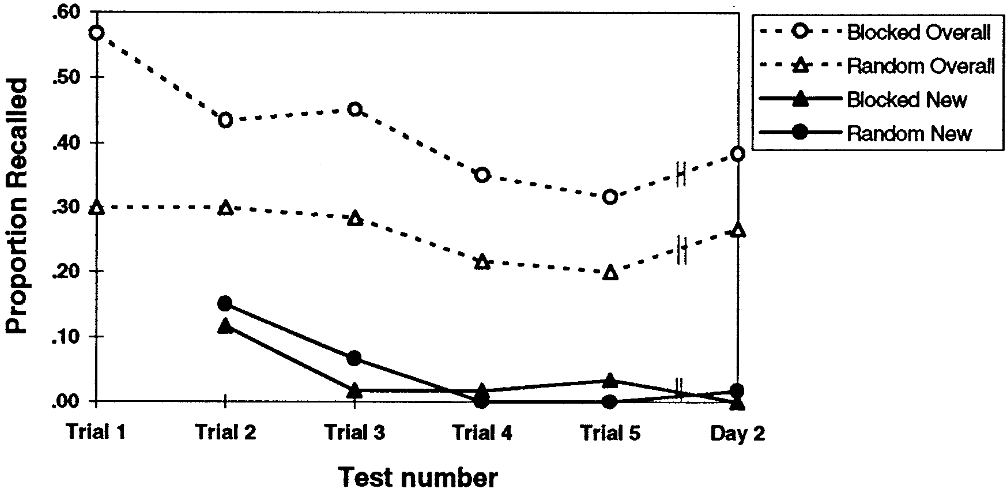

FIG. 6. Overall probability of recall of critical nonpresented items (dashed lines) and the probability of recalling ''new'' critical items as a function of test trial and presentation order.

multitrial learning paradigm, recall of the critical item dropped over trials but stabilized at 20 or 32% (depending on condition), indicating that subjects did not completely edit the erroneous responses even after five study/test cycles. An examination of conditionalized recall probabilities of the intrusions confirmed this claim. Fifth, blocked presentation enhanced accurate recall relative to random presentation; the finding for critical items was less clear-cut, as the blocked/random difference fell short of the criterion for significance. However, given that Toglia et al. (1995a, 1995b) obtained a robust blocked/random difference in a similar paradigm, the effect for critical items likely parallels the effect for studied items.

The present results speak to several issues in the development and maintenance of false memories. I discuss in turn how these results bear on three issues: (1) encoding factors that encourage development of false memories; (2) the critical role of retrieval factors in enhancing false memories; and (3) the persistence of the illusory memories created in this paradigm.

Encoding Factors

Blocked presentation enhances recall of both correct items and critical nonpresented items (see Experiment 2 reported here; Toglia et al., 1995a, 1995b). In addition, Toglia et al. (1995a) varied level of processing during encoding and showed that false recall (like correct recall) occurs more frequently after deep, semantic processing (a pleasantness rating task) than after shallow, graphemic processing (determining whether each word contained an ''a''). Both manipulations of list structure (blocked/random) and level of processing (graphemic, semantic) show that encoding factors affect later probability of false recall. Perhaps it is not too surprising that semantic analysis of meaningfully related words enhances false recall relative to graphemic analysis. More surprising is the outcome that blocked presentation leads to greater false recall than does random presentation. As mentioned in the lead-in to Experiment 2, one might hypothesize that repeated blocked presentation as in Experiment 2 would cause subjects to identify the nature of the list and more easily discern that a particular item they were expecting to hear had not occurred in the list. However, this did not happen. It would be interesting to see if the same pattern would occur if subjects received simultaneous visual presentation of the list, which would make the list structure more readily apparent (e.g., Elmes, Roediger, Wilkinson, & Greener, 1972).

Roediger and McDermott (1995) noted that the earliest idea used to explain false recognition in Underwood's (1965) paradigm also provides a workable account of the more recent findings. Underwood (1965) proposed that an implicit associative response occurring during encoding (see or hear *thread*, and then think *needle*) could explain why subjects would later falsely recognize *needle* as having occurred in the list. In the paradigm used here, subjects hear 15 words derived from a common nonpresented word. The nonpresented word is in turn an associate of many of the list words and therefore may be activated during list presentation. If the activation results in conscious awareness of the nonpresented word, then this might explain why, when Tulving's (1985) *remember/know* procedure is used, subjects claim to remember the occurrence of the critical nonpresented words (Roediger & McDermott, 1995, Experiment 2; Payne et al., 1996; Schacter et al., 1996). The difficulty would resemble that of reality monitoring, or internal source monitoring (Johnson & Raye, 1981; Johnson, Hashtroudi, & Lindsay, 1993), in which subjects must judge whether events they recognize were encountered externally or were internally generated. As Johnson and her colleagues have demonstrated, subjects often confuse imagined events with experienced events (Johnson, 1995).

These newer results can also be interpreted within this framework: If the associated items are presented together at study, then the context may make each presented item more likely to arouse the common associate than if the relevant items are dispersed throughout the list. Similarly, semantic analysis of list items would be more likely to lead to activation of links among associates than would shallower, graphemic analysis. Although the investigation into how study manipulations affect this memory illusion is just beginning, the present results and those of Toglia et al. (1995a, 1995b) fit comfortably into the ideas first proposed by Underwood (1965). However, other results indicate that encoding factors are only partly responsible for these effects. Retrieval factors, considered next, also play a large role.

Retrieval Factors

The study of the effect of taking an initial test on performance on a later test has a long history (e.g., Ballard, 1913; Spitzer, 1939). One general finding is that repeated testing can lead to greater overall recall of the target material under certain conditions, a finding called hypermnesia (Erdelyi & Becker, 1974; see Payne, 1987, for a review). A second consistent finding is the testing effect: taking a test leads to greater recall or recognition on a later test than found in a control condition in which no first test is given (e.g., Spitzer, 1939; Wheeler & Roediger, 1992). However, the effect of testing can also have deleterious consequences on later recall, as seen in these experiments (see too Schooler, Foster, & Loftus, 1988; Payne et al., 1996, Experiment 2). That is, the testing effect exists for false recall as well as for accurate recall; the present experiments show that taking a first test enhances later false recall just as it does veridical recall (see Fig. 3). The act of recall presumably provides retrieval practice for recalled items and makes them more accessible on a later test. Indeed, many experiments have shown that recall of events produces greater facilitation on a later test than does the actual re-presentation of the events (Hogan & Kintsch, 1971; Thompson, Wenger, & Bartling, 1978).

The results of Experiment 1 demonstrate that the same facilitation occurs for false recall as for accurate recall (see Roediger, Jacoby, and McDermott (1996) for further evidence on this point). The impact of initial testing on a later recognition test, however, is less clear-cut. Roediger and McDermott (1995, Experiment 2) found testing effects for studied items and critical nonpresented items on a recognition test. However, using the same general paradigm, Schacter et al. (1996) failed to find a testing effect for studied items or critical items. In addition, in an experiment that was essentially a replication of Roediger and McDermott's (1995) Experiment 2, Payne et al. (1996) failed to find a testing effect for critical items, although the effect was present for studied items. The reason for the discrepant findings within the recognition tests is unclear, but perhaps it is not too surprising that a testing effect for false memories is easier to obtain on a free recall test than a recognition test. This idea is based on the observation that for veridical recall, testing effects are robust on recall tests but difficult to obtain on recognition (see Roediger & Guynn, in press). The crucial point in the current experiments, though, is that in the paradigm used here, initial testing affects later false recall in much the same way that it enhances veridical recall.

An interesting finding emerged in Experiment 2 from a comparison of the last recall trial on Day 1 and the delayed test on Day 2. For both blocked and random conditions, subjects showed a drop in accurate recall but an absolute increase in false recall over time and repeated testing. This finding is reminiscent of the outcome of Posner and Keele's (1970) study of learning to classify dot patterns. Over delays subjects became less efficient in classifying the actual patterns that had been presented but still retained the ability to quickly classify the prototype from which the patterns had been generated (see too Solso & McCarthy, 1981; Spiro, 1980; and Sulin & Dooling, 1974, for similar observations). The present finding, if confirmed by later work, would show absolute increases in false recall over repeated testing, consistent with Bartlett's (1932) anecdotal results and speculations. Indeed, Payne et al. (1996) showed that false recall levels increased across three successive tests in the paradigm used here.

Other research also shows the importance of retrieval in creating false memories. Roediger, Wheeler, and Rajaram (1993) reported an experiment in which subjects studied 60 items and then were given either a free recall test, or they were forced to guess on an initial test (i.e., to produce 60 items that plausibly could have been presented). When asked to judge whether each item they had written actually had been presented on the list, subjects in the forced recall condition often claimed to recognize their intrusions as having been studied, whereas subjects in the free recall condition did not show this effect. Thus, the process of generating responses can potentially create memories for the presentation of these items. Relatedly, Hyman, Husband, and Billings (1995) reported experiments showing that repeatedly thinking about whether a nonoccurring childhood event had happened increased the likelihood that subjects believed that it actually had happened (see too Hyman & Pentland, 1996). These findings converge on the results reported here in demonstrating that the act of retrieval affects not only veridical retention but also false recall. Research in this area is new, but it seems likely from these first experiments that retrieval factors play a critical role in development and maintenance of false memories.

Persistence

The experiments reported here demonstrate that intrusions in free recall tests can be persistent across time and across multiple opportunities for subjects to realize their errors. In Experiment 1 (as well as in the papers by Roediger and McDermott (1995), Payne et al. (1996), and Schacter et al. (1996)), it was shown that on an immediate recall test, probability of recall of the critical nonpresented item approximated the probability of recall of items that had been presented in the middle portion of the list, or items usually thought to represent retrieval from long-term memory. After a 2-day delay, false recall probabilities exceeded recall probabilities from all list positions except the first few (see Fig. 3). Thus, the critical nonpresented items appear to have been retrieved from long-term memory much like presented items and, curiously, the ''forgetting'' of nonpresented events seems to have occurred more slowly than for events that did occur. Indeed, in Experiment 2, the probability of false recall increased over a 1-

day delay, whereas the probability of recall of list items decreased.

Experiment 2 was designed to examine directly the persistence of memory illusions as obtained in this paradigm by providing subjects with multiple study (and test) opportunities. Although subjects were able to decrease the number of false recalls across trials, they were not able to eliminate them completely. In the blocked condition, the probability of false recall greatly exceeded that of veridical recall on the first trial, but by the second trial, this pattern disappeared. In both blocked and random conditions, recall of list items appeared to be near asymptote by Trials 4 and 5, but the false recall probability did not drop below .32 and .20, respectively. Of course, whether additional study/test trials would cause the drop to continue remains to be seen, but clearly the illusion persists under the conditions used in Experiment 2 here.

Perhaps one reason that false recall was not eliminated in Experiment 2 is that the repeated testing aspect of the procedure somewhat counteracted the effect of repeated study presentations. That is, if recall tests serve to enhance procedures used in retrieval for both false and accurate recall, subjects might not have been able to effectively use the subsequent study presentations to eliminate errors. If subjects received only repeated study presentations (without intervening tests) they may be more likely to edit their responses than in the customary study/test procedure. Experiments are currently underway to examine this issue.

Concluding Remarks

In conclusion, the experiments reported here demonstrate robust memory illusions in which nonpresented items are retrieved from long-term memory with greater probability than recall of presented items. The effect is persistent across time as well as across repeated opportunities to learn the lists. Results demonstrate the important roles of both encoding and retrieval processes in producing memory illusions (see too Alba & Hasher, 1983; Hasher & Griffin, 1978; Johnson et al., 1993).

Finally, an interesting puzzle arises from examining the results of the current experiments and some prior experiments; the puzzle concerns the relation between false recall and accurate recall in this paradigm. Under certain conditions, false recall levels vary in the same manner as accurate recall levels, whereas under other conditions, false recall levels vary inversely with accurate recall. For example, enhancing accurate recall sometimes enhances false recall levels, as seen in (1) the effect of prior testing, (2) the benefit of semantic processing (relative to graphemic processing) during encoding, and (3) the benefit of blocked presentation (relative to random presentation). In all three cases, variables increasing the probability of accurate recall also increase that of false recall. However, under other conditions, variables enhancing accurate recall reduce false recall. For example, in the multitrial free recall paradigm of Experiment 2, as subjects learned the items in the lists, they also decreased their levels of false recall. Finally, a third pattern shows decreases in accurate recall can be accompanied by enhanced false recall. This pattern is demonstrated in Experiment 2, in which the probability of accurate recall dropped from Trial 5 on Day 1 to a later test on Day 2, whereas the probability of false recall rose between the two tests. Perhaps under conditions in which accurate and false recall vary together, the benefit shown in accurate recall might be a result of constructive processing. That is, the same processes that enhance veridical recall also drive up false recall. Resolution of when the various patterns are observed must await future research.

References

Alba, J. W., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin,* **93,** 203–231.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2, pp. 89–105). New York: Academic Press.

Ballard, P. B. (1913). Oblivescence and reminiscence. *British Journal of Psychology Monograph Supplements,* **1,** 1–82.

BARTLETT, F. C. (1932). *Remembering: A study in experimental and social psychology.* Cambridge, England: University Press.

BRANSFORD, J. D., & FRANKS, J. J. (1971). The abstraction of linguistic ideas. *Cognitive Psychology,* **2,** 331–350.

COFER, C. N. (1967). Does conceptual organization influence the amount retained in free recall? In B. Kleinmuntz (Ed.), *Concepts and the structure of memory.* New York: Wiley.

DALLETT, K. M. (1964). Number of categories and category information in free recall. *Journal of Experimental Psychology,* **68,** 1–12.

DEESE, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology,* **58,** 17–22.

ELMES, D. G. R., ROEDIGER, H. L., WILKINSON, W. C., & GREENER, W. I. (1972). Positive and negative part/whole transfer in free recall. *Journal of Verbal Learning and Verbal Behavior,* **11,** 251–256.

ERDELYI, M. H., & BECKER, J. (1974). Hypermnesia for pictures: Incremental memory for pictures but not words in multiple recall trials. *Cognitive Psychology,* **6,** 159–171.

GAULD, A. & STEPHENSON, G. M. (1967). Some experiments related to Bartlett's theory of remembering. *British Journal of Psychology,* **58,** 39–49.

GLANZER, M., & CUNITZ, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior,* **5,** 351–360.

HASHER, L., & GRIFFIN, M. (1978). Reconstructive and reproductive processes in memory. *Journal of Experimental Psychology: Human Learning and Memory,* **4,** 318–330.

HOGAN, R. M., & KINTSCH, W. (1971). Differential effects of study and test trials on long-term retention and recall. *Journal of Verbal Learning and Verbal Behavior,* **10,** 562–567.

HOWE, M. J. A. (1970). Repeated presentation and recall of meaningful prose. *Journal of Educational Psychology,* **61,** 214–215.

HOWE, M. J. A. (1972). Repeated presentation and retention of meaningful information. *Psychological Reports,* **31,** 840–842.

HYMAN, I. E., HUSBAND, T. H., & BILLINGS, F. J. (1995). False memories of childhood experiences. *Applied Cognitive Psychology,* **9,** 181–197.

HYMAN, I. E., & PENTLAND, J. (1996). The role of mental imagery in the creation of false childhood memories. *Journal of Memory and Language,* **35,** 101–117.

JOHNSON, M. K. (1995). *The relation between memory and reality.* Paper presented at the 103rd Annual convention of the American Psychological Association, New York, NY, August 1995.

JOHNSON, M. K., HASHTROUDI, S., & LINDSAY, D. S. (1993). Source monitoring. *Psychological Bulletin,* **114,** 3–28.

JOHNSON, M. K., & RAYE, C. L. (1981). Reality monitoring. *Psychological Review,* **88,** 67–85.

KAY, H. (1955). Learning and retaining verbal material. *British Journal of Psychology,* **46,** 81–100.

LOFTUS, E. F., MILLER, D. G., & BURNS, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory,* **4,** 19–31.

LOFTUS, E. F., & PALMER, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior,* **13,** 585–589.

PAYNE, D. G. (1987). Hypermnesia and reminiscence in recall: A historical and empirical review. *Psychological Bulletin,* **101,** 5–27.

PAYNE, D. G., ELIE, C. J., BLACKWELL, J. M., & NEUSCHATZ, J. S. (1996). Memory illusions: Recalling, recognizing, and recollecting events that never occurred. *Journal of Memory and Language,* **35,** 261–285.

POSNER, M. I., & KEELE, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology,* **83,** 304–308.

POSTMAN, L., & PHILLIPS, L. W. (1965). Short-term temporal changes in free recall. *Quarterly Journal of Experimental Psychology,* **17,** 132–138.

READ, J. D. (1996). From a passing thought to a false memory in 2 minutes: Confusing real and illusory events. *Psychonomic Bulletin and Review* **3,** 105–111.

ROEDIGER, H. L., & GUYNN, M. J. (in press). Retrieval processes. In E. L. Bjork and R. A. Bjork (Eds.), *Handbook of perception and cognition,* Vol. 10. New York: Academic Press.

ROEDIGER, H. L., JACOBY, J. D., & MCDERMOTT, K. B. (1996). Misinformation effects in recall: Creating false memories through repeated retrieval. *Journal of Memory and Language,* **35,** 300–318.

ROEDIGER, H. L., & MCDERMOTT, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **21,** 803–814.

ROEDIGER, H. L., & PAYNE, D. G. (1985). Recall criterion does not affect recall level or hypermnesia: A puzzle for generate/recognize theories. *Memory & Cognition,* **13,** 1–7.

ROEDIGER, H. L., WHEELER, M. A., & RAJARAM, S. (1993). Remembering, knowing, and reconstructing the past. In D. L. Medin (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 97–134). Academic Press: San Diego.

SCHACTER, D. L., VERFAELLIE, M., & PRADERE, D. (1996). The neuropsychology of memory illusions: False recall and recognition in amnesic patients. *Journal of Memory and Language,* **35,** 319–334.

SCHOOLER, J. W., FOSTER, R. A., & LOFTUS, E. F. (1988).

Some deleterious consequences of the act of recollection. *Memory & Cognition,* **16,** 243–251.

SOLSO, R. L., & MCCARTHY, J. E. (1981). Prototype formation of faces: A case of pseudomemory. *British Journal of Psychology,* **72,** 499–503.

SPIRO, R. J. (1980). Accomodative reconstruction in prose recall. *Journal of Verbal Learning and Verbal Behavior,* **19,** 84–95.

SPITZER, H. F. (1939). Studies in retention. *Journal of Educational Psychology,* **30,** 641–656.

SULIN, R. A., & DOOLING, D. J. (1974). Intrusion of a thematic idea in retention of prose. *Journal of Experimental Psychology,* **103,** 255–262.

THOMPSON, C. P., WENGER, S. K., & BARTLING, C. A. (1978). How recall facilitates subsequent recall: A reappraisal. *Journal of Experimental Psychology: Human, Learning, and Memory,* **4,** 210–221.

TOGLIA, M. P., NEUSCHATZ, J. S., GOODWIN, K. A., & LYON, M. L. (1995a). *Thematic abstraction and the creation of false memories.* Paper presented at the first meeting of the Society for Applied Research in Memory and Cognition (SARMAC), Vancouver.

TOGLIA, M. P., NEUSCHATZ, J. S., GOODWIN, K. A., & LYON, M. L. (1995b). *The influence of delayed recall on false memories.* Poster presented at the 36th Annual Meetings of the Psychonomic Society, Los Angeles, CA, 1995.

TULVING, E. (1985). Memory and consciousness. *Canadian Psychologist,* **26,** 1–12.

UNDERWOOD, B. J. (1965). False recognition produced by implicit verbal responses. *Journal of Experimental Psychology,* **70,** 122–129.

WHEELER, M. A., & ROEDIGER, H. L. (1992). Disparate effects of repeated testing: Reconciling Ballard's (1913) and Bartlett's (1932) results. *Psychological Science,* **3,** 240–245.

(Received July 31, 1995)
(Revision received December 21, 1995)