Journal of Experimental Psychology:
Learning, Memory, and Cognition
1995, Vol. 21, No. 4, 803-814

0278-7393/95/$3.00

Creating False Memories: Remembering Words Not Presented in Lists

Henry L. Roediger III and Kathleen B. McDermott
Rice University

Two experiments (modeled after J. Deese's 1959 study) revealed remarkable levels of false recall and false recognition in a list learning paradigm. In Experiment 1, subjects studied lists of 12 words (e.g., *bed, rest, awake*); each list was composed of associates of 1 nonpresented word (e.g., *sleep*). On immediate free recall tests, the nonpresented associates were recalled 40% of the time and were later recognized with high confidence. In Experiment 2, a false recall rate of 55% was obtained with an expanded set of lists, and on a later recognition test, subjects produced false alarms to these items at a rate comparable to the hit rate. The act of recall enhanced later remembering of both studied and nonstudied material. The results reveal a powerful illusion of memory: People remember events that never happened.

False memories—either remembering events that never happened, or remembering them quite differently from the way they happened—have recently captured the attention of both psychologists and the public at large. The primary impetus for this recent surge of interest is the increase in the number of cases in which memories of previously unrecognized abuse are reported during the course of therapy. Some researchers have argued that certain therapeutic practices can cause the creation of false memories, and therefore, the apparent "recovery" of memories during the course of therapy may actually represent the creation of memories (Lindsay & Read, 1994; Loftus, 1993). Although the concept of false memories is currently enjoying an increase in publicity, it is not new; psychologists have been studying false memories in several laboratory paradigms for years. Schacter (in press) provides an historical overview of the study of memory distortions.

Bartlett (1932) is usually credited with conducting the first experimental investigation of false memories; he had subjects read an Indian folktale, "The War of the Ghosts," and recall it repeatedly. Although he reported no aggregate data, but only sample protocols, his results seemed to show distortions in subjects' memories over repeated attempts to recall the story. Interestingly, Bartlett's repeated reproduction results never have been successfully replicated by later researchers (see Gauld & Stephenson, 1967; Roediger, Wheeler, & Rajaram, 1993); indeed, Wheeler and Roediger (1992) showed that recall of prose passages (including "The War of the Ghosts") actually improved over repeated tests (with very few errors) if short delays occurred between study and test.[1]

Nonetheless, Bartlett's (1932) contribution was an enduring one because he distinguished between *reproductive* and *reconstructive* memory. Reproductive memory refers to accurate, rote production of material from memory, whereas reconstructive memory emphasizes the active process of filling in missing elements while remembering, with errors frequently occurring. It generally has been assumed that the act of remembering materials rich in meaning (e.g., stories and real-life events) gives rise to reconstructive processes (and therefore errors), whereas the act of remembering more simplified materials (e.g., nonsense syllables, word lists) gives rise to reproductive (and thus accurate) memory. Bartlett (1932) wrote that "I discarded nonsense materials because, among other difficulties, its use almost always weights the evidence in favour of mere rote recapitulation" (p. 204).

The investigators of false memories have generally followed Bartlett's (1932) lead. Most evidence has been collected in paradigms that use sentences (Bransford & Franks, 1971; Brewer, 1977), prose passages (Sulin & Dooling, 1974), slide sequences (Loftus, Miller, & Burns, 1978), or videotapes (Loftus & Palmer, 1974). In all these paradigms, evidence of false memories has been obtained, although the magnitude of the effect depends on the method of testing (McCloskey & Zaragoza, 1985; Payne, Toglia, & Anastasi, 1994). The predominance of materials that tell a story (or can be represented by a script or schema) can probably be attributed to the belief that only such materials will cause false memories to occur.

There is one well-known case of false memories being produced in a list learning paradigm: Underwood (1965) introduced a technique to study false recognition of words in

This research was supported by Grant F49620-92-J-0437 from the Air Force Office of Scientific Research. We thank Ron Haas and Lubna Manal for aid in conducting this research. Also, we thank Endel Tulving for bringing the Deese (1959) report to our attention. The manuscript benefited from comments by Doug Hintzman, Steve Lindsay, Suparna Rajaram, and Endel Tulving.

Correspondence concerning this article should be addressed to Henry L. Roediger III or Kathleen B. McDermott, Department of Psychology, MS 25, Rice University, 6100 S. Main Street, Houston, Texas 77005-1892. Electronic mail may be sent via Internet to roddy@rice.edu or mcdermo@ricevm1.rice.edu.

[1] Bartlett's (1932) results from the serial reproduction paradigm—in which one subject recalls an event, the next subject reads and then recalls the first subject's report, and so on—replicates quite well (e.g., I. H. Paul, 1959). However, the repeated reproduction research, in which a subject is tested repeatedly on the same material, is more germane to the study of false memories in an individual over time. To our knowledge, no one has successfuly replicated Bartlett's observations in this paradigm with instructions that emphasize remembering (see Gauld & Stevenson, 1967).

lists. He gave subjects a continuous recognition task in which they decided if each presented word had been given previously in the list. Later words bore various relations to previously studied words. Underwood showed that words associatively related to previously presented words were falsely recognized. Anisfeld and Knapp (1968), among others, replicated the phenomenon. Although there have been a few reports of robust false recognition effects (Hintzman, 1988), in many experiments the false recognition effect was either rather small or did not occur at all. For example, in a study by L. M. Paul (1979), in which synonyms were presented at various lags along with other, unrelated lures, the false recognition effect was only 3% (a 20% false-alarm rate for synonyms and a 17% rate for unrelated lures). Gillund and Shiffrin (1984) failed to find any false recognition effect for semantically related lures in a similar paradigm. In general, most research on the false recognition effect in list learning does little to discourage the belief that more natural, coherent materials are needed to demonstrate powerful false memory effects. Interestingly, most research revealing false memory effects has used recognition measures; this is true both of the prose memory literature (e.g., Bransford & Franks, 1971; Sulin & Dooling, 1974) and the eyewitness memory paradigm (Loftus et al., 1978; McCloskey & Zaragoza, 1985). Reports of robust levels of false recall are rarer.

We have discovered a potentially important exception to these claims, one that reveals false recall in a standard list learning paradigm. It is represented in an experimental report published by Deese in 1959 that has been largely overlooked for the intervening 36 years, despite the fact that his observations would seem to bear importantly on the study of false memories. Deese's procedure was remarkably straightforward; he tested memory for word lists in a single-trial, free-recall paradigm. Because this paradigm was just gaining favor among experimental psychologists at that time and was the focus of much attention during the 1960s, the neglect of Deese's report is even more surprising. However, since the Social Science Citation Index began publication in 1969, the article has been cited only 14 times, and only once since 1983. Most authors mentioned it only in passing, several authors apparently cited it by mistake, and no one has followed up Deese's interesting observations until now, although Cramer (1965) reported similar observations and did appropriately cite Deese's (1959) article. (While working on this article, we learned that Don Read was conducting similar research, which is described briefly in Lindsay & Read, 1994, p. 291.)[2]

Deese (1959) was interested in predicting the occurrence of extralist intrusions in single-trial free recall. To this end, he developed 36 lists, with 12 words per list. Each list was composed of the 12 primary associates of a critical (nonpresented) word. For example, for the critical word *needle,* the list words were *thread, pin, eye, sewing, sharp, point, pricked, thimble, haystack, pain, hurt,* and *injection.* He found that some of the lists reliably induced subjects to produce the critical nonpresented word as an intrusion on the immediate free recall test. Deese's interest was in determining why some lists gave rise to this effect, whereas others did not. His general conclusion was that the lists for which the associations went in a backward (as well as forward) direction tended to elicit false recall. That is, he measured the average probability with which people produced the critical word from which the list was generated when they were asked to associate to the individual words in the list. For example, subjects were given *sewing, point, thimble,* and so on, and the average probability of producing *needle* as an associate was measured. Deese obtained a correlation of .87 between the probability of an intrusion in recall (from one group of subjects) and the probability of occurrence of the word as an associate to members of the list (from a different group). Our interest in Deese's materials was in using his best lists and developing his paradigm as a way to examine false memory phenomena.

Our first goal was to try to replicate Deese's (1959) finding of reliable, predictable extralist intrusions in a single-trial, free-recall paradigm. We found his result to be surprising in light of the literature showing that subjects are often extremely accurate in recalling lists after a single trial, making few intrusions unless instructed to guess (see Cofer, 1967; Roediger & Payne, 1985). As previously noted, most prior research on false memory phenomena has employed measures of recognition memory or cued recall. Deese's paradigm potentially offers a method to study false recollections in free recall. However, we also extended Deese's paradigm to recognition tests. In Experiment 1, we examined false recall and false recognition of the critical nonpresented words and the confidence with which subjects accepted or rejected the critical nonpresented words as having been in the study lists. In Experiment 2, we tested other lists constructed to produce extralist intrusions in single-trial free recall, to generalize the finding across a wider set of materials. In addition, we examined the extent to which the initial false recall of items led to later false recognition of those same items. Finally, we employed the remember–know procedure developed by Tulving (1985) to examine subjects' phenomenological experience during false recognition of the critical nonpresented items. We describe this procedure more fully below.

Experiment 1

The purpose of Experiment 1 was to replicate Deese's (1959) observations of false recall by using six lists that produced among the highest levels of erroneous recall in his experiments. Students heard and recalled the lists and then received a recognition test over both studied and nonstudied items, including the critical nonpresented words.

Method

Subjects. Subjects were 36 Rice University undergraduates who participated as part of a course project during a regular meeting of the class, Psychology 308, Human Memory.

Materials. We developed six lists from the materials listed in Deese's (1959) article. With one exception, we chose the six targets that produced the highest intrusion rates in Deese's experiment: *chair, mountain, needle, rough, sleep,* and *sweet.* As in Deese's experiment, for each critical word, we constructed the corresponding list by obtaining

[2] Some people know of Deese's (1959) paper indirectly because Appleby (1986) used it as the basis of a suggested classroom demonstration of déjà vu.

the first 12 associates listed in Russell and Jenkins's (1954) word association norms. For example, the list corresponding to *chair* was *table, sit, legs, seat, soft, desk, arm, sofa, wood, cushion, rest,* and *stool.* In a few instances, we replaced 1 of the first 12 associates with a word that seemed, in our judgment, more likely to elicit the critical word. (The lists for Experiment 1 are included in the expanded set of lists for Experiment 2 reported in the Appendix.)

The 42-item recognition test included 12 studied and 30 nonstudied items. There were three types of nonstudied items, or lures: (a) the 6 critical words, from which the lists were generated (e.g., *chair*), (b) 12 words generally unrelated to any items on the six lists, and (c) 12 words weakly related to the lists (2 per list). We drew the weakly related words from Positions 13 and below in the association norms; for example, we chose *couch* and *floor* for the *chair* list. We constructed the test sequence in blocks; there were 7 items per block, and each block corresponded to a studied list (2 studied words, 2 related words, 2 unrelated words, and the critical nonstudied lure). The order of the blocks corresponded to the order in which lists had been studied. Each block of test items always began with a studied word and ended with the critical lure; the other items were arranged haphazardly in between. One of the two studied words that were tested occurred in the first position of the study list (and therefore was the strongest associate to the critical item); the other occurred somewhere in the first 6 positions of the study list.

Procedure. Subjects were tested in a group during a regular class meeting. They were instructed that they would hear lists of words and that they would be tested immediately after each list by writing the words on successive pages of examination booklets. They were told to write the last few items first (a standard instruction for this task) and then to recall the rest of the words in any order. They were also told to write down all the words they could remember but to be reasonably confident that each word they wrote down did in fact occur in the list (i.e., they were told not to guess). The lists were read aloud by the first author at the approximate rate of 1 word per 1.5 s. Before reading each list, the experimenter said "List 1, List 2," and so on, and he said "recall" at the end of the list. Subjects were given 2.5 min to recall each list.

After the sixth list, there was brief conversation lasting 2–3 min prior to instructions for the recognition test. At this point, subjects were told that they would receive another test in which they would see words on a sheet and that they were to rate each as to their confidence that it had occurred on the list. The 4-point rating scale was 4 for *sure that the item was old* (or studied), 3 for *probably old,* 2 for *probably new,* and 1 for *sure it was new.* Subjects worked through the recognition test at their own pace.

At the end of the experiment, subjects were asked to raise their hands if they had recognized six particular items on the test, and the critical lures were read aloud. Most subjects raised their hands for several items. The experimenter then informed them that none of the words just read had actually been on the list and the subjects were debriefed about the purpose of the experiment, which was a central topic in the course.

Results

Recall. The mean probability of recall of the studied words was .65, and the serial position curve is shown in Figure 1. The curve was smoothed by averaging data from three adjacent points for each position because the raw data were noisy with only six lists. For example, data from the third, fourth, and fifth points contributed to the fourth position in the graph. The first and the last positions, however, were based only on the raw data. The serial position curve shows marked recency, indicating that subjects followed directions in recalling the last items

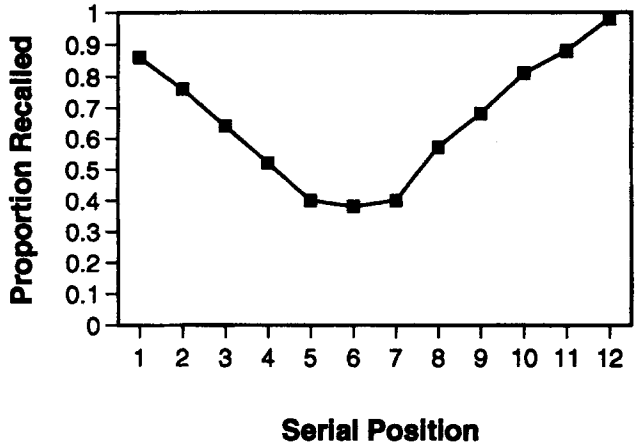

Figure 1. Probability of correct recall in Experiment 1 as a function of serial position. Probability of recall of the studied words was .65, and probability of recall of the critical nonpresented item was .40.

first. A strong primacy effect is also apparent, probably because the strongest associates to the critical target words occurred early in the list. The critical omitted word was recalled with a probability of .40, or with about the same probability as items that had been presented in the middle of the list (see Figure 1). Therefore, items that were not presented were recalled at about the same rate as those that were presented, albeit those in the least favorable serial positions.

The average output position for recall of the critical nonpresented word was 6.9 (out of 8.6 words written down in lists in which there was a critical intrusion). The cumulative production levels of the critical intrusion for those trials on which they occurred is shown in Figure 2 across quintiles of subjects' responses. The critical intrusion appeared only 2% of the time in the first fifth of subjects' output but 63% of the time in the last quintile. Thus, on average, subjects recalled the critical nonstudied item in the last fifth of their output, at the 80th percentile of recalled words (6.9 ÷ 8.6 × 100).

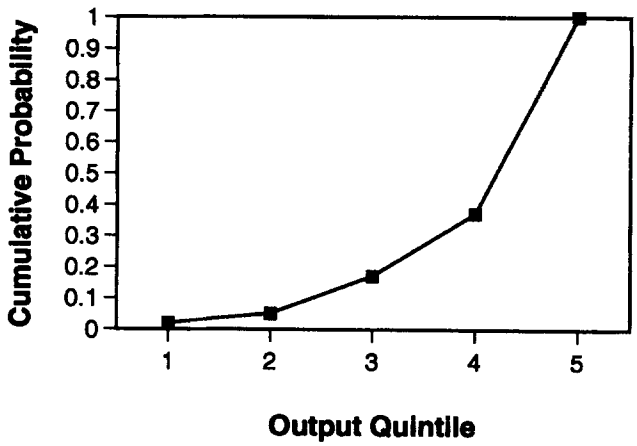

Figure 2. Recall of the critical intrusion as a function of output position in recall. Quintiles refer to the first 20% of responses, the second 20%, and so on.

Other intrusions also occurred in recall, albeit at a rather low rate. Subjects intruded the critical lure on 40% of the lists, but any other word in the English language was intruded on only 14% of the lists. Therefore, subjects were not guessing wildly in the experiment; as usual in single-trial free recall, the general intrusion rate was quite low. Nonetheless, subjects falsely recalled the critical items at a high rate.

Recognition. The recognition test was given following study and recall of all six lists, and thus the results were likely affected by prior recall. (We consider this issue in Experiment 2.) The proportion of responses for each of the four confidence ratings are presented in Table 1 for studied (old) items and for the three different types of lures: unrelated words, weakly related words, and the critical words from which the lists were derived. Consider first the proportion of items subjects called old by assigning a rating of 3 (*probably old*) or 4 (*sure old*). The hit rate was 86% and the false-alarm rate for the standard type of unrelated lures was only 2%, so by usual criteria subjects showed high accuracy. The rate of false alarms was higher for the weakly related lures (.21) than for the unrelated lures, $t(35) = 7.40$, $SEM = .026$, $p < .001$. This outcome replicates the standard false-recognition effect first reported by Underwood (1965). The false-recognition rate for weakly related lures was greater than obtained in many prior studies (e.g., L. M. Paul, 1979), and the rate for the critical nonpresented words was dramatically larger than the rate for the weakly related words. As shown in Table 1, the false-alarm rate for the critical nonstudied lures (.84) approached the hit rate (.86), $t(35) < 1$, $SEM = .036$, *ns*.

Consider next the results based on subjects high-confidence responses (i.e., when they were sure the item had appeared in the study list and rated it a "4"). The proportion of unrelated and weakly related lures falling into this category approached zero. However, subjects were still sure that the critical nonstudied items had been studied over half the time (.58). The hit rate for the studied items remained quite high (.75) and was reliably greater than the false-alarm rate for the critical lures, $t(35) = 3.85$, $SEM = .044$, $p < .001$. It is also interesting to look at the rates at which subjects classified items as *sure new*. Unrelated lures were correctly rejected with high confidence 80% of the time. Related lures received this classification only 44% of the time, and critical lures were confidently rejected at an even lower rate, 8%, which is similar to the rate for studied words (5%).

Table 1
Recognition Results for Experiment 1: The Proportion of Items Classified As Sure Old (a Rating of 4), Probably Old (3), Probably New (2), or Sure New (1) and the Mean Ratings of Items As a Function of Study Status

Study status	Old		New		Mean rating
	4	3	2	1	
Studied	.75	.11	.09	.05	3.6
Nonstudied					
Unrelated lure	.00	.02	.18	.80	1.2
Weakly related lure	.04	.17	.35	.44	1.8
Critical lure	.58	.26	.08	.08	3.3

Table 1 also presents the mean ratings for the four types of items on the 4-point scale. This measure seems to tell the same story as the other two: The mean rating of the critical lures (3.3) approached that of studied items (3.6); the difference did reach significance, $t(35) = 2.52$, $SEM = .09$, $p < .05$. In general, the judgments subjects provided for the critical lures appeared much more similar to those of studied items than to the other types of lures.

Discussion

The results of Experiment 1 confirmed Deese's (1959) observation of high levels of false recall in a single-trial, free-recall task, albeit with six lists that were among his best. We found that the critical nonpresented items were recalled at about the same level as items actually presented in the middle of the lists. This high rate of false recall was not due to subjects guessing wildly. Other intrusions occurred at a very low rate. In addition, we extended Deese's results to a recognition test and showed that the critical nonpresented items were called old at almost the same level as studied items (i.e., the false-alarm rate for the critical nonpresented items approximated the hit rate for the studied items). The false-alarm rate for the critical nonpresented items was much higher than for other related words that had not been presented. Finally, more than half the time subjects reported that they were sure that the critical nonstudied item had appeared on the list. Given these results, this paradigm seems a promising method to study false memories. Experiment 2 was designed to further explore these false memories.

Experiment 2

We had four aims in designing Experiment 2. First, we wanted to replicate and extend the recall and recognition results of Experiment 1 to a wider set of materials. Therefore, we developed twenty-four 15-item lists similar to those used in Experiment 1 and in Deese's (1959) experiment. (We included expanded versions of the six lists used in Experiment 1.) Second, we wanted to examine the effect of recall on the subsequent recognition test. In Experiment 1 we obtained a high level of false recognition for the critical nonpresented words, but the lists had been recalled prior to the recognition test, and in 40% of the cases the critical item had been falsely recalled, too. In Experiment 2, we examined false recognition both for lists that had been previously recalled and for those that had not been recalled. Third, we wanted to determine the false-alarm rates for the critical nonpresented items when the relevant list had not been presented previously (e.g., to determine the false-alarm rate for *chair* when related words had not been presented in the list). Although we considered it remote, the possibility existed that the critical nonpresented items simply elicit a high number of false alarms whether or not the related words had been previously presented.

The fourth reason—and actually the most important one—for conducting the second experiment was to obtain subjects' judgments about their phenomenological experience while recognizing nonpresented items. We applied the procedure developed by Tulving (1985) in which subjects are asked to

distinguish between two states of awareness about the past: remembering and knowing. When this procedure is applied in conjunction with a recognition test, subjects are told (a) to judge each item to be old (studied) or new (nonstudied) and (b) to make an additional judgment for each item judged to be old: whether they remember or know that the item occurred in the study list. A *remember* experience is defined as one in which the subject can mentally relive the experience (perhaps by recalling its neighbors, what it made them think of, what they were doing when they heard the word, or physical characteristics associated with its presentation). A *know* judgment is made when subjects are confident that the item occurred on the list but are unable to reexperience (i.e., remember) its occurrence. In short, remember judgments reflect a mental reliving of the experience, whereas know judgments do not. There is now a sizable literature on remember and know judgments (see Gardiner & Java, 1993; Rajaram & Roediger, in press), but we will not review it here except to say that evidence exists that remember–know judgments do not simply reflect two states of confidence (high and low) because variables can affect remember–know and confidence (sure–unsure) judgments differently (e.g., Rajaram, 1993).

Our purpose in using remember–know judgments in Experiment 2 was to see if subjects who falsely recognized the critical nonpresented words would report accompanying remember experiences, showing that they were mentally reexperiencing events that never occurred. In virtually all prior work on false memories, it has been assumed that subjects' incorrect responses indicated false remembering. However, if Tulving's (1985) distinction is accepted, then responding on a memory test should not be equated with remembering. Further metamemorial judgments such as those obtained with the remember–know procedure are required to determine if subjects are remembering the events. In fact, in most experiments using the remember–know procedure, false alarms predominantly have been judged as know responses (e.g., Gardiner, 1988; Jones & Roediger, 1995). This outcome would be predicted in our experiment, too, if one attributes false recognition to a high sense of familiarity that arises (perhaps) through spreading activation in an associative network. Therefore, in Experiment 2 we examined subjects' metamemorial judgments with respect to their false memories to see whether they would classify these memories as being remembered or known to have occurred.

In Experiment 2, subjects were presented with 16 lists; after half they received an immediate free recall test, and after the other half they did math problems. After all lists had been presented, subjects received a recognition test containing items from the 16 studied lists and 8 comparable lists that had not been studied. During the recognition test, subjects made old–new judgments, followed by remember–know judgments for items judged to be old.

Method

Subjects. Thirty Rice University undergraduates participated in a one hour session as part of a course requirement.

Materials. We developed 24 lists from Russell and Jenkins's (1954) norms in a manner similar to that used for Experiment 1. For each of 24 target words, 15 associates were selected for the list. These were usually the 15 words appearing first in the norms, but occasionally we substituted other related words when these seemed more appropriate (i.e., more likely to elicit the nonpresented target as an associate). The ordering of words within lists was held constant; the strongest associates generally occurred first. An example of a list for the target word *sleep* is: *bed, rest, awake, tired, dream, wake, night, blanket, doze, slumber, snore, pillow, peace, yawn, drowsy.* All the lists, corrected for a problem noted in the next paragraph, appear in the Appendix.

The 24 lists were arbitrarily divided into three sets for counterbalancing purposes. Each set served equally often in the three experimental conditions, as described below. The reported results are based on only 7 of the 8 lists in each set because the critical items in 2 of the lists inadvertently appeared as studied items in other lists; dropping 1 list in each of two sets eliminated this problem and another randomly picked list from the third set was also dropped, so that each scored set was based on 7 lists. With these exceptions, none of the critical items occurred in any of the lists.

Design. The three conditions were tested in a within-subjects design. Subjects studied 16 lists; 8 lists were followed by an immediate free recall test, and 8 others were not followed by an initial test. The remaining 8 lists were not studied. Items from all 24 lists appeared on the later recognition test. On the recognition test, subjects judged items as old (studied) or new (nonstudied) and, when old, they also judged if they remembered the item from the list or rather knew that it had occurred.

Procedure. Subjects were told that they would be participating in a memory experiment in which they would hear lists of words presented by means of a tape player. They were told that after each list they would hear a sound (either a tone or a knock, with examples given) that would indicate whether they should recall items from the list or do math problems. For half of the subjects, the tone indicated that they should recall the list, and the knock meant they should perform math problems; for the other half of the subjects, the signals were reversed. They were told to listen carefully to each list and that the signal would occur after the list had been presented; therefore, subjects never knew during list presentation whether the list would be recalled. Words were recorded in a male voice and presented approximately at a 1.5-s rate. Subjects were given 2 min after each list to recall the words or to perform multiplication and division problems. Recall occurred on 4 inch by 11 inch sheets of paper, and subjects turned over each sheet after the recall period, so the recalled items were no longer in view. The first part of the experiment took about 45 min.

The recognition test occurred about 5 min after the test or math period for the 16th list. During this time, subjects were given instructions about making old–new and remember–know judgments. They were told that they would see a long list of words, some of which they had heard during the earlier phase of the experiment. They were to circle either the word *old* or *new* next to each test item to indicate whether the item had been presented by means of the tape player. If an item was judged old, subjects were instructed that they should further distinguish between remembering and knowing by writing an *R* or *K* in the space beside the item. Detailed instructions on the remember–know distinction were given, modeled after those of Rajaram (1993). Essentially, subjects were told that a remember judgment should be made for items for which they had a vivid memory of the actual presentation; know judgments were reserved for items that they were sure had been presented but for which they lacked the feeling of remembering the actual occurrence of the words. They were told that a remember judgment would be made in cases in which they remembered something distinctive in the speaker's voice when he said the word, or perhaps they remembered the item presented before or after it, or what they were thinking when they heard the word. They were always told to make the remember–know judgment about a word with respect to its presentation on the tape recorder, not whether they

remembered or knew they had written it down on the free recall test. In addition, they were instructed to make remember–know judgments immediately after judging the item to be old, before they considered the next test item.

The recognition test was composed of 96 items, 48 of which had been studied and 48 of which had not. The 48 studied items were obtained by selecting 3 items from each of the 16 presented lists (always those in Serial Positions 1, 8, and 10). The lures, or nonstudied items, on the recognition test were 24 critical lures from all 24 lists (16 studied, 8 not) and the 24 items from the 8 nonstudied lists (again, from Serial Positions 1, 8, and 10). The 96 items were randomly arranged on the test sheet and beside each item were the words *old* and *new;* if subjects circled old, they made the remember–know judgment by writing *R* or *K* in the space next to the word. All subjects received exactly the same test sheet; counterbalancing of lists was achieved by having lists rotated through the three conditions (study + recall, study + arithmetic, and nonstudied) across subsets of 10 subjects.

After the recognition test, the experimenter asked subjects an open-ended question: whether they "knew what the experiment was about." Most subjects just said something similar to "memory for lists of words," but 1 subject said that she noticed that the lists seemed designed to make her think of a nonpresented word. She was the only subject who had no false recalls of the critical nonpresented words; her results were excluded from those reported below and replaced by the results obtained from a new subject. After the experiment, subjects were debriefed.

Results

Recall. Subjects recalled the critical nonpresented word on 55% of the lists, which is a rate even higher than for the 6 lists used in Experiment 1. The higher rate of false recall in Experiment 2 may have been due to the longer lists, to their slightly different construction, to the fact that 16 lists were presented rather than only 6, or to different signals used to recall the lists. In addition, in Experiment 1 the lists were read aloud by the experimenter, whereas in Experiment 2 they were presented by means of a tape player. Regardless of the reason or reasons for the difference, the false-recall effect was quite robust and seems even stronger under the conditions of Experiment 2.

The smoothed serial position curve for studied words is shown in Figure 3, where marked primacy and recency effects are again seen. As in Experiment 1, subjects recalled the critical nonpresented items at about the rate of studied items presented in the middle of the lists. Subjects recalled items in Positions 4–11 an average of 47% of the time, compared with 55% recall of nonpresented items. Therefore, recall of the critical missing word was actually greater than recall for studied words in the middle of the list; this difference was marginally significant, $t(29) = 1.80$, $SEM = .042$, $p = .08$, two-tailed.

Recognition. After subjects had heard all 16 lists, they received the recognition test and provided remember–know judgments for items that were called old on the test. We first consider results for studied words and then turn to the data for the critical nonpresented lures.

Table 2 presents the recognition results for items studied in the list. (Keep in mind that we tested only three items from each list [i.e., those in Positions 1, 8, and 10].) It is apparent that the hit rate in the study + recall condition (.79) was greater than in the study + arithmetic condition (.65), $t(29) = 5.20$, $SEM = .027$, $p < .001$, indicating that the act of recall enhanced later recognition. Further, the boost in recognition from prior recall was reflected in a greater proportion of remember responses, which differed reliably, $t(29) = 4.87$, $SEM = .033$, $p < .001$. Know responses did not differ between conditions, $t(29) < 1$. The false-alarm rate for items from the nonstudied lists was .11, with most false positives judged as know responses.

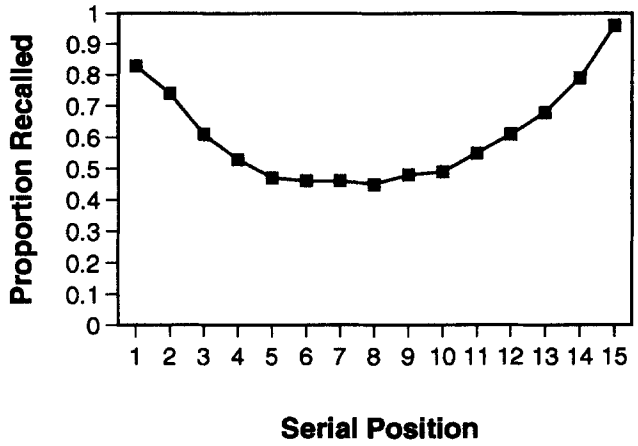

Figure 3. Probability of correct recall in Experiment 2 as a function of serial position. Probability of recall of the studied words was .62, and probability of recall of the critical nonpresented item was .55.

Recognition results for the critical nonpresented lures are also shown in Table 2. The first striking impression is that the results for false-alarm rates appear practically identical to the results for hit rates. Therefore, to an even greater extent than in Experiment 1, subjects were unable to distinguish items actually presented from the critical lures that were not presented. Table 2 also shows that the act of (false) recall in the study + recall condition enhanced later false recognition relative to the study + arithmetic condition, in which the lists were not recalled. In addition, after recalling the lists subjects were much more likely to say that they remembered the items from the list, with remember judgments being made 72% of the time (i.e., .58 ÷ .81 × 100) for words that had never been presented. When the lists were presented but not recalled, the

Table 2
Recognition Results for Studied Items and Critical Lures in Experiment 2

Item type and condition	Proportion of Old responses		
	Overall	R	K
Studied			
Study + recall	.79	.57	.22
Study + arithmetic	.65	.41	.24
Nonstudied	.11	.02	.09
Critical lure			
Study + recall	.81	.58	.23
Study + arithmetic	.72	.38	.34
Nonstudied	.16	.03	.13

Note. R = remember judgment; K = know judgment.

rate of remember judgments dropped to 53%, although this figure is still quite high. Interestingly, the corresponding percentages for items actually studied were about the same: 72% for remember judgments for lists that were recalled and 63% for lists that were not recalled.

One point that vitiates the correspondence between the results for studied and nonstudied items in Table 2 is that the false-alarm rates for the types of items differed when the relevant lists had not been studied. The rate for the regular list words was .11, whereas the rate for the critical lures (when the relevant prior list had not been studied) was .16, $t(29) = 2.27$, $SEM = .022$, $p = .03$, two-tailed. However, the difference was not great, and in both cases false alarms gave rise to more know responses than remember responses.

One further analysis is of interest. In the study + recall condition, we can consider recognition results for items that were produced in the recall phase (whether representing correct responding or false recall) relative to those that were not produced. Although correlational, such results provide an interesting pattern in comparing the effects of prior correct recall to prior false recall on later recognition. Table 3 shows the results of this analysis, including the means for studied items and for the critical items. For the studied items, recognition of items that had been correctly recalled was essentially perfect, and most old responses were judged to be remembered. Items not produced on the recall test were recognized half the time, and responses were evenly divided between remember and know judgments. These effects could have been due to the act of recall, to item selection effects, or to some combination. Nonetheless, they provide a useful point of comparison for the more interesting results about the fate of falsely recalled items, as shown in Table 3.

The recognition results for the falsely recalled critical items closely resemble those for correctly recalled studied items. The probability of recognizing falsely recalled items was quite high (.93), and most of these items were judged to be remembered (.73) rather than known (.20). More remarkably, the critical items that were not produced were later (falsely) recognized at a higher rate (.65) than were items actually studied but not produced (.50); this difference was marginally significant, $t(29) = 1.81$, $SEM = .083$, $p = .08$, two-tailed. In addition, these falsely recognized items were judged to be remembered in 58% of the cases (i.e., .38 ÷ .65 × 100), or at about the same rate as for words that were studied but not produced (52%). These analyses reveal again the powerful false memory effects at work in this paradigm, with people falsely remembering the critical nonstudied words at about the same levels (or even greater levels) as presented words.

Table 3
Proportion of Items Judged to be Old on the Recognition Test in the Study + Recall Condition of Experiment 2 As a Function of Whether the Items Were Produced on the Immediate Free Recall Test

Condition	Production rate of free recall	Recognition		
		Overall	R	K
Studied				
Produced	.62	.98	.79	.19
Not produced	.38	.50	.26	.24
Critical lure				
Produced	.55	.93	.73	.20
Not produced	.45	.65	.38	.27

Note. R = remember judgment; K = know judgment.

General Discussion

The primary results from our experiments can be summarized as follows: First, the paradigm we developed from Deese's (1959) work produced high levels of false recall in single-trial free recall. In Experiment 1, with 12-word lists, subjects recalled the critical nonstudied word after 40% of the lists. In Experiment 2, with 15-word lists, false recall increased, occurring on 55% of the occasions. Second, this paradigm also produced remarkably high levels of false recognition for the critical items; the rate of false recognition actually approached the hit rate. Third, the false recognition responses were frequently made with high confidence (Experiment 1) or were frequently accompanied by remember judgments (Experiment 2). Fourth, the act of recall increased both accurate recognition of studied items and the false recognition of the critical nonstudied items. The highest rates of false recognition and the highest proportion of remember responses to the critical nonstudied items occurred for those items that had been falsely recalled.

We discuss our results (a) in relation to prior work and (b) in terms of theories that might explain the basic effects. We then discuss (c) how the phenomenological experience of remembering events that never happened might occur, and (d) what implications our findings might have for the wider debates on false memories.

Relation to Prior Work

Prior work by Underwood (1965) has shown false recognition for lures semantically related to studied words, but as we noted in the introduction, these effects were often rather small in magnitude. In our experiments, we found very high levels of false recall and false recognition. Our recognition results are similar to those obtained by investigators in the 1960s and 1970s who used prose materials and found erroneous recognition of related material. For example, Bransford and Franks (1971) presented subjects with sentences that were related and created a coherent scene (e.g., The rock rolled down the mountain and crushed the hut. The hut was tiny). Later, they confidently recognized sentences that were congruent with the meaning of the complex idea, although the sentences had not actually been presented (e.g., The rock rolled down the mountain and crushed the tiny hut). Similarly, Posner and Keele (1970) showed subjects dot patterns that were distortions from a prototypic pattern. Later, they recognized the prototype (that had never been presented) at a high rate, and forgetting of the prototype showed less decline over a week than did dot patterns actually presented. Jenkins, Wald, and Pittenger (1986) reported similar observations with pictorial stimuli.

In each of the experiments just described, and in other related experiments (see Alba & Hasher, 1983, for a review),

subjects recognized events that never happened if the events fit some general schema derived from the study experiences. A similar interpretation is possible for our results, too, although most researchers have assumed that schema-driven processes occur only in prose materials. Yet the lists for our experiments were generated as associates to a single word and therefore had a coherent form (e.g., words related to sleep or to other similar concepts). The word *sleep,* for example, may never have been presented in the list, but was the "prototype" from which the list was generated, and therefore our lists arguably encouraged schematic processing.

Although our results are similar to those of other research revealing errors in memory, several features distinguish our findings. First, we showed powerful false memory effects in both recall and recognition within the same paradigm. The findings just cited, and others described below, all used recognition paradigms. Although some prior studies have reported false recall (e.g., Brewer, 1977; Hasher & Griffin, 1979; Spiro, 1980), these researchers used prose materials. Second, we showed that subjects actually claimed to remember most of the falsely recognized events as having occurred on the list. The items did not just evoke a feeling of familiarity but were consciously recollected as having occurred. Third, we showed that the effect of prior recall increased both accurate and false memories and that this effect of recall was reflected in remember responses.

Explanations of False Recall and False Recognition

How might false recall and false recognition arise in our paradigm? Actually, the earliest idea about false recognition—the implicit associative response—still seems workable in helping to understand these phenomena, although today we can elaborate on the idea with new models now available. Underwood (1965) proposed that false recognition responses originated during encoding when subjects, seeing a word such as *hot,* might think of an associate (*cold*). Later, if *cold* were presented as a lure, they might claim to recognize its occurrence in the list because of the earlier implicit associative response.

Some writers at the time assumed that the associative response had to occur consciously to the subject during study, so it was implicit only in the sense that it was not overtly produced. Another possible interpretation is that the subject never even becomes aware of the associative response during study of the lists, so that its activation may be implicit in this additional sense, too. Activation may spread through an associative network (e.g., Anderson & Bower, 1973; Collins & Loftus, 1975), with false-recognition errors arising through residual activation. That is, it may not be necessary for subjects to consciously think of the associate while studying the list for false recall and false recognition to occur. On the other hand, the predominance of remember responses for the critical lures on the later recognition test may indicate that the critical nonpresented words do occur to subjects during study of the list. That may be why subjects claim to remember them, through a failure of reality monitoring (Johnson & Raye, 1981).

In further support of the idea that associative processes are critically important in producing false recall, Deese (1959) showed that the likelihood of false recall in this paradigm was predicted well by the probability that items presented in the list elicited the critical nonpresented word in free association tests. In other words, the greater the likelihood that list members produced the critical nonpresented target word as an associate, the greater the level of false recall (see also Nelson, Bajo, McEvoy, & Schreiber, 1989). It is worth noting that some of Deese's lists that contained strong forward associations—including the famous "butterfly" list used in later research—did not lead to false recall. The particular characteristics of the lists that lead to false memories await systematic experimental study, but in general Deese reported that the lists that did not lead to false recall contained words that did not produce the critical targets as associates. The butterfly list did not elicit even one false recall in Deese's experiment.

If false recall and false recognition are produced by means of activation of implicit associative responses, then the reason our false-recognition results were more robust than those usually reported may be that we used lists of related words rather than single related words. Underwood (1965) and others had subjects study single words related to later lures on some dimension, and they showed only modest levels of false recognition, or in some cases none at all (Gillund & Shiffrin, 1984). In the present experiments, subjects studied lists of 12–15 items and the false-recognition effect was quite large. Hall and Kozloff (1973), Hintzman (1988), and Shiffrin, Huber, and Marinelli (1995) have shown that false recognition is directly related to the number of related words in a list. For example, Hintzman (1988, Experiment 1) presented from 0 to 5 items from a category in a list and showed that both accurate recognition of studied category members, as well as false recognition of lures from that category, increased as a function of category size. False recognition increased from about 8% when no category members were included in the list to around 35% when five category members occurred in the list. (These percentages were estimated from Hintzman's Figure 11.) Our lists were not categorized, strictly speaking, but the words were generally related. For our 15-item lists in Experiment 2 that did not receive recall tests, false recognition was 72%; the corresponding figure for recalled lists was 81%. It will be interesting to see if longer versions of standard categorized lists will produce false recognition at the same levels as the lists we have used and whether the average probability that items in the list evoke the lure as an associate will predict the level of false recognition. We are now conducting experiments to evaluate these hypotheses.

If the errors in memory occurring on both recall and recognition tests arise from associative processes, then formal models of associative processing might be expected to predict them. At least at a general level, they would seem to do so. For example, the search of associative memory (SAM) model, first proposed by Raaijmakers and Shiffrin (1980) and later extended to recognition by Gillund and Shiffrin (1984), provides for the opportunity of false recognition (and presumably recall) by means of associative processes. Although it was not the main thrust of their paper, Shiffrin et al. (1995) demonstrated that the SAM model did fit their observation of an increased tendency to produce false alarms to category mem-

bers with increases in the number of category exemplars presented.

Recently, McClelland (in press) has extended the parallel distributed processing (PDP) approach to explaining constructive memory processes and memory distortions. This model assumes that encoding and retrieval occur in a parallel distributed processing system in which there are many simple but massively interconnected processing units. Encoding an event involves the activation of selected units within the system. Retrieval entails patterns of reactivation of the same processing units. However, because activation in the model can arise from many sources, a great difficulty (for the model and for humans) lies in the failure to differentiate between possible sources of prior activation (McClelland, in press). Therefore, because what is encoded and stored is a particular pattern of activity, subjects may not be able to reconstruct the actual event that gave rise to this activity. For example, if presenting the words associated with *sleep* mimics the activity in the system as occurs during actual presentation of the word *sleep,* then the PDP system will be unable to distinguish whether or not the word actually occurred. Consequently, the PDP system would give rise to false memory phenomena, as McClelland (in press) describes.

As the examples above show, associative models can account for false-recall and false-recognition results, although we have not tried fitting specific models to our data. To mention two other models based on different assumptions, Hintzman's (1988) MINERVA 2 model, which assumes independent traces of events, modeled well the effect of increasing category size on the probability of identifying an item from the category as old; this was true both for correct recognition and false recognition. In addition, Reyna and Brainerd (1995) have also applied their fuzzy-trace theory to the problem of false memories.

Although most theorists have assumed that the false memory effects arise during encoding, all remembering is a product of information both from encoding and storage processes (the memory trace) and from information in the retrieval environment (Tulving, 1974). Indeed, false remembering may arise from repeated attempts at retrieval, as shown in Experiment 2 and elsewhere (e.g., Ceci, Huffman, Smith, & Loftus, 1994; Hyman, Husband, & Billings, 1995; Roediger et al., 1993). Retrieval processes may contribute significantly to the false recall and false recognition phenomena we have observed. Subjects usually recalled the critical word toward the end of the set of recalled items, so prior recall may trigger false recall, in part. Also, in the recognition test, presentation of words related to a critical lure often occurred prior to its appearance on the test; therefore, activation from these related words on the test may have enhanced the false recognition effect by priming the lure (Neely, Schmidt, & Roediger, 1983). The illusion of memory produced by this mechanism, if it exists, may be similar to illusions of recognition produced by enhanced perceptual fluency (Whittlesea, 1993; Whittlesea, Jacoby, & Girard, 1990). Indeed, one aspect of our results on which the theories outlined above remain mute is the phenomenological experience of the subjects: They did not just claim that the nonpresented items were familiar; rather, they claimed to remember their occurrence. We turn next to this aspect of the data.

Phenomenological Experience

In virtually all previous experiments using the remember–know procedure, false alarms have been predominantly labeled as know experiences (e.g., Gardiner & Java, 1993; Jones & Roediger, 1995; Rajaram, 1993). The typical assumption is that know responses arise through fluent processing, when information comes to mind easily, but the source of the information is not readily apparent (Rajaram, 1993). In addition, Johnson and Raye (1981) have noted that memories for events that actually occurred typically provide more spatial and temporal details than do memories for events that were only imagined. For these reasons, when we conducted Experiment 2 we expected that the false alarms in our recognition tests would, like other recognition errors, be judged by subjects to be known but not remembered. Yet our results showed that, in our paradigm, this was not so. Subjects frequently reported remembering events that never happened. Clearly, false memories can be the result of conscious recollection and not only of general familiarity.

Furthermore, in our current experiments we found that the act of recall increased both overall recognition and remembering of presented items and of the critical nonpresented items. We assume that generation of an item during a free recall test solidifies the subject's belief that memory for that item is accurate and increases the likelihood of later recognition of the item; why, however, should recall enhance the phenomenological experience of remembering the item's presentation? The enhanced remember responses may be due to subjects' actually remembering the experience of recalling the item, rather than studying it, and confusing the source of their remembrance; similarly, it could be that subjects remember thinking about the item during the study phase and confuse this with having heard it. Each of these mistakes would represent a source monitoring error (Johnson, Hashtroudi, & Lindsay, 1993). Note that our instructions to subjects about their remember–know responses specified that they were to provide remember judgments only when they remembered the item's actual presentation in the list (i.e., not simply when they remembered producing it on the recall test). Nonetheless, despite this instruction, subjects provided more remember responses for items from lists that had been recalled in Experiment 2.

The most promising approach to explaining such false remembering comes from an attributional analysis of memory, as advocated by Jacoby, Kelley, and Dywan (1989). They considered cases in which the aftereffects of past events were misattributed to other sources, but more importantly for present concerns, they considered cases in which subjects falsely attributed current cognitive experience to a concrete past event when that event did not occur. They hypothesized that the ease with which a person is able to bring events to mind increases the probability that the person will attribute the experience to being a memory. They also argued that the greater the vividness and distinctiveness of the generated event, the greater the likelihood of believing that it represents

a memory (Johnson & Raye, 1981). Thus, in our paradigm, if subjects fluently generate (in recall) or process (in recognition) the word *sleep* (on the basis of recent activation of the concept) and if this fluency allows them to construct a clear mental image of how the word would have sounded if presented in the speaker's voice, then they would likely claim to remember the word's presentation. The act of recall increases the ease of producing an event and may thereby increase the experience of remembering. Jacoby et al.'s (1989) analysis offers promising leads for further research.

Implications

The results reported in this article identify a striking memory illusion. Just as perceptual illusions can be compelling even when people are aware of the factors giving rise to the illusion, we suspect that the same is true in our case of remembering events that never happened. Indeed, informal demonstration experiments with groups of sophisticated subjects, such as wily graduate students who knew we were trying to induce false memories, also showed the effect quite strongly.

Bartlett (1932) proposed a distinction between reproductive and reconstructive memory processes. Since then, the common assumption has been that list learning paradigms encourage rote reproduction of material with relatively few errors, whereas paradigms using more coherent (schematic) material (e.g., sentences, paragraphs, stories, or scenes) are necessary to observe constructive processes in memory retrieval. Yet we obtained robust false memory effects with word lists, albeit with ones that contain related words. We conclude that any contrast between reproductive and reconstructive memory is ill-founded; all remembering is constructive in nature. Materials may differ in how readily they lead to error and false memories, but these are differences of a quantitative, not qualitative, nature.

Do our results have any bearing on the current controversies raging over the issue of allegedly false memories induced in therapy? Not directly, of course. However, we do show that the illusion of remembering events that never happened can occur quite readily. Therefore, as others have also pointed out, the fact that people may say they vividly remember details surrounding an event cannot, by itself, be taken as convincing evidence that the event actually occurred (Johnson & Suengas, 1989; Schooler, Gerhard, & Loftus, 1986; Zaragoza & Lane, 1994). Our subjects confidently recalled and recognized words that were not presented and also reported that they remembered the occurrence of these events. A critic might contend that because these experiments occurred in a laboratory setting, using word lists, with college student subjects, they hold questionable relevance to issues surrounding more spectacular occurrences of false memories outside the lab. However, we believe that these are all reasons to be more impressed with the relevance of our results to these issues. After all, we tested people under conditions of intentional learning, with very short retention intervals, in a standard laboratory procedure that usually produces few errors, and we used college students—professional memorizers—as subjects. In short, despite conditions much more conducive to veridical remembering than those that typically exist outside the lab, we found dramatic evidence of false memories. When less of a premium is placed on accurate remembering, and when people know that their accuracy in recollecting cannot be verified, they may even be more easily led to remember events that never happened than they are in the lab.

References

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

Anderson, J. R., & Bower, G. H. (1973). *Human associative memory.* Washington, DC: V. H. Winston.

Anisfeld, M., & Knapp, M. (1968). Association, synonymity, and directionality in false recognition. *Journal of Experimental Psychology, 77,* 171–179.

Appleby, D. (1986). Déjà vu in the classroom. *Network, 4,* 8.

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

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

Brewer, W. F. (1977). Memory for the pragmatic implications of sentences. *Memory & Cognition, 5,* 673–678.

Ceci, S. J., Huffman, M. L. C., Smith, E., & Loftus, E. F. (1994). Repeatedly thinking about non-events. *Consciousness and Cognition, 3,* 388–407.

Cofer, C. N. (1967). Does conceptual organization influence the amount retained in free recall? In B. Kleinmuntz (Ed.), *Concepts and the structure of memory* (pp. 181–214). New York: Wiley.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82,* 407–428.

Cramer, P. (1965). Recovery of a discrete memory. *Journal of Personality and Social Psychology, 1,* 326–332.

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

Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition, 16,* 309–313.

Gardiner, J. M., & Java, R. I. (1993). Recognizing and remembering. In A. Collins, S. Gathercole, & P. Morris (Eds.), *Theories of memory* (pp. 168–188). Hillsdale, NJ: Erlbaum.

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

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1–67.

Hall, J. F., & Kozloff, E. E. (1973). False recognitions of associates of converging versus repeated words. *American Journal of Psychology, 86,* 133–139.

Hasher, L., & Griffin, M. (1979). Reconstructive and reproductive processes in memory. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 318–330.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95,* 528–551.

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

Jacoby, L. L., Kelley, C. M., & Dywan, J. (1989). Memory attributions. In H. L. Roediger III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 391–422). Hillsdale, NJ: Erlbaum.

Jenkins, J. J., Wald, J., & Pittenger, J. B. (1986). Apprehending pictorial events: An instance of psychological cohesion. In V. McCabe & G. J. Balzano (Eds.), *Event cognition: An ecological perspective* (pp. 117–133). Hillsdale, NJ: Erlbaum.

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.

Johnson, M. K., & Suengas, A. G. (1989). Reality monitoring judgments of other people's memories. *Bulletin of the Psychonomic Society, 27,* 107–110.

Jones, T. C., & Roediger, H. L., III. (1995). The experiential basis of serial position effects. *European Journal of Cognitive Psychology, 7,* 65–80.

Lindsay, D. S., & Read, J. D. (1994). Psychotherapy and memories of childhood sexual abuse: A cognitive perspective. *Applied Cognitive Psychology, 8,* 281–338.

Loftus, E. F. (1993). The reality of repressed memories. *American Psychologist, 48,* 518–537.

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.

McClelland, J. L. (in press). Constructive memory and memory distortions: A parallel-distributed processing approach. In D. L. Schacter, J. T. Coyle, G. D. Fischbach, M. M. Mesulam, & L. E. Sullivan (Eds.), *Memory distortion.* Cambridge, MA: Harvard University Press.

McCloskey, M., & Zaragoza, M. (1985). Misleading postevent information and memory for events: Arguments and evidence against memory impairment hypotheses. *Journal of Experimental Psychology: General, 114,* 1–16.

Neely, J. H., Schmidt, S. R., & Roediger, H. L., III. (1983). Inhibition from related primes in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 196–211.

Nelson, D. L., Bajo, M., McEvoy, C. L., & Schreiber, T. A. (1989). Prior knowledge: The effects of natural category size on memory for implicitly encoded concepts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 957–967.

Paul, I. H. (1959). Studies in remembering: The reproduction of connected and extended verbal material. *Psychological Issues, 1* (Monograph 2), 1–152.

Paul, L. M. (1979). Two models of recognition memory: A test. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 45–51.

Payne, D. G., Toglia, M. P., & Anastasi, J. S. (1994). Recognition performance level and the magnitude of the misinformation effect in eyewitness memory. *Psychonomic Bulletin & Review, 1,* 376–382.

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

Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 14, pp. 207–262). New York: Academic Press.

Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition, 21,* 89–102.

Rajaram, S., & Roediger, H. L., III. (in press). Remembering and knowing as states of consciousness during recollection. In J. D. Cohen & J. W. Schooler (Eds.), *Scientific approaches to the question of consciousness.* Hillsdale, NJ: Erlbaum.

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences, 7,* 1–75.

Roediger, H. L., III, & 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., III, 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). San Diego, CA: Academic Press.

Russell, W. A., & Jenkins, J. J. (1954). *The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test.* (Tech. Rep. No. 11, Contract N8 ONR 66216, Office of Naval Research). University of Minnesota.

Schacter, D. L. (in press). Memory distortion: History and current status. In D. L. Schacter, J. T. Coyle, G. D. Fischbach, M. M. Mesulam, & L. E. Sullivan (Eds.), *Memory distortion.* Cambridge, MA: Harvard University Press.

Schooler, J. W., Gerhard, D., & Loftus, E. F. (1986). Qualities of the unreal. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 171–181.

Shiffrin, R. M., Huber, D. E., & Marinelli, K. (1995). Effects of category length and strength on familiarity in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 267–287.

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

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

Tulving, E. (1974). Cue-dependent forgetting. *American Scientist, 62,* 74–82.

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., III. (1992). Disparate effects of repeated testing: Reconciling Ballard's (1913) and Bartlett's (1932) results. *Psychological Science, 3,* 240–245.

Whittlesea, B. W. A. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 1235–1253.

Whittlesea, B. W. A., Jacoby, L. L., & Girard, K. (1990). Illusions of immediate memory: Evidence of an attributional basis for feelings of familiarity and perceptual quality. *Journal of Memory and Language, 29,* 716–732.

Zaragoza, M. S., & Lane, S. M. (1994). Source misattributions and the suggestibility of eyewitness memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 934–945.

(Appendix follows on next page)

Appendix

The Twenty-Four 15-Word Lists Used in Experiment 2

Within-lists words were presented in this order, which generally follows the association norms. (We replaced two words here for reasons described in the *Method* section of Experiment 2.)

Anger	Black	Bread	Chair	Cold	Doctor	Foot	Fruit
mad	white	butter	table	hot	nurse	shoe	apple
fear	dark	food	sit	snow	sick	hand	vegetable
hate	cat	eat	legs	warm	lawyer	toe	orange
rage	charred	sandwich	seat	winter	medicine	kick	kiwi
temper	night	rye	couch	ice	health	sandals	citrus
fury	funeral	jam	desk	wet	hospital	soccer	ripe
ire	color	milk	recliner	frigid	dentist	yard	pear
wrath	grief	flour	sofa	chilly	physician	walk	banana
happy	blue	jelly	wood	heat	ill	ankle	berry
fight	death	dough	cushion	weather	patient	arm	cherry
hatred	ink	crust	swivel	freeze	office	boot	basket
mean	bottom	slice	stool	air	stethoscope	inch	juice
calm	coal	wine	sitting	shiver	surgeon	sock	salad
emotion	brown	loaf	rocking	Arctic	clinic	smell	bowl
enrage	gray	toast	bench	frost	cure	mouth	cocktail

Girl	High	King	Man	Mountain	Music	Needle	River
boy	low	queen	woman	hill	note	thread	water
dolls	clouds	England	husband	valley	sound	pin	stream
female	up	crown	uncle	climb	piano	eye	lake
young	tall	prince	lady	summit	sing	sewing	Mississippi
dress	tower	George	mouse	top	radio	sharp	boat
pretty	jump	dictator	male	molehill	band	point	tide
hair	above	palace	father	peak	melody	prick	swim
niece	building	throne	strong	plain	horn	thimble	flow
dance	noon	chess	friend	glacier	concert	haystack	run
beautiful	cliff	rule	beard	goat	instrument	thorn	barge
cute	sky	subjects	person	bike	symphony	hurt	creek
date	over	monarch	handsome	climber	jazz	injection	brook
aunt	airplane	royal	muscle	range	orchestra	syringe	fish
daughter	dive	leader	suit	steep	art	cloth	bridge
sister	elevate	reign	old	ski	rhythm	knitting	winding

Rough	Sleep	Slow	Soft	Spider	Sweet	Thief	Window
smooth	bed	fast	hard	web	sour	steal	door
bumpy	rest	lethargic	light	insect	candy	robber	glass
road	awake	stop	pillow	bug	sugar	crook	pane
tough	tired	listless	plush	fright	bitter	burglar	shade
sandpaper	dream	snail	loud	fly	good	money	ledge
jagged	wake	cautious	cotton	arachnid	taste	cop	sill
ready	snooze	delay	fur	crawl	tooth	bad	house
coarse	blanket	traffic	touch	tarantula	nice	rob	open
uneven	doze	turtle	fluffy	poison	honey	jail	curtain
riders	slumber	hesitant	feather	bite	soda	gun	frame
rugged	snore	speed	furry	creepy	chocolate	villain	view
sand	nap	quick	downy	animal	heart	crime	breeze
boards	peace	sluggish	kitten	ugly	cake	bank	sash
ground	yawn	wait	skin	feelers	tart	bandit	screen
gravel	drowsy	molasses	tender	small	pie	criminal	shutter

Received August 17, 1994
Revision received December 12, 1994
Accepted December 14, 1994 ■