JOURNAL OF MEMORY AND LANGUAGE **39,** 508–520 (1998)
ARTICLE NO. ML982582

Attempting to Avoid Illusory Memories: Robust False Recognition of Associates Persists under Conditions of Explicit Warnings and Immediate Testing

Kathleen B. McDermott and Henry L. Roediger, III

Washington University

Roediger and McDermott (1995) showed that presentation of associated words can induce false recognition of a related, but nonpresented, associate. In three experiments, we placed this related associate in half of the study lists (but not in the other half) in an attempt to determine whether (and under what conditions) subjects could distinguish between cases in which the critical associate was and was not presented. Results suggest that subjects are quite poor at performing this straightforward task, even when explicitly informed of the false recognition phenomenon, instructed to pay careful attention to whether or not the critical linking associate was presented in the list, and given a 1-item recognition test immediately following each 15-word list. Although subjects were not able to perform accurately under these conditions, the warning instruction did attenuate the false recognition effect (relative to an uninformed condition). This illusion of memory appears to be remarkably robust and little affected by the instructional manipulations.

Roediger and McDermott (1995) reported experiments showing that presentation of words (e.g., *door, glass, pane, shade, ledge*) that are associates of a nonpresented word (e.g., *window*) can induce high levels of false recall (see also Deese, 1959) and false recognition of the nonpresented associate. This outcome demonstrates a strong memory illusion (Roediger, 1996) about which there has been a recent spate of research (see Roediger, McDermott, & Robinson, 1998, for a review).

A key question, but one that is difficult to answer conclusively, concerns the source of the associatively induced false memories: Do the nonpresented associates consciously come to mind during list presentation, with subjects later facing the difficult decision of determining whether the word was originally perceived or imagined? Or are the nonpresented associates highly primed during the study phase, in the absence of any conscious awareness of the word? In the former case, false memories would be said to occur from a failure of reality monitoring (see Johnson & Raye, 1981); subjects have difficulty discriminating between list words and the word that was generated but never presented. In the latter case, critical words would be primed due to spreading activation in a semantic network (e.g., Collins & Loftus, 1975), and this unconscious priming would be sufficient to induce later false recall and false recognition of the words.

Several lines of evidence suggest that the nonpresented associate[1] often comes consciously to mind during the study phase. For example, Roediger and McDermott's (1995, Experiment 2) finding that these words are predominantly classified as being *remembered* in Tulving's (1985) remember/know paradigm

The authors thank Kate Pfeifer for testing subjects in Experiments 2 and 3 and helping to code the data for all three experiments.

Address correspondence and reprint requests to Kathleen McDermott, Division of Radiological Sciences, Box 8225, Washington University School of Medicine, 4525 Scott Ave., St. Louis, MO 63110, or to Henry L. Roediger, III, Department of Psychology, Box 1125, Washington University, One Brookings Drive, St. Louis, MO 63130-4899. The authors can also be reached via internet at kmcd@npg.wustl.edu or roediger@artsci.wustl.edu.

[1] The nonpresented associate will also be referred to as the *critical word* or the *critical linking associate* because all the studied words are highly associated to this word; see Roediger and McDermott (1995) for a more detailed discussion of how the lists were constructed.

0749-596X/98 $25.00

supports this position. That is, when subjects are asked for each item judged old (i.e., previously studied) on a recognition test to report whether they can recollect some specific aspect of the presentation episode (a *remember* experience) or whether they just know the item was presented but cannot recollect anything specific about the original presentation (a *know* experience), subjects often claim that they can indeed recollect the original presentation of the critical nonpresented associates (see too Norman & Schacter, 1997; Read, 1996; Payne, Elie, Blackwell, & Neuschatz, 1996; Schacter, Verfaellie, & Pradere, 1996). This finding is probably most easily explained by assuming that the critical associates are consciously evoked during the study episode and that subjects later misattribute memories for the internally generated events as being memories for presented words. Qualitative studies of these remember judgments have shown that people also claim to remember rehearsing the critical nonpresented item to the same extent that they claim to remember rehearsing items actually presented (Mather, Henkel, & Johnson, 1997).

McDermott (1997) has argued that if the critical words come consciously to mind during the study episode, one would expect to see priming for these words on perceptual implicit memory tests; however, if implicit activation is producing the false memories, there should be no perceptual priming. McDermott (1997) showed that perceptual priming does indeed occur for the critical nonpresented items, although to a lesser extent than for words actually presented. This finding, like the results with respect to remember judgments, points to the explanation that false memories in this paradigm occur largely as a result of a breakdown in the reality monitoring process, with subjects confusing perceived and imagined (or internally generated) events (Johnson & Raye, 1981).

Because a breakdown in reality monitoring is probably an important source of the false memories in this paradigm, in the current experiments we decided to turn the task into a direct reality monitoring judgment to assess the extent to which subjects can distinguish between cases in which the critical word is and is not presented. Therefore, we informed subjects about the false recall and recognition phenomena that occur with these associative lists and gave subjects a sample list to demonstrate the phenomenon; we then presented the critical linking word in half of the lists but left it out of the other half of the lists. Subjects were informed that sometimes the critical linking word would be present and other times it would not, and their job was to determine whether it was presented and to remember this observation for a memory test.

Three experiments are reported in the present paper. In Experiment 1, all subjects were warned about the nature of the false memory phenomenon; they heard 12 associative lists, followed by a final recognition test. The critical item appeared in half of the lists but not in the other half. We wanted to explore the extent to which, on a final yes/no recognition test, subjects would be able to determine whether the critical item had been presented. This question can be asked separately for cases in which the item is and is not present. That is, to what extent can subjects accurately recognize the presence of the critical item when it is present? And to what extent can subjects determine the absence of the critical item when it is not present? Because subjects were not very accurate in their judgments with respect to the critical nonpresented associates in Experiment 1, the recognition test in Experiment 2 was changed such that subjects received a one-item recognition test immediately following each list. The item tested was always the critical item, which had been presented in only half of the lists. Here, too, the question was whether subjects could discriminate between instances in which the critical word was and was not presented. Again, the nature of the task was explained fully to subjects: They were made aware of the tendency of associates to elicit false recognition of nonpresented related associates, and their job was to determine for each list whether the linking associate had appeared in the list. Because the levels of false recognition remained high in this experiment, too, in Experiment 3 we warned only half of our subjects (and gave standard instructions to the other half) to determine

whether the explicit warning was at all effective in reducing the false recognition probabilities on an immediate test.

EXPERIMENT 1

The primary goal of Experiment 1 was to assess the ability of subjects to encode accurately and to remember whether the critical linking words had been presented in the lists when directly asked to do so; subjects were warned about the false recognition phenomenon, given an example, and told to monitor for the presence/absence of the critical words.

Method

Subjects. Twenty-four students from Rice University participated in this experiment in partial fulfillment of a course requirement.

Materials and design. Eighteen of Roediger and McDermott's (1995) lists, which are reported in their Appendix, were selected as stimuli for this experiment. (We chose the 18 lists that had produced the highest probabilities of false recall in a previous experiment.) Each subject studied 12 of the associative lists. In 6 of these studied lists, the critical linking item was not presented. This is the condition commonly used for producing false recall and false recognition and will be referred to as the Nonpresented condition in this report. (Note that the list itself is presented in this condition; however, the critical associate is not presented.) In 6 of the lists, the critical associate was presented; it replaced the word in serial position 4 for 3 of the lists and the word in position 10 for 3 of the lists. This condition will be referred to as the Presented condition. Therefore, of the 12 studied lists, 6 lists occurred in the Presented condition and 6 lists in the Nonpresented condition. Six lists were held out as Nonstudied and served as a base-rate measure of false alarms on the recognition test.

Counterbalancing was accomplished by dividing the lists into 3 sets (of 6 lists/set) and rotating the sets across the different conditions, such that each set of lists served in each study condition (Presented, Nonpresented, Nonstudied) an equal number of times across subjects.

Lists were recorded digitally in stereo in a female voice using a 22,050 Hz sampling rate and 16-bit resolution. Lists were presented in immediate succession via computer speakers. The words were presented at approximately a 1 s rate (mean list length = 15.34 s, SD = .45 s). For lists in which the critical word was presented, the appropriate list word (item 4 or 10) was spliced out (using Creative WaveStudio software, v.2.0, Creative Technology, Ltd.) and replaced with a recording of the critical linking associate. Therefore, with the exception of this one word, lists were exactly the same across all conditions.

Response sheets for the memory test contained 18 test words, each of which was a critical linking word. For each subject, 6 of these words had occurred in the Presented condition, 6 were from the Nonpresented condition, and the relevant list had not been studied for 6 of the words (the Nonstudied condition). Therefore, only 33% of the items on the test had actually been presented during the study phase. There was one random ordering of critical words on the test sheet.

Procedure. Subjects were tested individually or in pairs. The study instructions were read to subjects; because instructions were a critical aspect of the experiment, they are presented here verbatim:

> In this experiment, you will hear lists of words presented on the computer speakers. Your job is to try to remember the words you hear. Your memory for each list will be tested after all lists have been presented. You will hear 12 lists in all, with 15 words per list. You should pay close attention to the lists because the words will be presented quickly—about 1 word per second.
>
> Each list is comprised of related words. All the words in each list are associated to one common word. Sometimes the common word is present in the list, and other times it is not. For each list, you should try to figure out what the word is that ties all the other words together and note whether or not that word is present in the list.
>
> For example, you might hear the following words: *queen, England, crown, prince, George, dictator, palace, throne, chess, rule, subject, monarch, royal, leader, reign.*
>
> In this case, *king* is the word that links all the above words. You would be asked whether or not *king* was actually presented in the list. In this case, *king* was *not* present in the list. As mentioned above, sometimes

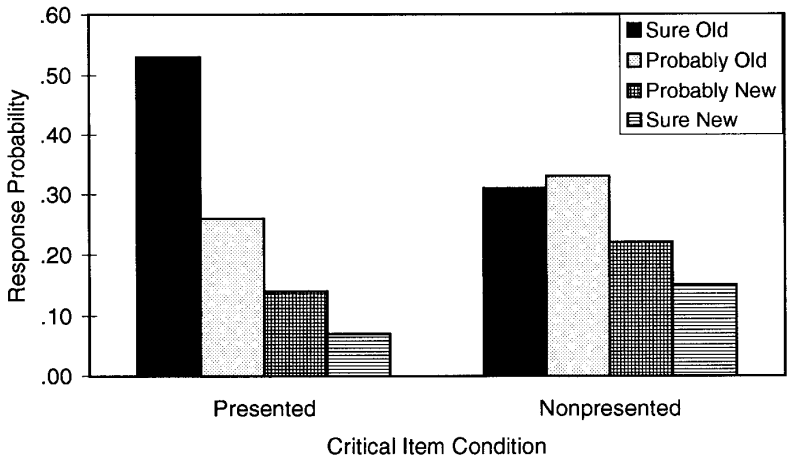

FIG. 1. The probabilities of responding Sure Old, Probably Old, Probably New, and Sure New to the critical linking associates as a function of whether the critical associate had occurred in the list (the Presented condition), or the critical associate had not been presented but the relevant list had been studied (the Nonpresented condition). Data reflect performance on a final recognition test, Experiment 1.

> the common word, like *king,* will be present and other times it will not. Sometimes people mistakenly remember the critical word that links together all the others (e.g., *king*) even when it was not presented. Do not make this error. *Be sure to notice and remember whether or not the linking word is present.*

Subjects were encouraged to ask questions about the procedure. Once the experimenter had determined that subjects understood the study instructions, list presentation was begun.

Presentation of the 12 lists took approximately 3 minutes. Following presentation of the last list, subjects were given the test sheets; instructions for the test appeared at the top of the sheets, and were as follows:

> For each of the words below, please (A) circle Y (for Yes) or N (for No) to indicate whether you heard a list of words that was related to that concept (irrespective of whether or not the word was actually present in the list) and (B) if you did hear a list of words related to that concept, indicate whether or not the word itself was present in the list. Circle 4 if you're SURE it WAS in the list, 3 if you think it PROBABLY WAS in the list, 2 if it PROBABLY was NOT in the list, and 1 if you're SURE it was NOT. If you have any questions, please ask the experimenter at this time.

After questions were answered, the experimenter summarized the test instructions and directed subjects to begin. The test was subject-paced.

Results and Discussion

Subjects were quite accurate on the initial question ("Did you hear words related to the concept [*sleep*]?"); the hit rate was .89 and the correct rejection rate .85. We turn now to the results of primary interest: the follow-up question regarding whether the linking word had appeared in the lists that subjects recognized as having been studied. Thus, the following results are conditional upon the subjects correctly recognizing having heard the list.

The results for this question are represented in Fig. 1. The left half of the graph shows the results for the lists in which the critical word had appeared, the Presented condition. Subjects correctly recognized having heard the word with a probability of .79 (and 53% of the time they claimed to be sure they had heard the word). Only 7% of the time did subjects (erroneously) claim that they were sure this word had not been presented in the list.

The accuracy in the Presented condition can be contrasted with the poor performance in the Nonpresented condition (shown on the right of Fig. 1). If subjects had shown a similar level of accuracy in the two conditions, the pattern of responses for the Nonpresented condition would have been the exact opposite of that for

the Presented condition; that is, the predominant response would have been Sure New, with successively fewer Probably New, Probably Old, and Sure Old responses. Obviously, however, this pattern did not occur. Subjects gave predominantly Old responses (probability .64), and 31% of the time they claimed to be sure the nonpresented critical item had been present in the list (Sure Old).

Mean ratings for the Critical Presented and Nonpresented words were calculated. For the Critical Presented words, the mean rating was 3.25, which indicates confidence somewhat higher than Probably Old (3). The mean rating for the Critical Nonpresented words was lower (2.78), which indicates confidence just below Probably Old (3). A paired comparisons *t* test indicated that the difference between mean rating scores was reliable, $t(23) = 3.90$, $SEM = .121$. Subjects in the Presented condition were fairly accurate at labeling the words as having been presented. Perfect retention (with high confidence) in this condition would have been a mean rating of 4; the obtained mean rating (of 3.25) was not too far off this mark. Performance in the Nonpresented condition was much poorer, however. Perfect retention (with high confidence) in this condition would have led to a mean rating of 1; the obtained mean of 2.78 was not close; indeed, it was closer to Probably Old than to Probably New.

In summary, when the critical word had not occurred in the study list, subjects were surprisingly poor at determining so: Over half of the Critical Nonpresented words were erroneously classified as having been presented, even though subjects had been fully briefed about the false recognition phenomenon before list presentation. This finding could indicate that subjects were poor at discriminating whether or not the items had been presented in the lists; however, it is also possible that the delay between hearing the lists and taking the test played a role in these high rates of false alarms. That is, immediately after presentation of each individual list, subjects may have been able to discriminate between presented and nonpresented critical items; over the course of hearing multiple lists and then receiving test instructions, they may have forgotten this information. This possibility is supported by the finding that reality monitoring becomes increasingly less accurate as a function of retention interval (e.g., Hamersley & Read, 1986) (but note that in Experiment 1 the delay between study and test was only about 5–7 minutes). In addition, there have been reports showing that the probability of false recall and recognition in this paradigm is very robust after a delay—sometimes even rising over intervals in which forgetting of studied words is observed (McDermott, 1996). In Experiment 2, we changed the test to an immediate 1-item recognition test following each list, in an effort to eliminate the possible effects of delay and output interference.

EXPERIMENT 2

The basic question in Experiment 2 was the same as in Experiment 1: To what extent can forewarned subjects discriminate between Critical Presented and Nonpresented words in this paradigm? The test in Experiment 2 was made even easier than Experiment 1 by placing it immediately after each 15-item list.

Method

Subjects. Sixteen subjects from Washington University were tested individually. They were each paid $5 for participation.

Materials and design. Twenty associative lists were selected for this experiment. These lists consisted of 16 of Roediger and McDermott's lists plus 4 others, which were taken from our expanded set of materials, reported by Stadler, Roediger, and McDermott (in press). Lists were recorded digitally and played over computer speakers, as described in the *Method* section of Experiment 1. The lists were divided into 2 sets. For each subject, the critical linking item appeared in half of the lists (1 set). Because the recognition test immediately followed presentation of each list, there was no reason to hold out some lists as nonstudied in this experiment. Each list occurred in the Presented and Nonpresented conditions an equal number of times across subjects.

A test sheet was constructed on which there were 20 numbered rows of the following re-

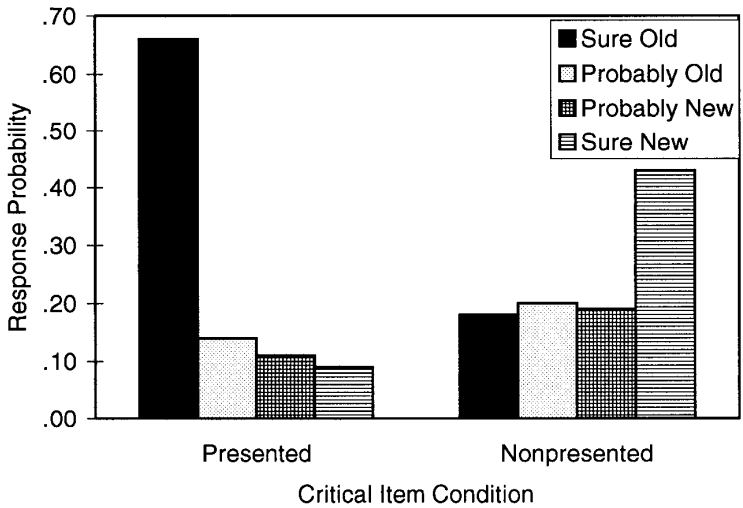

FIG. 2. The probabilities of responding Sure Old, Probably Old, Probably New, and Sure New to the critical linking associates as a function of whether the critical associate had occurred in the list (the Presented condition), or the critical associate had not been presented but the relevant list had been studied (the Nonpresented condition). Data reflect performance on 1-item recognition tests given immediately after presentation of each list in Experiment 2.

sponse possibilities: Sure Old, Probably Old, Probably New, Sure New.

Procedure. Subjects were given instructions much like those used in Experiment 1. The instructions were modified slightly, as necessitated by the procedural changes; for example, there were 20—not 12—lists in this experiment, and subjects were told that tests would follow the presentation of each individual list. That is, the critical linking word would be read by the experimenter, and subjects should determine the appropriate response (Sure Old, etc.) to circle on the response sheet. The experimenter explained that Old items referred to ones that had been presented in the list, and New items were those that had not been presented in the list.

Lists were then played over the speakers. After each list, the experimenter read aloud the appropriate critical associate. When the subject had responded by circling the appropriate alternative on the test sheet, the experimenter initiated presentation of the next list. The entire experiment took about 15 minutes.

Results and Discussion

Figure 2 shows the results for Experiment 2. The pattern of results for the Presented words (on the left side of the figure) looks similar to the pattern for those words in Experiment 1: On most trials (.80) subjects correctly classified the words as having been presented, and the response with the greatest probability was Sure Old (.66). Subjects rarely claimed that they were sure the item had not been present in the list (.09). The mean rating for this condition was 3.37.

For the lists in which the linking word had not been presented (shown in the right half of Fig. 2), subjects classified the critical words as New with a probability of .62, collapsing across level of confidence. The modal response was Sure New, which occurred with a probability of .43. Nevertheless, the false alarm rate for the critical words that had not been presented was still substantial (.38); a Sure Old classification occurred with a probability of .18.[2] The mean rating for this condition was 2.12, which differs

[2] Note that we are assuming that the false alarm rate for unrelated words would approximate .00. That is, we assume that subjects are not predisposed to label items old for reasons other than memorial-based ones. As the results of Experiment 3 will show, this seems to be a valid assumption. Thus, the high false alarm rates (.38 for Experiment 2 and .64 in Experiment 1) do seem to be tapping illusory memories.

reliably from that of the Presented condition (3.37), $t(15) = 7.34$, $SEM = .17$.

Unlike the results of Experiment 1, the pattern of responses in this condition clearly shows a reversal in responses for critical Nonpresented words (relative to Critical Presented words). That is, most Critical Presented words were (correctly) classified in the Sure Old category, and most Critical Nonpresented words were (correctly) classified in the Sure New category. However, the finding that on over $\frac{1}{3}$ of the trials subjects produced a false alarm to the Critical Nonpresented words is somewhat surprising. After a full explanation of the phenomenon, and with a 1-item immediate recognition test following each 15-word list, subjects still claimed to have heard the Critical Nonpresented items on 38% of the trials. This finding led us to wonder whether our warning instructions were exerting any attenuating effect on the false alarm rates. Perhaps the memory illusion in this paradigm is simply not under conscious control. To assess this possibility, we performed Experiment 3.

EXPERIMENT 3

To address whether warning instructions are effective at dampening the memory illusion in this paradigm, we warned only half of the subjects in Experiment 3, and we provided standard, general study and test instructions to the other half of the subjects. Gallo, Roberts, and Seamon (1997) reported a similar experiment (but with the critical word absent from all the lists); they obtained an attenuation (but not an elimination) of false recognition following a warning.

In the present experiment, it seemed likely that subjects in the No Warning condition might be confused initially and then catch on to the phenomenon if we adopted the procedure of Experiment 2 (a 1-item immediate recognition test following each list); therefore, we expanded the number of items tested after each list to 4. The 4 items tested were: the critical word (presented in half of the lists and not presented in half); a standard studied word; a very low associate of the critical word, which was not present in the list; and a word generally unrelated to any of the studied words.

Method

Subjects. Thirty-two Washington University undergraduates participated in the experiment. One to three subjects were tested in each session.

Materials and design. The study materials were identical to those used in Experiment 2. Instructional set (Warning/No Warning) was manipulated between subjects (16 subjects per condition). A test sheet was constructed with 80 numbered rows; each row contained the following response possibilities: Sure Old, Probably Old, Probably New, Sure New. For the recognition test, the Standard Studied items (1 per list) were drawn from serial positions 2, 3, 5–9, 11, and 12 in the lists (with 1–3 words drawn from each position). Low Associates were obtained by consulting Russell and Jenkins's (1954) word norms for words associated to the critical linking word but not contained in our associative lists. Unrelated words were simply words that seemed unrelated to the studied words. Materials for this experiment can be found in the Appendix.

Procedure. The procedure was virtually identical to that of Experiment 2, with the exception that 4 words were tested after each list. Subjects in the Warning condition received instructions highly similar to those for Experiment 1 (but, as in Experiment 2, modified as necessitated by procedural differences). The subjects in the No Warning condition were given nonspecific instructions concerning the nature of the lists: They were told that they were participating in a memory experiment, that they would hear 20 15-word lists, and that each list would be followed by a short, 4-item recognition test. No mention was made of the associative nature of the lists or that there were linking words that might or might not be presented.

After presentation of each list, the experimenter read aloud the appropriate row number and the test word, waited for subjects to respond by circling an option on the test sheet (Sure New, etc.), and then read the next row number and test item until all 4 words had been tested.

TABLE 1

Probability of Assigning Items Sure Old, Probably Old, Probably New, and Sure New Responses as a Function of Item Type and Instructional Condition in Experiment 3

		Response category			
Item type	Instruction	Sure Old	Probably Old	Probably New	Sure New
Standard Studied	No warning	.62	.12	.10	.17
	Warning	.47	.21	.16	.17
Critical Pres.	No warning	.79	.13	.03	.06
	Warning	.73	.19	.05	.03
Critical Nonpres.	No warning	.57	.23	.06	.14
	Warning	.38	.21	.15	.26
Low Associate	No warning	.04	.06	.13	.77
	Warning	.02	.05	.19	.74
Unrelated	No warning	.00	.00	.01	.99
	Warning	.00	.00	.03	.97

Test words occurred in one fixed random order for all subjects, and the ordering of item type (Critical Presented, Critical Nonpresented, Standard Studied, and Low Associate) differed across lists.

Results and Discussion

Results of Experiment 3 are presented in Table 1; a subset of the data are also depicted graphically in Fig. 3, in which hits and false alarms (collapsed across degree of confidence) are displayed. Data in the last two rows of the table show that subjects were highly adept at identifying Unrelated words as such: These words were never misclassified as Old and were almost always given a rating of Sure New. Likewise, subjects were quite good at labeling the Low Associates as New items; collapsing across level of confidence (see Fig. 3), they were assigned New responses with a probability of .90 in the No Warning condition and .93 in the Warning condition; this difference was not reliable, $t(30) = 1.10$, $SEM = .028$. The Sure New responses were also used with a high prob-

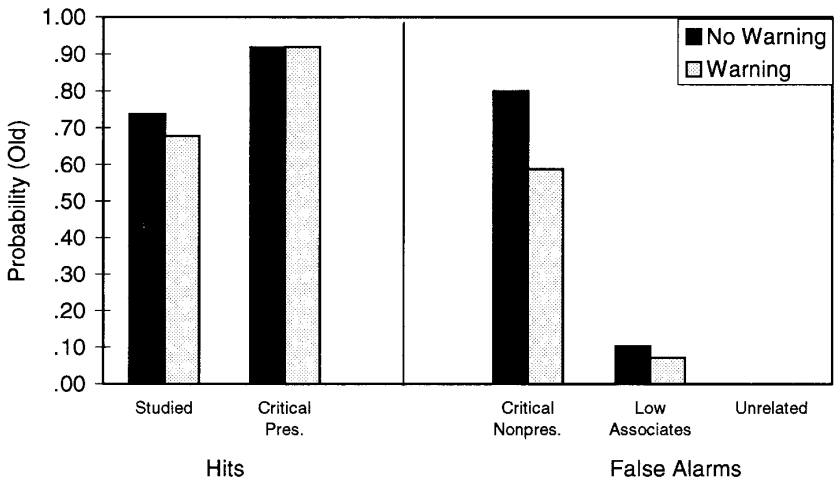

FIG. 3. The probabilities of labeling a word as Old (collapsing across level of confidence) as a function of item type and whether the subject had been forewarned about the false memory effect.

ability for these items (.77 and .74 for the No Warning and Warning conditions, respectively, $t < 1$.) Therefore, as we proposed in footnote 2, subjects were extremely capable of identifying standard lures as being new items; this finding supports the position that the high false alarm rates to the Critical Nonpresented words demonstrate a compelling memory illusion and are not simply reflecting a general bias toward positive responses.

In summary, the results from the Low Associate and Unrelated lure conditions show that (a) subjects do not simply classify most words as being Old in this paradigm and (b) the warning instruction did not affect performance for the nonstudied Low Associates or the Unrelated words (although ceiling effects cloud interpretation of this finding, especially in the Unrelated condition).

Results for the Standard Studied, Critical Presented, and Critical Nonpresented conditions will now be considered together. These data can be seen in the top of Table 1 and in Fig. 3. We discuss first the Old responses, collapsing across level of confidence, as shown in Fig. 3. An inspection of these data suggests that the warning instruction did affect the response criterion for the Standard Studied words (.74 and .68 for the No Warning and Warning conditions, respectively, $t(30) = 1.45$, $SEM = .04$, $p = .07$) and for the Critical Nonpresented items (.80 and .59 for the No Warning and Warning conditions, respectively, $t(30) = 3.56$, $SEM = .060$). However, there is no discernible difference between the two conditions for the Critical Presented words, although ceiling problems cloud this finding; the probability of an Old response was .92 in both conditions, $t < 1$. We performed a median split on the basis of hit rate for the Standard Studied words to see if performance in the Critical Presented condition would come down off ceiling for the subjects with lower hit rates to the Standard Studied words; it did not. Low performers showed very high hit rates to the Critical Presented words (.93 and .90 in the No Warning and Warning conditions, respectively). High performers, too, were at ceiling here (.91 and .94 in the No Warning and Warning conditions, respectively). More will be said about this finding in the General Discussion.

In general, these data show that subjects were quite adept at determining when the Critical word had been presented (as shown by the high hit rates); somewhat surprisingly, subjects were no less adept at this task after receiving a warning with respect to the nature of the lists, although we interpret this result cautiously. Most importantly, the purpose of the experiment was to determine whether a warning would affect false recognition rates; results show conclusively that the probability of false recognition of the Critical Nonpresented words was attenuated by the warning instructions.

We turn now to the Sure Old results, which appear in the leftmost column of Table 1. The probability of a Sure Old classification for the Standard Studied items differed as a function of instructional set (.62 for the No Warning condition and .47 for the Warning condition, $t(30) = 2.75$, $SEM = .054$). The Critical Presented items demonstrated no reliable effect of instructions (.79 for the No Warning condition and .73 for the Warning condition, $t(30) < 1$), although numerically the difference was in the predicted direction. Therefore, accurate recognition did not change as a function of instruction for the Critical Presented words when confidence was ignored (as discussed previously); when only the Sure Old responses were examined, there was a small but nonsignificant effect. In general, the Critical Presented words were very well recognized (and with similarly high levels of confidence across the instructional sets). We will return to this point in the General Discussion.

The Critical Nonpresented items, like the Standard Studied items, did show a large instructional effect in the Sure Old responses (.57 for No Warning and .38 for Warning, $t(30) = 2.59$, $SEM = .07$). Therefore, the Critical Nonpresented words and the Standard Studied words behaved similarly as a function of instructional set: Fully informed subjects produced fewer false alarms to the critical associates but at the expense of lower hit rates.

A secondary question with respect to these data is whether the effect of the warning instruction on the Critical Nonpresented items simply represents a general shift in criterion or if there is some benefit of warning for these items over and above the general criterion shift accompanying the warning instruction. Because interpretation of data for the Critical Presented words is clouded by a ceiling effect, we will use the Standard Studied words for comparison. Collapsing across level of confidence, there is a probability change of .06 for the hit rate of Standard Studied items as a function of instructional set. The corresponding change in the false alarm rate for the Critical Nonpresented words is .21. This interaction is reliable, $t(30) = 2.06$, $SEM = .07$. Therefore, warning instructions did serve to substantially decrease the level of false recognition in this experiment, and this effect seems not to be a general shift in criterion because it was greater for the Critical Nonpresented words than for Standard Studied words (see too Gallo et al., 1997).

Sure New responses are also informative with respect to this question (see Table 1). The instructional manipulation exerted no influence on Sure New responses for the Standard Studied items (.17 in both conditions, $t < 1$) or the Critical Presented items (.06 and .03 in the No Warning and Warning conditions, respectively, $t(15) = 1.43$, $SEM = .021$). The only condition for which instructions affected Sure New responses was the Critical Nonpresented condition, in which the probability of a (correct) Sure New response increased with the warning (.14 and .26 for the No Warning and Warning conditions, respectively, $t(15) = 2.03$, $SEM = .059$). Further, the interaction of instruction with the presented/nonpresented variable was reliable, $t(15) = 2.33$, $SEM = .067$. These results, when combined with those discussed above, converge upon the conclusion that false recognition of the Critical Nonpresented items is under some amount of conscious control: False recognition shows an effect of warning instruction over and above a general shift in criterion.

One puzzling finding in this experiment concerns why the general pattern of results for the Critical Nonpresented items in Experiment 3 seems more like that of Experiment 1 (with a delayed test) than of Experiment 2 (with an immediate test, as in Experiment 3). That is, when the recognition test occurred immediately after list presentation in Experiment 2, the modal response to the Critical Nonpresented words was a (correct) Sure New response (see Fig. 2). In the current experiment, the test also occurred immediately after each list, and yet the modal response for Critical Nonpresented words was Sure Old. The most likely source of this discrepancy lies in the number of items tested: In Experiment 2 we used a 1-item recognition test, whereas in Experiment 3 we tested 4 items. (As noted above, this change was deemed necessary to ensure that the No Warning subjects did not immediately become aware of the nature of the experiment.) Although this issue cannot be conclusively resolved with the current data set, it seems plausible that the delay and interference brought on by the 4-item test might be the source of this discrepancy, making Experiment 3 more closely analogous to Experiment 1, in which the test was a final recognition test.

In summary, the primary finding of Experiment 3 was that warning subjects did indeed attenuate the false recognition rate of the Critical Nonpresented items. However, the effect of warning was surprisingly slight; the modal response for these items was an incorrect judgment of Sure Old. Nevertheless, although only a slight effect, the effect of warning subjects represented more than just a general shift in criterion.

GENERAL DISCUSSION

The primary question addressed by these experiments is whether subjects are able to distinguish between cases in which a critical associate is and is not presented. Two experiments showed that, even under optimal conditions for making this judgment (i.e., when fully informed about the false recognition phenomenon), subjects had substantial difficulty in performing this task.

Thus, the principal findings of the experiments reported here are as follows. When subjects were explicitly forewarned about the nature of the associatively induced false recognition effect, they still falsely recognized 64% of the critical nonpresented words as having been presented on a final recognition test (Experiment 1). When the delay between list presentation and the recognition test was minimized and subjects were tested on only one item, they were better at labeling the critical nonpresented items as lures: The modal response in this experiment was for subjects to claim correctly that they were sure the item had not been present in the list. Yet even under these conditions, the critical nonpresented words were called Old with a probability of .38 (Experiment 2). Therefore, Experiment 3 examined whether an explicit warning effects lower levels of false recognition in the immediate testing paradigm. The primary finding of this experiment was that warning subjects did diminish the false recognition effect. Further, this reduction was slightly greater than the general shift in response criterion. Subjects falsely recognized the Critical Nonpresented words with a high probability in both conditions, leading us to conclude that although the illusory memory effect in the Roediger–McDermott paradigm is extremely robust, it is somewhat amenable to conscious control.

In an experiment somewhat similar to Experiment 3 of this series, Gallo et al. (1997) showed that a warning to subjects was sufficient to reduce (but not eliminate) the false recognition effect on a final recognition test given after presentation of 8 associative lists. However, in Gallo et al.'s experiment, the critical linking word was never presented in any of the lists; therefore, if subjects were able to confidently identify the English word that most effectively linked all the presented words, they could be confident that the word had not been presented. That is, the effect of the warning manipulation observed by Gallo et al. might have been overestimated by the fact that subjects did not necessarily have to perceive or remember whether or not the critical item had been presented in any of the lists; rather, it was sufficient to identify the linking word, and that it had not been presented could then be inferred. The results reported in Experiment 3 in the present report confirm and extend Gallo et al.'s conclusion under conditions in which this criticism cannot be made: Warning subjects reduced (but by no means eliminated) false recognition of the critical nonpresented items under conditions in which subjects had to actively determine whether or not the linking word was presented in the list.

We conclude by making several general observations about the data reported here. First, it is interesting that when the critical linking items are included in the study lists, subjects are extremely effective at recognizing these items as having been presented. In Experiment 3, subjects were better at classifying the Critical Presented words as being Old (in the .90 probability range) than they were at classifying the Standard Studied items as being Old, which they did in the .70 probability range. This finding of extremely high levels of accurate recognition of the Critical Presented words confirms McDermott's (1997, Experiment 1) similar observations in free recall. McDermott reported that the Critical Presented items are recalled with an unusually high probability (.42, relative to .16 for other studied words occurring in similar serial positions). Therefore, as described in McDermott's (1997) report, although the Critical Nonpresented items are recalled and recognized with about the same probability as Standard Studied items, they are not recalled and recognized as if they had been present in the list, as might be inferred from Roediger and McDermott's (1995) data and many of the subsequent studies using this paradigm. When the critical words are presented in the list, they are unusually well remembered.

Other data lead us to believe that the hit rate for the Critical Presented words will tend to hover around the ceiling (Robinson, 1998). For example, when the studied words were presented for only 20 ms, the hit rate for the Critical Presented words was .88, whereas

the corresponding hit rates for other list words did not approach ceiling (.70), nor did the critical false alarm rate (.75). Thus, just as there is something special about the critical words when they are not presented (such that they are often recognized as having been presented), there is also something special about them when they are presented (i.e., they are virtually always recognized).

The finding that the critical false alarm rate is, to a small extent, amenable to conscious control is an interesting one and is consistent with the existing literature on false alarms using this paradigm. For example, older adults are assumed to have deficits in consciously controlled components of recognition memory and source discrimination. Thus, to the extent that these processes are useful in counteracting the critical false alarm rate, we might expect older adults to show an elevated false alarm rate to these items. This pattern is indeed observed (Norman & Schacter, 1997).

In addition, the source monitoring framework would lead us to predict that people should be able to distinguish between studied and critical nonpresented words on the basis of amount of perceptual information retrieved with the item. When words previously heard or seen are retrieved, more perceptual details are available than when people retrieve words previously generated (Johnson & Raye, 1981). Roediger and McDermott's (1995) comparable levels of accurate and false memories as measured by recall, recognition, and remember judgments indicate that subjects often do not take advantage of this information when encountering memory tests in this paradigm. However, Mather et al. (1997) and Norman and Schacter (1997) have provided evidence that when subjects are specifically asked about the degree of perceptual detail accompanying studied and critical nonpresented words, more such details accompany studied words. Schacter, Reiman, Curran, Yun, Bandy, McDermott, & Roediger (1996) have reported positron emission tomography data that converge upon this conclusion; specifically, they reported greater temporoparietal activation (in or near auditory cortex) while subjects performed a yes/no recognition judgment on previously heard words relative to performing the same judgment on the critical nonpresented words.

The present experiments complement these findings in a case in which subjects are not directly asked about the degree of perceptual detail. Rather, they are informed of the phenomenon and are asked to minimize false alarms. The question is whether people can apply the information that sometimes distinguishes true and false memories to guide them in determining whether or not the word was studied. The source monitoring framework would predict that subjects should be able to do so. In addition, Mather et al. (1997) have argued that changes in criterion should disproportionately affect false (relative to accurate) recognition in this paradigm. Thus when subjects are instructed to adopt a high criterion, they should be better able to reject false memories. Our experiments show that this is true, but to a limited extent.

In conclusion, it is noteworthy that, as in many previous studies (e.g., McDermott, 1996; Norman & Schacter, 1997; Payne et al., 1996; Roediger & McDermott, 1995), the level of endorsements of Critical Nonpresented words resembled that of presented words more than it resembled other lures (e.g., the Low Associates and Unrelated Words in Experiment 3). This finding was surprising when first reported; however, it is considerably more surprising under the conditions reported here, in which subjects are specifically asked not to make this mistake. Although the memory illusion is, to a limited extent, amenable to conscious control, there seems to be a component to these illusions that is almost perceptual in nature: Subjects have only limited success in bringing the illusion under conscious control, and they appear unable to eliminate it. Further work is clearly needed to disentangle the consciously controlled and automatic components of these illusory memories.

APPENDIX

Materials for Experiment 3: Critical Linking Items, Studied Words, Low Associates (Not Contained in the Study Lists), and Unrelated Words are Presented for Each Studied List)

Critical	Studied	Low Associate	Unrelated
needle	point	yarn	religion
soft	furry	mattress	lick
fruit	ripe	lemon	sleeve
sleep	awake	lay	carpenter
trash	crowded	discard	southern
city	state	parks	painting
lion	mane	ferocious	beach
anger	temper	sorrow	anatomy
bread	flour	stale	shrill
chair	wood	lamp	cabbage
smoke	fire	thick	khaki
thief	money	jewels	tennis
foot	toe	sore	dishes
river	tide	bend	paper
mountain	glacier	plateau	opera
spider	fly	ant	rectangle
cold	chilly	cough	marriage
slow	delay	move	thirst
black	bottom	wool	scissors
rough	rugged	harsh	vitamin

Note. Complete lists may be obtained in Stadler et al. (in press).

REFERENCES

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

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

Gallo, D. A., Roberts, M. J., & Seamon, J. G. (1997). Remembering words not presented in lists: Can we avoid creating false memories? *Psychonomic Bulletin & Review,* **4**, 271–276.

Hammersley, R., & Read, J. D. (1986). What is integration? Remembering a story and remembering false implications about the story. *British Journal of Psychology,* **77,** 329–341.

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

Mather, M., Henkel, L. A., & Johnson, M. K. (1997). Evaluating characteristics of false memories: Remember/know judgments and memory Characteristics Questionnaire compared. *Memory & Cognition,* **25,** 826–837.

McDermott, K. B. (1997). Priming on perceptual implicit memory tests can be achieved through presentation of associates. *Psychonomic Bulletin & Review,* **4,** 582–586.

McDermott, K. B. (1996). The persistence of false memories in list recall. *Journal of Memory and Language,* **35,** 212–230.

Norman, K. A., & Schacter, D. L. (1997). False recognition in younger and older adults: Exploring the characteristics of illusory memories. *Memory & Cognition,* **25,** 838–848.

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.

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

Robinson, K. J. (1998). *Automatic mechanisms that produce false memories.* Unpublished masters thesis, Washington University.

Roediger, H. L. (1996). Memory illusions. *Journal of Memory and Language,* **35,** 76–100.

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., McDermott, K. B., & Robinson, K. J. (1998). The role of associative processes in creating false memories. In M. A. Conway, S. E. Gathercole, and C. Cornoldi (Eds.), *Theories of memory II* (pp. 187–245). Hove, Sussex: Psychological 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., 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.

Schacter, D. L., Reiman, E., Curran, T., Yun, L. S., Bandy, D., McDermott, K. B., & Roediger, H. L. (1996). Neuroanatomical correlates of veridical and illusory recognition memory: Evidence from positron emission tomography. *Neuron,* **17,** 267–274.

Stadler, M. A., Roediger, H. L., & McDermott, K. B. (in press). Norms for word lists that create false memories. *Memory & Cognition.*

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

(Received March 13, 1998)
(Revision received April 13, 1998)