Psychological Review
1999, Vol. 106, No. 2, 406–410

0033-295X/99/$3.00

False Alarms About False Memories

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

M. B. Miller and G. L. Wolford (1999) make two contributions. First, they add conditions to the basic Roediger–McDermott (1995) procedure and find that critical items are recalled and recognized more often if they are presented in the list than if they are not presented. These results agree with our own, which are briefly reviewed. Second, they apply signal detection theory to the paradigm and conclude that false responding is caused by a shift of response criterion. They present no evidence that requires this interpretation, however, and we discuss several reasons why their account is implausible. For example, even when fully informed subjects are asked to use a very strict criterion in responding, the false recognition phenomenon persists at high levels. Further, some variables produce opposite effects on veridical and false recall and recognition; this pattern is difficult to accommodate solely by a shift in response criterion. Therefore, little evidence exists that a simple shift of response criterion can explain false recall and false recognition phenomena in the Roediger–McDermott paradigm.

Roediger and McDermott (1995) presented subjects with lists of words (e.g., *bed, rest, awake,*) that were associates of a critical nonpresented word (e.g., *sleep*); subjects were given both immediate free recall tests (see also Deese, 1959) and a delayed recognition test. Despite instructions warning the subjects not to guess during the recall tests (i.e., to be certain that all words produced had been heard), subjects recalled the critical nonpresented word with about the same probability (Experiment 1) or with even higher probabilities (Experiment 2) than they recalled the words presented in the middle of the list. This general outcome has been widely replicated (see Roediger, McDermott, & Robinson, 1998, for a general review). In addition, on the recognition test, the critical nonpresented items were labeled "old" with probabilities similar to those items that had been studied; further, subjects often claimed to remember the specific instance of presentation of the critical nonpresented words. Roediger and McDermott (1995) argued that these phenomena constituted a striking memory illusion in which people recall, recognize, and claim to remember the moment of occurrence of events that did not occur.

Miller and Wolford (1999) make two primary points. First, they noted that the experiments reported in Roediger and McDermott's (1995) article did not include a condition in which the critical word was presented. Second, they applied a signal detection analysis and concluded that false alarms in our paradigm were largely due to criterion effects. We comment on these points in turn.

We thank Dave Balota, Jim Neely, Tram Neill, Richard Marsh, Endel Tulving, and John Wixted for helpful comments.

Correspondence concerning this article should be addressed to Henry L. Roediger III, Department of Psychology, Box 1125, Washington University, One Brookings Drive, St. Louis, Missouri 63130-4899, or to Kathleen B. McDermott, Division of Radiological Sciences, Box 8225, Washington University School of Medicine, 4525 Scott Avenue, St. Louis, Missouri 63110. Electronic mail may be sent to roediger@artsci.wustl.edu or to kmcd@npg.wustl.edu.

Recall and Recognition of the Critical Item

Roediger and McDermott (1995) did not include conditions in which the critical item was presented in the lists because we were interested in whether list context by itself could elicit false recall and false recognition of the critical item. The error rates we observed approximated correct recall and recognition of the list items that had been presented. Miller and Wolford (1999) tested the hypothesis that critical items would be remembered as well when they were not presented as when they were presented. Of course, this was their own hypothesis; we never predicted this pattern of results or claimed to have found it; after all, we did not have data relevant to that issue. However, McDermott (1997) did provide the relevant conditions, as we have in more recent experiments, also (McDermott & Roediger, 1998). Our findings agree with Miller and Wolford's (1999) outcome that the critical items are better remembered when presented than when not presented. Although our own inclusion of the critical items within the lists was done for different reasons than providing a signal detection analysis, we review the relevant findings here briefly.

McDermott (1997) either presented or did not present the critical item in the Roediger–McDermott paradigm, and her results from final free recall (given after presentation of 18 lists) are shown in Figure 1. When presented, the critical item replaced the fourth item in the list; as can be seen, she obtained much higher recall when the item had been presented than when it had not. As she pointed out, "although the critical nonpresented words behaved much like words that had been studied, it is not true that these critical words behaved as though they themselves had been studied" (pp. 585–586). An additional relevant finding from this experiment was that even though McDermott (1997) included the critical item in half her lists, the level of false recall approximated the level of veridical recall for most items presented in the list (i.e., those after the primacy effect), unlike the pattern predicted by Miller and Wolford (1999).

In general, it is hardly a surprise that studied items are recalled with higher probabilities than nonstudied items, but the power of this associatively induced memory illusion is so great that the

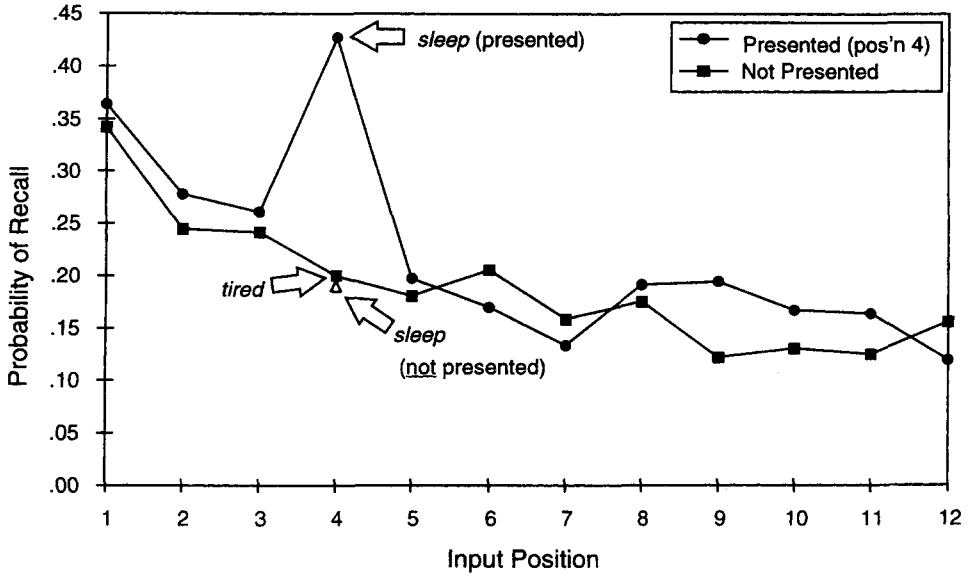

Figure 1. Performance on a final free recall test as a function of serial position and whether the critical associate (e.g., *sleep*) had been presented in the list (in place of the fourth associate) or not (data from McDermott's, 1997, Experiment 1). When the critical associate had appeared in the study list, it was extremely well remembered (recalled with a probability of .42). When it had not appeared in the study list, it was recalled at a level (.19) approximately equal to the probability of recalling the items that had been presented (e.g., *tired*), in the fourth serial position, recalled with a probability of .20.

comparison is of interest. McDermott and Roediger (1998) reported similar results with recognition, which are reviewed below. The finding that critical items are better remembered when presented than when not presented is quite secure, if not particularly novel.

Signal-Detection Analysis of False Recognition

Miller and Wolford (1999) report a signal detection analysis of the false recognition phenomenon; they conclude that a criterion shift is responsible for the memory illusion because their experiment showed relatively small differences in measures of sensitivity [$d(a)$] and large differences in the measure of bias [$c2$].[1] Although this evidence is consistent with the idea of a criterion shift, it is by no means definitive. The same findings could arise if the subjects' criteria stayed the same, but presentation of the lists produced a change in distributions. Miller and Wolford (1999) clearly recognize this point, which was made by a reviewer; Miller and Wolford note that "the differences that we observe could result from these shifted distributions rather than shifted criteria" (p. 403).

Wixted and Stretch (1998) made the point even more forcefully that the Miller–Wolford results are equally consistent with shifts in criteria or shifts in distributions. As Wixted and Stretch noted, when subjects change response criteria, measures of bias must change. However, when measures of bias change, it does not logically follow that subjects changed their response criterion; changes in distributions could lead to the same outcome (i.e., changes in the "bias" parameter). Therefore, the data reported by Miller and Wolford (1999) simply do not speak to the issue. Although none of the results in Miller and Wolford's article require the conclusion that false recognition is due to criterion shifts, other evidence reviewed next is more directly relevant to the issue.

Direct Manipulation of Response Criteria

A powerful way to gauge the importance of the response criterion is to manipulate criteria across groups of subjects by varying instructions. Miller and Wolford (1999) provide a thought experiment in which they imagine a group of subjects instructed to adopt a liberal response criterion. We do not doubt that if Miller and Wolford were to tell subjects to say "old" to every item, no matter what they think its status is, then Miller and Wolford would obtain many false alarms. Similarly, if subjects were instructed to say "new" to every item, then they would never report false memories; in fact, they would never report any memories at all. The same outcome would occur in within-subjects experiments if subjects were told to respond to, say, every other item in these same two ways (e.g., calling odd items *old* and even items *new*). However, demonstrating that subjects are obedient to instructions seems unrelated to the question of whether or not they rely on metaknowledge to voluntarily shift their response criterion from item to item.

The standard assumption within classical signal-detection theory (Green & Swets, 1966; Hirshman, 1995) is that subjects set a

[1] Actually, measures of sensitivity were greater for related (i.e., standard list) items than for critical or for unrelated items in both of Miller and Wolford's (1999) experiments. No inferential statistics were applied to support the claim of equivalent measures of sensitivity or, for that matter, to support the claimed differences in measures of bias.

single criterion when taking a test and they do not change it dramatically from item to item, as Miller and Wolford (1999) assume they do. Of course, minor perturbations from sequence effects (e.g., a long series of *yes* trials) can lead to slight changes. Hirshman and Arndt (1997) argued that some standard signal detection assumptions must be modified to apply the theory to the issue of false memories, although they did not argue for single-item criterion shift. Miller and Wolford argue that subjects analyze what category of item each test word represents and then set different criteria for the different categories, which leads to the three different measures of *c2* in their Table 2. We doubt that subjects are capable of making such sophisticated discriminations between critical items and lists of words, although Miller and Wolford (1999) argue that the subjects probably possess such metaknowledge about critical items. Even if subjects could distinguish critical items from other list items, other evidence indicates that they probably would not use this ability to adjust response criteria: Stretch and Wixted (1998) attempted to induce item-by-item criterion shifts in several experiments using obvious item characteristics and found that subjects adopted a single criterion and did not adjust responses on an item-by-item basis.

Rather than liberalize response criteria as in Miller and Wolford's (1999) thought experiment, we believe it is more informative to subject the criterion interpretation to an empirical test to see whether the phenomenon of false recognition survives with even stricter criteria than usually invoked on standard recognition tests. Roediger and McDermott's (1995) standard instructions, widely adopted by others, warned subjects against guessing on free recall tests and asked subjects to be as accurate as possible on recognition tests. Nonetheless, two sets of researchers have tried to make the response criterion even stricter by informing subjects about the phenomenon in question and asking them not to make errors.

Gallo, Roberts, and Seamon (1997) compared subjects tested in the Roediger–McDermott paradigm with two groups given more specialized instructions. In the most stringent condition, Gallo et al., (1997) informed some subjects about the false recognition phenomenon and strongly warned them not to make such errors. Relative to the standard condition used by Roediger and McDermott (1995), the forewarned subjects did reduce both their hit rates for studied items and false alarm rates for critical nonpresented items, revealing a general criterion shift. However, these subjects, who were warned not to make errors, still showed a critical false alarm rate of .46, which was much higher than the false alarm rate of unrelated items under similar conditions (.14).

McDermott and Roediger (1998) went further with the same general idea. In three experiments, the phenomena of false recall and false recognition in the Roediger–McDermott paradigm were explained to subjects, and an example was provided. Subjects were warned not to make these errors. In another condition, subjects were given our standard instructions, which include the general admonition to be as accurate as possible. We converted the paradigm to a reality-monitoring situation (Johnson & Raye, 1981) by having the critical item appear in a random half of the lists and not the other half, with the warned subjects instructed that the critical word would only sometimes be in the list.[2] Both groups of subjects were given a series of our lists and were tested on recognition (with confidence judgments) immediately after the list had been presented. The primary question was whether subjects could eliminate the false recognition effect when fully instructed and warned not to make errors; that is, would subjects be able to shift their response criteria in such a way to eliminate the false recognition effect?

The answer we obtained from our experiments (in agreement with Gallo et al., 1997) was that manipulating criterion could somewhat reduce but certainly not eliminate the false recognition effect. Figure 2 shows results from Experiment 3 for a standard (not warned) group and an informed (warned) group on recognition of regular list items when presented (left columns) and of the critical items when presented (middle columns) or not presented (right columns). Warning subjects had the expected effect of making them more conservative, as can be seen by reductions in the hit rate for studied items and false alarm rate for critical nonpresented items. In addition, the interaction between instruction (warning or no warning) and item type (studied and critical) apparent in the figure was significant: The warning decreased the false alarm rate for critical nonpresented items more than it decreased the hit rate for studied items. (We have adopted the "standard studied" condition as the control condition here rather than the "critical presented" condition, because the latter condition was compromised by a ceiling effect.) Therefore, manipulating response criterion does affect the false recognition phenomenon; however, even with stringent response criteria with fully informed subjects, the false recognition effect is robust: 57% of the critical lures were erroneously labeled "old," as shown in Figure 2. Further, 38% of the critical lures were assigned a high confidence judgment of having occurred in the list (i.e., "sure old"); the comparable rates for low associates of list items and for unrelated items were 2% and 0%, respectively. We should point out that the strict instructional conditions used by Gallo et al., (1997) and by McDermott and Roediger (1998) are quite unlike conditions typically experienced either inside or outside the laboratory. However, the illusion survives even this most stringent test.

Other results are inconsistent with the criterion shift interpretation of the false recall and false recognition phenomena. Any result that has opposite effects on correct responding and false responding is difficult for an account based solely on shifting criteria. McDermott (1996) found that repeated study–test trials increased correct recall and reduced false recall. Norman and Schacter (1997) compared performance of healthy older adults with that of younger adults. Older adults showed poorer free recall of studied items than did younger adults, as expected; older adults also showed enhanced recall of critical nonpresented items relative to younger adults, an effect replicated by Balota et al., (1998). Such a decrease in one measure and an increase in the other is difficult to explain in terms of simple shifts in response criteria. Indeed, McDermott (1996) pointed out that a major theoretical challenge is to explain why some variables affect veridical and false remembering in the same way and other variables affect them in opposite ways. Shifts in response criteria may play some role when both measures are affected in similar ways, but the account is less likely to explain opposite effects of the same variable.

[2] In Gallo et al.'s (1997) procedure, the critical item was never presented in the lists. Therefore, if subjects could discern the identity of the critical item, they would know to judge it new, regardless of whether they remembered it. Our procedure does not permit this possibility because half the time the critical item was included in the list.

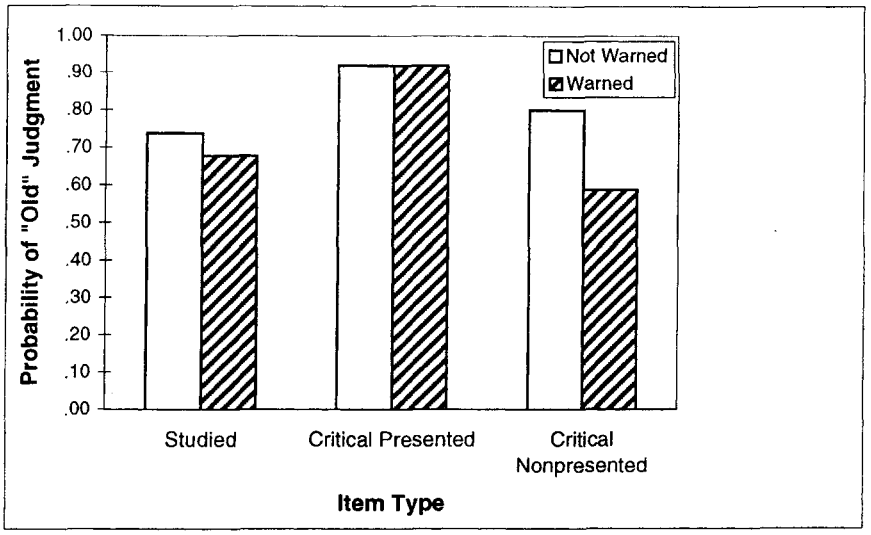

Figure 2. Probability of "old" judgments on recognition tests immediately after each list of associates (15 words per list) as a function of instructional set. Half of the subjects were informed of the false recognition phenomenon and instructed to avoid making such errors; the other half of the subjects received standard recognition instructions. The differences between the two instruction conditions reflect shifts in criteria.

Even if criterion shifts play some role in explaining false recognition in the Roediger–McDermott (1995) paradigm, the application to recall seems more tenuous, except as a redescription of the phenomenon under question, as being a one-item criterion shift. Roediger and McDermott (1995) showed high levels of false recall of the critical item on an immediate test under conditions of single-trial free recall. Subjects have always been assumed to set high criteria in this test (Cofer, 1967; Roediger & Payne, 1985), and the instructions Roediger and McDermott (1995) used warned explicitly against guessing and stressed accurate responding.

Part of the power of the memory illusion produced by the Roediger–McDermott procedure is in the experiential characteristics subjects report. As far as we know, it is the first procedure to be discovered in which, when subjects make false alarms, between 50% and 70% of these false alarms are given remember judgments when Tulving's (1985) remember/know procedure is used. In all other recognition memory experiments of which we are aware, the great preponderance of false alarms are deemed to be known, not remembered. Following the lead of others (e.g., Donaldson, 1996), Miller and Wolford (1999) argue that remember judgments are simply high-confidence recognition judgments and of no special consequence. The issue is certainly not settled, however, because, as Gardiner and his colleagues have pointed out in a series of publications, the alignment of remember/know and confidence judgments does not work in some situations. Data falsify predictions of the detection model with respect to remember/know judgments (Gardiner & Gregg, 1997); in addition, variables create dissociations between remember/know and sure/unsure judgments, revealing that they do not always have the same basis (see Gardiner & Conway, 1999; Rajaram, 1999).

In addition to claiming to remember the critical nonpresented words in the Roediger–McDermott paradigm, subjects also report a number of other compelling experiences that make the nonpresented events seem real. Payne, Elie, Blackwell, and Neuschatz (1996) showed that subjects reported remembering the voice in which illusory words had been presented. Also, subjects reported having rehearsed critical items during list presentation (Mather, Henkel, & Johnson, 1997). In short, subjects find recall and recognition of the critical nonpresented words a compelling experience and, rather like the case of perceptual illusions, one that cannot be overcome easily by instructions or shifting criteria.

Conclusion

Miller and Wolford (1999) argue that "most of the false memories" obtained in the Roediger–McDermott paradigm "could be ascribed to criterion shifts" (p. 398) and that they "believe that many people would be surprised, and a bit disappointed, to learn that some forms of false memories result from criterion shifts" (pp. 404, 405). We suspect they can rest easy on this score. We are not afflicted by this feeling ourselves, partly because Miller and Wolford (1999) provide no direct evidence for shifting criteria as responsible for the phenomenon. If subjects change their strategies to become more liberal, measures of bias must differ; however, the finding that measures of bias differ does not require the conclusion that subjects changed strategies, which is what Miller and Wolford (1999) claim. Of course, other evidence does lead us to believe that response criterion can play some limited role in false recognition (Gallo et al., 1997; McDermott & Roediger, 1998), but as reviewed previously, the illusion is largely resistant to criterion manipulations.

References

Balota, D. A., Cortese, M. J., Duchek, J. M., Adams, D., Roediger, H. L., McDermott, K. B., & Yerys, B. E. (in press). Veridical and false memories in healthy older adults and in dementia of the Alzheimer's type. *Cognitive Neuropsychology.*

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

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

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition, 24,* 523–533.

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.

Gardiner, J. M., & Conway, M. A. (1999). Levels of awareness and varieties of experience. In B. H. Challis & B. M. Velichovsky (Eds.), *Stratification of consciousness and cognition* (pp. 237–254). Amsterdam: John Benjamin.

Gardiner, J. M., & Gregg, V. H. (1997). Recognition memory with little or no remembering: Implications for a detection model. *Psychonomic Bulletin & Review, 4,* 474–479.

Green, D. M., Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Hirshman, E. (1995). Decision processes in recognition memory: Criterion shifts and the list strength paradigm. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 21,* 302–313.

Hirshman, E., & Arndt, J. (1997). Discriminating alternative conceptions of false recognition: The cases of word concreteness and word frequency. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 23,* 1306–1323.

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

Mather, M., Henkel, L. A., & Johnson, M. J. (1997). Evaluating characteristics of false memories: Remember/know judgments and Memory Characteristics Questionnaire compared. *Memory & Cognition, 25,* 826–836.

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

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., & Roediger, H. L. (1998). Attempting to avoid illusory memories: Robust false recognition of associates persists under conditions of explicit warnings and immediate testing. *Journal of Memory and Language, 39,* 508–520.

Miller, M. B., & Wolford, G. L. (1999). The role of criterion shift in false memory. *Psychological Review, 106,* 398–405.

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.

Rajaram, S. (1999). Accessing the nature of retrieval experience: Advances and challenges. In B. H. Challis & B. M. Velichovsky (Eds.), *Stratification of consciousness and cognition* (pp. 255–275). Amsterdam: John Benjamin.

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. Gathecole, & C. Cornoldi (Eds.), *Theories of memory II* (pp. 187–245). Hove, Sussex: Psychology Press.

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

Stretch, V., & Wixted, J. T. (1998). On the difference between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 24,* 1379–1396.

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

Wixted, J. T., & Stretch, V. (1998). *The evidence against a criterion shift account of false memory.* Manuscript submitted for publication.

Received October 1, 1998
Revision received December 14, 1998
Accepted December 14, 1998 ■