

Test-induced priming of false memories

ELIZABETH J. MARSH

Duke University, Durham, North Carolina

AND

PATRICK O. DOLAN

Drew University, Madison, New Jersey

Of interest was whether prior testing of related words primes false memories in the Deese/Roediger–McDermott (DRM) paradigm. After studying lists of related words, subjects made old–new judgments about zero, three, or six related items before being tested on critical nonpresented lures. When the recognition test was self-paced, prior testing of list items led to faster false recognition judgments, but did not increase the rate of false alarms to lures from studied lists. Critically, this pattern changed when decision making at test was speeded. When forced to respond quickly—presumably precluding the use of monitoring processes—clear test-induced priming effects were observed in the rate of false memories. The results are consistent with an activation–monitoring explanation of false memories and support that retrieving veridical memories can be a source of memory error.

In the Deese/Roediger–McDermott (DRM) paradigm, studying lists of related words such as *sour*, *candy*, and *sugar* leads to high rates of false recall and recognition of critical nonpresented words, such as *sweet* (Deese, 1959; Roediger & McDermott, 1995). Subjects are confident in their memories; they claim to “remember” studying the nonpresented words (Roediger & McDermott, 1995), and they are willing to attribute their memories to a speaker (Payne, Elie, Blackwell, & Neuschatz, 1996; Roediger, McDermott, Pisoni, & Gallo, 2004). False memories created via this method are remarkably persistent, even when subjects are explicitly forewarned about the illusion (Gallo, Roberts, & Seamon, 1997; Gallo, Roediger, & McDermott, 2001; McDermott & Roediger, 1998) or have multiple chances to study the list (McDermott, 1996).

Because of the robustness of the DRM illusion, it is a useful method for testing theories about false memory more generally. One popular explanation is the activation–monitoring account, which posits that false memories are actually encoded at study as they are activated via a spreading activation mechanism. Namely, activation from studied words spreads to other related items in the semantic network (Collins & Loftus, 1975); in the DRM paradigm, this is seen in the study of many related words leading to heightened activation of the critical lure (see, e.g., Roediger et al., 2001). Whether or not this activation leads to the conscious encoding of the critical lure is still open for debate (e.g., Marsh & Bower, 2004; Seamon, Luo, & Gallo, 1998), but the activation–monitoring framework can accommodate either outcome. Rather, a key part of the framework is that subjects attempt to monitor the

source of activations at test. Monitoring takes time, and hence speeded retrieval disrupts these processes (Benjamin, 2001; Dodson & Hege, 2005; Heit, Brockdorff, & Lamberts, 2004).

Most discussions of the activation–monitoring framework focus on whether or not critical lures are activated during the study phase, with strategic processes to reduce false memories at test. However, in their 1995 article, Roediger and McDermott pointed to the possibility that false memories might also be created at test. This hypothesis was driven by the observation that the critical lure is often recalled near the end of the list in free recall, suggesting it may have benefited from priming from the recall of other related study items. Similarly, in recognition, the critical lure is often tested after other list items, again allowing for test-induced priming (see, e.g., Neely, Schmidt, & Roediger, 1983). However, the two experiments we know of that directly tested this hypothesis failed to find strong evidence for test-induced priming in the DRM paradigm (Dodd, Sheard, & MacLeod, 2006; Marsh, McDermott, & Roediger, 2004). In the earlier study by Marsh et al., subjects made old–new recognition judgments about critical lures from studied and nonstudied lists after they had been tested on zero, three, or six related list items. Test-induced priming was only found for critical lures from nonstudied lists. That is, when the lists had *not* been studied (when subjects had not studied *sour*, *candy*, and *sugar*), prior testing of related (new) words, such as *sour* and *candy*, increased false alarms to *sweet*. When the lists had been studied, prior testing of list items did not increase false memories. When subjects had actually studied words such as *sour*, *candy*, and

E. J. Marsh, emarsh@psych.duke.edu

sugar, then prior testing of related words such as *sour* and *candy* did not increase false alarms to *sweet*. False recognition of critical lures from studied lists was not at ceiling. Rather, the proposed explanation was that even if the test increased lure activation, subjects were able to monitor the source of this activation and thus avoid an increase in false memories; the current study tests that hypothesis.

The Dodd et al. (2006) study¹ used a different methodology to reach a similar conclusion: no role for test-induced priming in the typical DRM paradigm. They examined false alarms in an immediate 10-item recognition memory test after each list of 12 items had been presented. Of particular interest is their Experiment 2 in which subjects were required to respond to each test item within 750 msec. Surprisingly, given the predictions of the activation–monitoring theory, they still found no effect of their test manipulation in a speeded response condition: False alarms were similarly high regardless of how many just-studied items were tested prior to the lure. However, false alarms appeared very high in their speeded condition and did not allow a lot of room for observing priming; the baseline false alarm rate to critical lures was .84. In the present work, we used a delayed recognition test (as was used in Marsh et al., 2004) to move false alarm rates away from ceiling.

A demonstration of test-induced priming in the DRM paradigm would be important for several reasons. First, the majority of test manipulations in the DRM paradigm to date have been targeted at demonstrating the role of monitoring processes; the contribution of activation to increasing the illusion at test has not been documented. In fact, related work suggests that subjects may benefit from prior testing of studied items since the retrieval of actual memories will provide a strong standard for what should be considered “old” (see, e.g., Brainerd, Reyna, & Kneer, 1995). Second, the demonstration of test-induced priming would be important theoretically, given its consistency with the activation–monitoring account of the DRM illusion. That is, the same mechanism leading to the activation of the critical lure during study should increase its activation at test, with consequences for false memory. The failure of Marsh et al. (2004) and Dodd et al. (2006) to find test-induced priming of false memories is problematic for the activation–monitoring account, meaning that it is an important test of activation–monitoring theory to further examine this issue. Thus, the experiment is important both for its theoretical implications and as a demonstration.

The present research reexamines the test-induced priming issue. The self-paced testing procedure of Marsh et al. (2004) was extended with one important methodological change: Reaction times (RTs) were collected to see if a more sensitive measure demonstrated priming even if false alarm data did not. Second, a speeded responding condition was added to allow examination of priming in a condition that removed potentially camouflaging monitoring behavior.

To preview: Subjects studied lists of related words and then made old–new decisions either at their own pace or under deadline. In both conditions, critical lures from

studied and nonstudied lists were tested after zero, three, or six related list items. Of interest was test-induced priming in false alarms and RTs.

METHOD

Subjects

Forty-eight Drew University students participated in the study for course credit. Five subjects were excluded from the analyses because their postexperiment questionnaires indicated that they knew the study was about false memories. Seven subjects were excluded because they were unable to discriminate between studied and new items. Thus, data from 36 subjects was included in the analyses.

Materials

The materials were exactly the same as those used in Marsh et al. (2004). Subjects studied half of the 36 lists (each containing 15 words) from Stadler, Roediger, and McDermott (1999). Each list of words was related to a critical nonpresented lure, and words within a list were presented in order from the strongest to the weakest associate (e.g., one list contained *sour*, *candy*, *sugar*, *bitter*, *good*, *taste*, *tooth*, *nice*, *honey*, *soda*, *chocolate*, *heart*, *cake*, *tart*, and *pie*, but not the critical lure, *sweet*). The lists were presented in random order on a computer screen, using MediaLab and DirectRT (Jarvis, 2004a, 2004b).

For counterbalancing purposes, the lists were split into two sets of 18 lists (Sets 1 and 2). Half the subjects studied the lists in Set 1, and the other half studied the lists in Set 2. The two sets were broken down further into three sets each (1A, 2A, 1B, 2B, 1C, 2C), which contained six lists each. The six sets were needed to counterbalance which items were studied (Set 1 or Set 2) and the position in which items were tested (lures were tested after zero, three, or six related items across subjects; thus, lures were rotated through the three conditions). Subjects who studied Set 1 studied the lists in Sets 1A, 1B, 1C, while subjects who studied Set 2 studied lists in Sets 2A, 2B, and 2C.

Procedure

Each subject was seated at a computer and was instructed that 18 lists of 15 words would be presented, that each word would be presented for 1 sec, and that they should try to remember the words for a later memory test.

After the presentation of the 18 lists, the instructions for the memory test appeared. All subjects were instructed to respond by pressing either the key labeled *old* or the one labeled *new*. Each new trial was initiated via a press of the space bar. In the *control* condition, subjects responded to each test item at their own pace. In the *speeded* condition, subjects had to respond to each item within 750 msec. When a response was made after 750 msec had elapsed, the words “Please try to respond faster” appeared briefly.² To ensure understanding of the test parameters, all subjects studied a practice list and took a practice test on those words prior to the actual memory test.

Following the memory test, all subjects wrote a response to the question “What do you think the experiment was about?” They were then thanked and debriefed.

RESULTS

All results were significant at the .05 level unless otherwise noted.

Recognition of List Items

We computed a 2 (list status: old or new) \times 2 (test responding: self-paced or speeded) ANOVA on proportion of list items called “old.” As was expected, subjects

Table 1
Average Proportion of List Items and Critical Lures Called "Old" in the Control and Speeded Responding Conditions

	Primes	Test Responding			
		Self-Paced		Speeded	
		Studied	Nonstudied	Studied	Nonstudied
List items		.69	.18	.51	.25
Critical lures	0	.70	.16	.44	.32
	3	.72	.31	.67	.37
	6	.76	.30	.69	.34
	Mean	.73	.25	.60	.35

Note—For critical lures, the data are shown as a function of test position (after zero, three, or six related list items) in addition to mean probabilities.

were more likely to correctly identify list items as "old" if the lists had been studied ($M = .60$) than if they had not ($M = .22$) [$F(1,34) = 257.34$, $MS_e = .01$]. However, this effect was qualified by an interaction with test responding [$F(1,34) = 25.67$, $MS_e = .01$]. Speeded responding at test significantly decreased the hit rate from 69% to 51% [$t(34) = 3.89$, $SE = .04$] but did not have a significant effect on the false alarm rate (18% vs. 25%) [$t(34) = 1.68$, $SE = .04$].

False Recognition of Critical Lures

We computed a 2 (list status: old or new) \times 3 (number of test primes: 0, 3, 6) \times 2 (test responding: self-paced or speeded) ANOVA on proportion of critical lures called "old." The data are shown in Table 1.

An examination of Table 1 reveals opposite patterns in the self-paced and speeded conditions. Statistics supported this observation; the three-way interaction between list status, number of test primes, and speed of test responding was significant [$F(2,68) = 4.22$, $MS_e = .04$]. The data in the self-paced condition replicated the findings of Marsh et al. (2004): When the lists had been studied, false alarms were unaffected by prior testing of list items ($ts < 1$). When the lists had not been studied, false alarms increased above baseline after testing of three [$t(17) = 2.05$, $SE = .07$, $p = .057$] or six list items [$t(17) = 3.07$, $SE = .045$]. In contrast, in the speeded condition, the number of previously tested items affected only critical lures from studied lists. When the lists had been studied, false alarms in the speed responding condition increased above baseline after testing of three [$t(17) = 3.57$, $SE = .06$] or six list items [$t(17) = 3.91$, $SE = .07$]. When responding was speeded, the number of test primes had no effect on false alarms to critical lures from nonstudied lists ($ts < 1$).

In short, we replicated the findings of Marsh et al. (2004) that test-induced priming only occurred for critical lures when the lists had been studied—but this pattern reversed when old–new judgments were made under deadline.

Reaction Times

RTs from the control self-paced condition were analyzed on the basis of whether the lures were accepted as old (false memories) or correctly rejected (called "new"). These data are shown in Table 2. Separate one-way ANO-

VAs were computed on false alarms and correct rejections of critical lures from studied lists. The priming effect was significant for false alarms to critical lures from studied lists [$F(2,34) = 5.02$, $MS_e = 184,989.69$, $p < .02$]. Prior testing of studied items increased the speed with which subjects made memory errors. Priming was not significant when the lures were correctly rejected. Similar analyses were conducted on false alarms and correct rejections of critical lures from nonstudied lists, but did not reach significance.

There were no priming effects in the RT data from the speeded condition. These data are not presented here since the response deadline meant that there was little variability.

In summary, a priming effect was found in RTs of subjects in the self-paced condition. Even though their false alarms did not increase significantly after testing related items, when errors were made, they were committed quickly.

DISCUSSION

The self-paced condition replicated that of Marsh et al. (2004): Prior testing of list items increased false alarms to critical lures from nonstudied lists, but had no reliable effect on rates of false alarms to critical lures from studied lists. However, prior testing of related words did speed subjects' acceptance of critical lures from studied lists. Using a more sensitive measure of priming (RTs) allowed us to demonstrate test-induced priming in the standard (self-paced) DRM test condition.

Before turning to the data from the speeded condition, we wish to further discuss the priming effect found in the RT data. Prior testing sped false alarms to critical lures—a result consistent with the concept of spreading activation and similar to standard semantic priming results (see, e.g., Meyer & Schvaneveldt, 1971). But could an argument be made that prior testing should have slowed responding to critical lures? That is, to the extent that monitoring a more activated lure is difficult, one might have expected prior testing to have slowed the evaluation of critical lures. Key is that the priming effect occurred for false alarms—items for which monitoring failed or was not engaged. A systematic priming

Table 2
Average Times (in Milliseconds) to Respond "Old" Versus "New" to List Items and Critical Lures in the Self-Paced Test Condition

	Primes	Studied		Nonstudied	
		"Old"	"New"	"Old"	"New"
List items		1,147.2	1,405.4	1,542.6	1,278.5
Critical lures	0	1,450.9	1,533.9 (14)	1,397.4 (8)	1,471.2
	3	1,157.3	1,945.0 (15)	1,288.9 (14)	1,292.9
	6	1,003.7	1,570.0 (11)	1,290.0 (13)	1,502.3
	Mean	1,203.9	1,794.8 (16)	1,305.5	1,422.2

Note—For critical lures, the data are also shown as a function of test position (after zero, three, or six related list items). The number in parentheses represents the number of participants with observations in a cell if n was not 18.

effect was not observed on trials in which subjects successfully rejected the critical lures, presumably because any increase in activation (that would lead to speeding) was successfully offset by monitoring processes (which would lead to slowing).

One other puzzle is that subjects in the self-paced condition only appeared to monitor the lures from studied lists, not from nonstudied lists. This result—found both in the current study and in that of Marsh et al. (2004)—is somewhat surprising because, almost by definition, the monitoring task should be more difficult in the condition in which subjects had studied the list. As one reviewer of an earlier version of this paper asked, “Why doesn’t activation monitoring work for the nonstudied lures?” One possibility is that subjects may have based their decisions on different evidence for items from studied versus nonstudied lists. When subjects encounter items related to a studied theme, they may evaluate these items differently than ones not related to any of the study themes. This is not the same as arguing for a different criterion for lures versus studied items (as was argued by Miller & Wolford, 1999, and argued against by Wixted, 2000); rather, the argument is that subjects may evaluate distinct classes of items on different bases. For example, Dobbins and Kroll (2005) recently showed that people used different criterion for making old–new recognition memory judgments about pictures from their home campus than for pictures from an away campus (samples of both campuses had been studied during the first part of the experiment). Their argument was that “with sufficient response time and recent encoding, observers demand more evidence for conceptually distinctive items, perhaps because such items typically foster vivid recollection during retrieval” (p. 1186). Returning to the terms of activation–monitoring, the claim would be that subjects would be more likely to actively monitor test items that relate to a familiar theme (lures from studied lists) than those from an unfamiliar theme (lures from unstudied lists). Hence, when there is time to monitor, priming is only found in the less monitored class, unstudied lists. When there is not enough time to monitor (the speeded condition), priming can be revealed in the studied list condition as well.

Overall, the data support the hypothesis that retrieval factors contribute to the creation of false memories. That is, rather than showing how retrieval factors can reduce the illusion (e.g., by encouraging subjects to engage in monitoring processes or by raising the standard for “old” items), the current results point to how the test phase can create an illusion. Even standard DRM tests can speed acceptance of false memories, and a response deadline can increase the rate of false memories.

The test-induced priming effect is a clear extension of the activation–monitoring framework, with the same mechanisms increasing activation at study and test. In both the self-paced and speeded conditions, critical lures accrued activation from the primes (as supported by the RT data from the control condition). However, this activa-

tion only translated into a higher rate of false memories when memory monitoring was reduced.

AUTHOR NOTE

We thank Veronica Zak and Eric Kleeman for collecting the data, Holli Sink for programming the experiment, and Barbie Huelser for assistance with data analysis and manuscript preparation. Address correspondence to E. J. Marsh, Psychology and Neuroscience, Duke University, 9 Flowers Drive, Box 90086, Durham, NC 27708-0086 (e-mail: emarsh@psych.duke.edu).

Note—Accepted by David A. Balota’s editorial team.

REFERENCES

- BENJAMIN, A. S. (2001). On the dual effects of repetition on false recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *27*, 941-947.
- BRAINERD, C. J., REYNA, V. F., & KNEER, R. (1995). False-recognition reversal: When similarity is distinctive. *Journal of Memory & Language*, *34*, 157-185.
- 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.
- DOBBINS, I. G., & KROLL, N. E. A. (2005). Distinctiveness and the recognition mirror effect: Evidence for an item-based criterion placement heuristic. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *31*, 1186-1198.
- DODD, M. D., SHEARD, E. D., & MACLEOD, C. M. (2006). Re-exposure to studied items at test does not influence false recognition. *Memory*, *14*, 115-126.
- DODSON, C. S., & HEGE, A. C. G. (2005). Speeded retrieval abolishes the false-memory suppression effect: Evidence for the distinctiveness heuristic. *Psychonomic Bulletin & Review*, *12*, 726-731.
- 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.
- GALLO, D. A., ROEDIGER, H. L., III, & MCDERMOTT, K. B. (2001). Associative false recognition occurs without strategic criterion shifts. *Psychonomic Bulletin & Review*, *8*, 579-586.
- HEIT, E., BROCKDORFF, N., & LAMBERTS, K. (2004). Strategic processes in false recognition memory. *Psychonomic Bulletin & Review*, *11*, 380-386.
- JARVIS, B. G. (2004a). DirectRT (Version 2004.1.0.55) [Computer software]. New York: Empirisoft Corporation.
- JARVIS, B. G. (2004b). Medialab (Version 2004.2.87) [Computer software]. New York: Empirisoft Corporation.
- MARSH, E. J., & BOWER, G. H. (2004). The role of rehearsal and generation in false memory creation. *Memory*, *12*, 748-761.
- MARSH, E. J., MCDERMOTT, K. B., & ROEDIGER, H. L., III (2004). Does test-induced priming play a role in the creation of false memories? *Memory*, *12*, 44-55.
- MCDERMOTT, K. B. (1996). The persistence of false memories in list recall. *Journal of Memory & Language*, *35*, 212-230.
- MCDERMOTT, K. B., & ROEDIGER, H. L., III (1998). Attempting to avoid illusory memories: Robust false recognition of associates persists under conditions of explicit warnings and immediate testing. *Journal of Memory & Language*, *39*, 508-520.
- MEYER, D. E., & SCHVANEVELDT, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*, 227-234.
- MILLER, M. B., & WOLFORD, G. L. (1999). The role of criterion shift in false memory. *Psychological Review*, *106*, 398-405.
- NEELY, J. H., SCHMIDT, S. R., & ROEDIGER, H. L., III (1983). Inhibition from related primes in recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *9*, 196-211.
- 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 & Language*, **35**, 261-285.
- ROEDIGER, H. L., III, BALOTA, D. A., & WATSON, J. M. (2001). Spreading activation and arousal of false memories. In H. L. Roediger III, J. S. Nairne, I. Neath, & S. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 95-115). Washington, DC: American Psychological Association.
- ROEDIGER, H. L., III, & McDERMOTT, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 803-814.
- ROEDIGER, H. L., III, McDERMOTT, K. B., PISONI, D. B., & GALLO, D. A. (2004). Illusory recollection of voices. *Memory*, **12**, 586-602.
- SEAMON, J. G., LUO, C. R., & GALLO, D. A. (1998). Creating false memories of words with or without recognition of list items: Evidence for non-conscious processes. *Psychological Science*, **9**, 20-26.
- STADLER, M. A., ROEDIGER, H. L., III, & McDERMOTT, K. B. (1999). Norms for word lists that create false memories. *Memory & Cognition*, **27**, 494-500.
- WIXTED, J. T. (2000). The case against a criterion-shift account of false memory. *Psychological Review*, **107**, 368-376.

NOTES

1. We learned of the Dodd et al. (2006) article only after our own was in the review process.
2. In the speeded condition, on average 9.8% of trials were eliminated because responding was slower than 750 msec. Slow responding did not vary as a function of study status of the list or number of primes (all $F_s < 1$).

(Manuscript received December 31, 2005;
revision accepted for publication June 27, 2006.)