MEMORY, 2004, *12* (1), 44–55

Does test-induced priming play a role in the creation of false memories?

Elizabeth J. Marsh, Kathleen B. McDermott, and Henry L. Roediger, III
Washington University in St. Louis, MO, USA

We investigated the role of test-induced priming in creating false memories in the Deese/Roediger-McDermott (DRM) paradigm, in which subjects study lists of related words (*bed*, *rest*, *awake*) and then falsely recall or recognise a related word (*sleep*) on a later test. However, in experiments using three different procedures, we found that the number of related words tested prior to the critical word had surprisingly little impact on false recall and recognition. We manipulated the location of the critical item in tests of yes/no recognition, word-stem cued recall, and part-set cued recall. We consistently obtained high probabilities of false recall and recognition, but the probability was unaffected by the number of related items presented prior to the test of the critical item. Surprisingly, test-induced priming of the critical item does not seem to play a large role in this memory illusion.

The Deese-Roediger-McDermott (DRM) paradigm provides a straightforward means for creating false memories. Subjects study lists of related words such as *sour*, *candy*, *sugar*, *bitter*, *good*, *taste*, *tooth*, *nice*, *honey*, *soda*, *chocolate*, *heart*, *cake*, *tart*, and *pie*. When later tested, subjects falsely recall and recognise non-presented critical lures (*sweet*) with unusually high probabilities—often with a probability equivalent to studied items (Deese, 1959; Roediger & McDermott, 1995). Subjects are confident in their memories, and believe that they ‘‘remember’’ studying the non-presented lures (Roediger & McDermott, 1995). They are willing to describe their false memories in some detail, attributing the critical lure to having been presented early in the list (Read, 1996), and to one of two speakers (Payne, Elie, Blackwell, & Neuschatz, 1996). It is as if subjects are remembering a real event.

The DRM paradigm has been used successfully to create false memories in hundreds of subjects (see Roediger, McDermott, & Robinson, 1998, for a partial review). Indeed, it is remarkably difficult to eliminate or even significantly reduce the effect (Gallo, Roberts, & Seamon, 1997; Gallo, Roediger, & McDermott, 2001; McDermott, 1996; McDermott & Roediger, 1998; Neuschatz, Payne, Lampinen, & Toglia, 2001). How is it that such strong false memories are created? One possibility (among several) raised in Roediger and McDermott’s original paper is that false memories may be at least partially created during the test phase. Roediger and McDermott (1995, p. 811) argued that

> … retrieval processes may contribute significantly to the false recall and false recognition phenomena we have observed. Subjects usually

Correspondence should be addressed to Elizabeth J. Marsh, Psychological and Brain Sciences, Duke University, 9 Flowers Drive, Box 90086, Durham, NC, 27708-0086, USA. Email: emarsh@psych.duke.edu

We thank Ameet Parikh and Elyce Kirschenbaum for their assistance with data collection, and Kristina Olson for her help with data input. Keith Payne provided assistance with data analyses. We thank Dave Gallo for discussion of data and theory, and assistance with the ERTS program. This paper has benefited from the insightful reviews of Dan Wright and Alan Richardson-Klavehn. These data were presented at the 2000 meeting of the Psychonomic Society held in New Orleans. The first author was supported by an NRSA postdoctoral fellowship from the National Institute of Mental Health, #1F32MH12567, and the second author was supported by NIMH grant MH62514.

http://www.tandf.co.uk/journals/pp/09658211.html DOI:10.1080/09658210244000405

recalled the critical word toward the end of the set of recalled items, so prior recall may trigger false recall. Also, in the recognition test, presentation of words related to a critical lure often occurred prior to its appearance on the test; therefore, activation from these related words on the test may have enhanced the false recognition effect by priming the lure (Neely, Schmidt, & Roediger, 1983).

Such an account of test-induced priming is consistent with activation-based accounts of the illusion (e.g., Roediger, Balota, & Watson, 2001a; Seamon, Luo, & Gallo, 1998). The spreading activation mechanism posits that activation from the studied list items spreads to other related words in the semantic network (Collins & Loftus, 1975), and that a false memory occurs for the critical lure because its high activation level is misattributed as arising from prior presentation of the word (i.e., a source error; Johnson, Hashtroudi, & Lindsay, 1993; Johnson & Raye, 1981). We also note that current theorising invokes a second process, a monitoring process (e.g., Israel & Schacter, 1997; McDermott & Watson, 2001; Smith & Hunt, 1998). During retrieval, this monitoring process is hypothesised to be active as subjects make decisions about the source of an item's activation. We will return to the monitoring process in our general discussion; for now, we wish to focus on how spreading activation might be involved in the illusion, both during study and at test.

Several results suggest that spreading activation plays a role in the creation of false memories. First, false memories increase with the number of related words studied (Robinson & Roediger, 1997); longer lists allow activation from more related concepts to spread to the critical lure. Second, priming effects are often interpreted as the result of activation spreading through a semantic network, and false memories occur on implicit memory tests such as word fragment completion, word stem completion, and tests of word association (McDermott, 1997; see also McKone & Murphy, 2000). Third, populations such as older adults and those with dementia show equivalent (or even increased) levels of false memories even though they are impaired at remembering list items (Balota et al., 1999; Norman & Schacter, 1997; Watson, Balota, & Sergent-Marshall, 2001). False memories in such populations can be attributed to activation spreading through a still-intact semantic network. Fourth, associative strength is an excellent predictor of false memories. A recent regression analysis yielded backward associative strength as the most important predictor of false recall (Roediger, Watson, McDermott, & Gallo, 2001b; see also Deese, 1959; Gallo & Roediger, 2002; McEvoy, Nelson, & Komatsu, 1999). Backward associative strength (BAS) is a measure of how likely list words are to elicit the critical lure as an associate on a free association test. High BAS implies a high relation between list items and the critical word, which in turn implies that activation will spread to the critical word and result in false memories.

Spreading activation is not limited to the study phase; indeed, as described already, Roediger and McDermott (1995) proposed it to occur both during study and at test. In the current set of experiments, we examined three situations in which it was possible for activation to spread from test items to the critical lures, to determine the extent to which test-induced priming seems to contribute to the false memory effect. All three experiments shared a similar structure; subjects studied DRM lists and then took a memory test. Of interest across the experiments was whether the presence of related cues at test increased false memories, presumably via test-induced priming or activation. In Experiment 1, we manipulated the number of studied items tested prior to critical lures on a recognition test. In Experiment 2, we manipulated the number of studied items tested prior to critical lures on a word-stem cued recall test. In Experiment 3, we examined the effect of intra- and extra-list cues on recall of critical lures. To our surprise, across all three experiments, we found no evidence for test-induced priming of critical lures for studied lists.

EXPERIMENT 1: RECOGNITION

Method

Participants. A total of 36 Washington University students participated in the study. They were recruited through the psychology department's subject pool and received either course credit or monetary compensation for their participation. Participants were tested either individually or in groups of up to three people.

Materials. The stimuli were the 36 lists (each containing 15 words) from Stadler, Roediger, and

McDermott (1999). Each list of 15 words converged onto a critical non-presented lure. The order of words within the list was from strongest to weakest associate. For example, one list contained *bed*, *rest*, *awake*, *tired*, *dream*, *wake*, *snooze*, *blanket*, *doze*, *slumber*, *snore*, *nap*, *peace*, *yawn*, *drowsy*, but not the critical lure *sleep*.

For counterbalancing purposes, the 36 lists from Stadler et al. (1999) were divided into two sets of 18 lists (Sets 1 and 2). Each list in Set 1 was paired with a list in Set 2; each set was then further broken down into three sets that preserved the original pairings (Sets 1A, 2A, 1B, 2B, 1C, 2C). As will be described more fully in the next two paragraphs, these six sets were needed to counterbalance which items were studied (subjects studied lists in Set 1 or Set 2) and in which position items were tested (lures were rotated through three conditions so that across subjects each was tested after zero, three, or six related items).

Half the participants studied the lists in Set 1, whereas the remaining participants studied the lists in Set 2. For example, if a subject studied Set 1, they studied the lists in Sets 1A, 1B, and 1C. If a subject studied Set 2, they studied the lists in Sets 2A, 2B, and 2C. Thus each participant studied 18 lists. Two different random study orders were created for each of the sets. Thus, there were four study conditions that were counterbalanced across participants.

The memory test was constructed to fulfil a number of constraints. No more than two items from the same list were tested consecutively, nor were there more than five consecutive ‘‘old’’ (or ‘‘new’’) items. The recognition test contained 297 items: 6 list items from each of the 36 lists (18 studied, 18 non-studied), the 36 critical lures, and 45 unrelated fillers. A filler item was always tested immediately before each of the critical lures. The remaining fillers were placed primarily at the beginning and end of the test.

The experimental manipulation involved the placement of the critical lure in relation to its corresponding six list items. Across three different versions of the recognition test, the position of the critical lure was manipulated so that it was tested after zero, three, or six list items. Figure 1 illustrates how a subset of the recognition test was structured. Across subjects, three critical lures (one each from the A, B, and C sets) were rotated through three positions; the position of list items never changed. Three list items from each of the three related lists were always tested between the first and second critical lure positions. The remaining three list items were always tested between the second and third critical lure positions. Thus, the critical lure tested first was tested before any list items. The critical lure tested second was tested after three list items, and the critical lure tested third was tested after six list items. The length of the test (297 items) and filler words interspersed throughout the test helped to disguise the test structure. In addition, testing of list triplets overlapped so that words from other lists were interspersed between each pairing of critical lures.

Thus, study order, study set, and test position were counterbalanced across subjects, leading to 12 different versions of the experiment.

Procedure. Each participant was seated at a computer. The experiment was presented from a DOS platform via ERTS software. All instructions were presented visually on the computer screen; the experimenter read these aloud while the subject read them silently. Subjects were told they would see 18 lists of 15 words each, and that they should try to remember the words for a later memory test. Participants were told the words would be presented at a rate of one word per second, and that the lists would be visually delineated via a ‘‘next list’’ message on the screen.

Each word was presented for 1 second in light grey text on a black background; the inter-stimulus interval was 300 ms. ‘‘Next list’’ prompts were presented for 2 seconds. After the presentation of the 18 lists, a message appeared informing subjects to ask the experimenter for additional instructions.

The experimenter then read aloud the instructions for the recognition memory test. Participants were told that they would be tested on 297 items, each of which would be numbered and would correspond to a number on their answer sheets. Participants controlled the rate of testing by pressing the enter key to receive the next item. For each word they made an old–new judgement. Participants were instructed to circle ‘‘old’’ on their answer sheets if they thought the word had been presented during the study phase and ‘‘new’’ if they did not think it had been studied. Participants then made a second judgement for words called ‘‘old’’. For old words, they circled ‘‘remember’’, ‘‘know’’, or ‘‘guess’’ to describe the recollected experience (Gardiner & Conway, 1999). ‘‘Remember’’ was defined as memory for the word accompanied by conscious recollection of specific aspects of its presentation. ‘‘Know’’ was

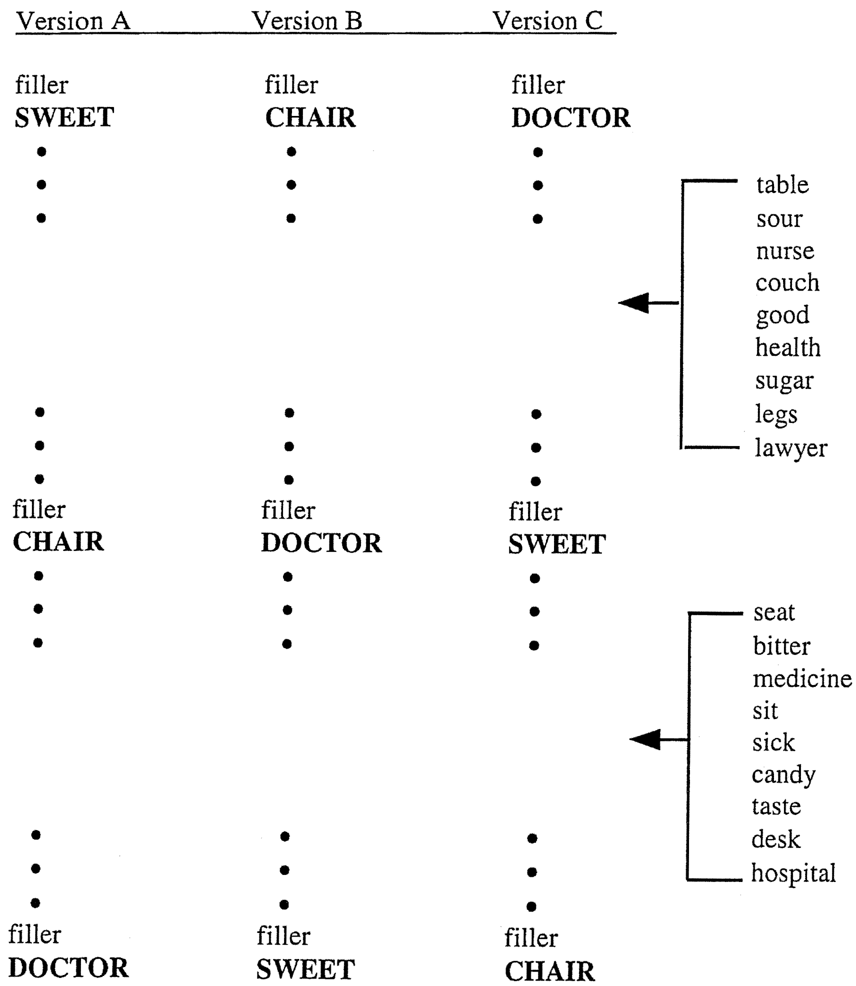

Figure 1. A simplified example of the recognition memory test used in Experiment 1. Versions A, B, and C allowed for counterbalancing of critical lure position (following zero, three, or six related words). Half the subjects studied the lists associated with the critical lures *sweet*, *chair*, and *doctor*; the remaining subjects did not. A filler item always immediately preceded each critical lure. The same nine list items were always tested between each pairing of critical lures, interspersed with filler items and words from other lists.

defined as the belief that the word had been studied, but without conscious recollection of its occurrence. Finally, "guess" was defined as simply guessing the word had been presented without a firm belief that it had been studied.

Following the recognition test, participants answered a series of questions about the experiment designed to assess their pre-experimental knowledge and current awareness of the false memory paradigm. Finally, participants were informed about the purpose of the study and thanked for their participation.

Results

A .05 level of significance was used for each test. All results are significant at that level unless otherwise noted.

Recognition. As expected, participants were most likely to say "old" to words that had actually been presented during the study phase (.78, see Table 1). Critical lures from studied lists were falsely recognised significantly more often than either critical lures from unstudied lists (.65 vs .43), $t(35) = 4.41$, $SEM = .05$, or unrelated filler words (.27), $t(35) = 7.82$, $SEM = .05$. However, false recognition of critical lures from studied lists (.65) was still reliably lower than the hit rate (.78), $t(35) = 2.73$, $SEM = .05$.

The critical question is whether test position affected false memories. The data are shown in Table 1. A 2 (study status of list) × 3 (test position) ANOVA model was conducted on the proportion of items falsely recognised as "old". Both main effects were significant. Subjects were much more likely to falsely recognise lures when

TABLE 1
Experiment 1

	Studied lists		*Non-studied lists*	
Item-type	*"Old"*	*(R/K/G)*	*"Old"*	*(R/K/G)*
List items	.78	(.39/.24/.14)	.32	(.04/.10/.19)
Critical lures				
Test position:				
zero	.63	(.28/.25/.10)	.31	(.02/.09/.20)
three	.64	(.24/.27/.13)	.49	(.04/.15/.29)
six	.69	(.23/.27/.19)	.49	(.08/.19/.22)
M	.65	(.25/.27/.14)	.43	(.05/.14/.24)
Unrelated fillers		.27	(.03/.09/.15)	

Average proportion of list items, critical items, and unrelated filler items recognised ("Old") in Experiment 1 with corresponding "Remember" (R), "Know" (K), and "Guess" (G) responses in parentheses. For critical lures, the data are also shown as a function of test position (after zero, three, or six related items) in addition to mean probabilities.

Instances where "remember", "know", and "guess" judgements do not sum to the proportion of items recognised reflect rounding error.

the corresponding list had been studied, $F(1, 35) = 19.41$, $MSe = .14$. "Old" judgements also increased with test position, $F(2, 70) = 6.60$, $MSe = .04$. However, these main effects were qualified by a significant interaction between study status and test position, $F(2, 70) = 5.30$, $MSe = .03$. As shown in Table 1, while false recognition of lures related to non-studied lists increased with test position, little or no such priming effect was obtained for lures related to studied lists. To further understand the interaction, separate ANOVAs were conducted on the data for critical lures from studied versus unstudied lists. Regarding the lures related to studied lists, there was no main effect of test position on false recognition, $F(2, 70) = 1.37$, $MSe = .02$. However, a different pattern emerged for the critical lures from non-studied lists. Test position exerted a reliable effect on false recognition probability, $F(2, 70) = 8.26$, $MSe = .05$. When tested prior to related items, these lures were falsely recognised with about the same probability as the unrelated filler lures. When tested after three or six associates, false recognition increased significantly, $t(35) = 3.19$, $SEM = .06$. Obviously, the increase from three to six preceding associates did not increase false recognition ($t < 1$).

Phenomenology of recognition. The "remember"/"know"/"guess" results are presented in parentheses in Table 1. The data clearly show that subjects were most likely to make "remember" judgements for words they had actually studied (.39), and this probability was higher than for critical lures from studied lists (.25), $t(35) = 2.97$, $SEM = .04$. However, subjects did make many more "remember" judgements for critical lures from studied lists (.25) than from non-studied lists (.05), $t(35) = 6.30$, $SEM = .04$, and unrelated filler lures (.03), $t(35) = 5.85$, $SEM = .04$.

Of interest is how the phenomenology of false memories changed as a function of test position. We will first consider false memories related to studied lists. There were no significant differences in "remember" responses across test positions, $F(2, 70) = 1.41$, $MSe = .03$; subjects were equally likely to "remember" the lure if it had been tested after zero (.28), three (.24), or six related words (.23). However, "guess" responses did significantly increase with test position. Compared to when lures were tested first (.10), "guess" responses increased when critical lures were tested after three (.13) or six related studied items (.19), $F(2, 70) = 3.83$, $MSe = .08$. Thus, an increase in "guess" responses is responsible for the slight increase in false alarms to critical lures following six studied items.

A different pattern was found for false recognition of critical lures from non-studied lists. Prior responding to related words significantly increased both "remember", $F(2, 70) = 3.31$, $MSe = .03$, and "know" responses, $F(2, 70) = 3.70$, $MSe = .03$.

''Remember'' responses were almost non-existent in the zero condition (.02), and increased slightly following three (.04) or six related items (.08). A similar pattern occurred for the ''know'' responses; they were infrequent in the zero condition (.09) but increased following testing of three (.15) or six related items (.19). There was also a decrease in ''guess'' responses following prior testing of six (.22) as compared with three related items (.29), $t(35) = 2.30$, $SEM = .03$, which is not surprising because ''remember'' and ''know'' responses increased.

Discussion

In Experiment 1, only critical lures from non-studied lists showed priming from prior testing of list items. Subjects were more likely to incorrectly label these items as ''old'' if they were tested after three or six related items than if they were tested prior to the related items. It is interesting to note that this pattern becomes linear if guesses are removed; false recognition of critical lures from non-studied lists then increases from .11 following testing of zero related items, to .19 after testing of three related items, to .27 after testing of six related items. Prior testing of related items changed the phenomenology of false memories; there was a significant increase in false ''remember'' and ''know'' responses.

The pattern was very different for critical lures related to studied lists. Although such memories occurred frequently, they were unaffected by their position in the recognition test. Not only did these false memories fail to increase after testing of related items, but their phenomenology was also unaffected. If anything, prior testing of related items served only to increase guesses.

Our results are similar to those recently reported by Fernández, Diez, Alonso, and Beato (2001). In three experiments, they reported no significant effect of test position on the probability of false recognition. For example, in one experiment, the probability of false recognition of critical lures from studied lists was the same regardless of whether prior testing involved five strong associates, five weak associates, or no prior associates. However, as in our study, Fernández et al. also reported that a different pattern emerged for critical lures from non-studied lists. False recognition of these lures increased when tested after weak or strong associates.

We wish to draw the reader's attention to four points supporting our conclusion of a null result for critical lures from studied lists. Although there was a numerical hint of an increase in false recognition with the number of prior list items tested, the direct comparison between false recognition of lures preceded by zero (.63) versus six studied items (.69) failed to reach significance, $t(35) = 1.48$, $SEM = .03$. Four reasons compel us to believe we should accept the null hypothesis in this case. First, as noted earlier, the experiment was sensitive enough to detect an effect for critical lures from non-studied lists. Second, as described in the previous paragraph, our results are not isolated ones; they directly parallel those reported by Fernández et al. Third, there is sufficient power in the experiment to detect an effect. A statistical power analysis revealed that the power to detect a .09 effect (half the size of the effect found for non-studied lists) is .72. In order for the .06 difference (between the zero and six prior items conditions) to become significant at a reasonable level of power (.70), 99 subjects would have to be tested. Fourth, and most critically, the subjective report data show that guesses underlie any observed tendency for an increase in false recognition of critical lures from studied lists following testing of six related items. The DRM illusion is interesting in large part because of the phenomenology reported by the subjects—it is as if subjects are remembering real events. If ''guess'' responses are removed from Table 1, the data are unambiguous. False recognition of critical lures from non-studied lists would then increase from .11 (.02/.09, remember/know) after zero related words to .19 after three related words to .27 after six related words. If anything, the opposite pattern occurs for critical lures from studied lists. False recognition drops from .53 (.28/.25) after zero related words to .51 (.24/.27) after three related words to .50 (.23/.27) after six related words. The same pattern occurs if only ''remember'' responses are considered.

Thus, in Experiment 1 we successfully demonstrated test-induced priming by manipulating test order, but only when subjects had not studied the related list items. Why was there no test-induced priming of critical lures from studied lists? False recognition was not at ceiling, even for lures tested following six related items (.69). Thus, it seems there could have been room for an increase in activation levels with consequent effects on levels of false recognition.

EXPERIMENT 2: WORD-STEM CUED RECALL

In Experiment 2, we tried again to find test-induced priming of critical lures from studied lists. In Experiment 1, subjects' heavy memory load and the length of the test period may have worked to reduce test-induced priming. To minimise these problems in Experiment 2, a different procedure was used to look for effects of prior testing on false memories. Immediately after presentation of each list of words, subjects took a word-stem cued recall test in which critical lures were tested in the first, third, seventh, or thirteenth output position. This manipulation was stronger than the one we used previously; critical lures could be tested after as many as 12 items had been tested, and list items were tested closer in time to lures than in Experiment 1. However, as we shall describe, this stronger test manipulation still had no effect on the probability of false cued recall.

Method

Participants. A total of 44 Washington University undergraduates participated in the experiment in exchange for course credit.

Materials. A total of 24 lists from the Stadler et al. (1999) norms were used. These 24 lists were randomly divided into four sets. For each participant, one list set was assigned to each of the four test conditions (position 1, 3, 7, or 13). Across participants, test position was counterbalanced as each of the four list sets was rotated through each of the four test positions.

Test booklets contained 24 pages, one for each studied list. For each studied list, 15 word-stem cues were presented. Each cue consisted of the first three letters of the target list item, followed by blank underlined spaces. Of the 15 stem cues, 12 were items from the list, 2 were unrelated to the studied list, and 1 corresponded to the critical lure. The experimental manipulation involved the position of the critical lure stem cue. This critical lure stem cue was tested in the first, third, seventh, or thirteenth position. When the critical lure was not tested in the first position, it was always immediately preceded by the same two list item cues. The unrelated stems were tested at random following the testing of the critical lure.

Procedure. All participants were tested individually or in small groups. They were told they would see 24 lists of words and that each list would consist of 15 words presented at a rate of one per second. Participants were informed about the upcoming memory tests and thus encouraged to try and remember as many words as possible.

Following presentation of each list, subjects took a memory test. Each memory test consisted of a single page listing the 15 to-be-completed stems; the test booklet contained a total of 24 pages (one for each list). Subjects were given explicit cued recall instructions; they were told that if possible they should complete stem cues with studied words. They were warned that not all cues would correspond to studied words and thus they should only fill in those items that corresponded to studied words. They were also instructed not to guess.

After completing the memory test on a list, participants turned the page in their booklets and signalled their readiness to continue in the experiment. After all 24 study–test cycles were completed, participants were informed about the nature of the experiment and thanked for their participation.

Results

Overall, participants recalled 73% of the studied words and 64% of the critical lures. Although the false recall probability was well above that observed for unrelated filler items (5%), $t(43) = 16.22$, $SEM = .04$, it was significantly below the hit rate, $t(43) = 2.19$, $SEM = .04$. More importantly, false cued recall did not vary as a function of test position, with false recall varying unsystematically across test positions ($F < 1$). False cued recall was .66 when cued in the first test position, .63 when cued in the third test position, .66 when cued in the seventh test position, and .64 when cued in the thirteenth test position.

Discussion

As in Experiment 1, we failed to see an effect of our test manipulation on false memory. False cued recall was equivalent across the four output positions; it did not matter whether the critical lure was tested in the first, third, seventh, or thirteenth position. Two points strengthen this null conclusion. First, there was not even a trend towards an increase in false cued recall with later test position. Across the four test positions, false cued recall varied unsystematically between .63 and .66; the highest rate was not associated with the thirteenth position (but rather was obtained when the lure was tested either first or seventh). Second, the power to detect a true effect of .10 was .72. We believe a hypothesised effect size of .10 (the difference between false cued recall in the first versus thirteenth position) is reasonable given that testing after six lures in Experiment 1 led to effects of .18 and .06.

Before turning to the next experiment, we wish to comment on an interesting facet of the experiment and the consequent data. The test used, cued-recall, has been used in the DRM paradigm only by McKone and Murphy (2000). In our experiment, subjects completed 64% of word stems corresponding to critical lures even though they clearly understood they were only to complete word stems with studied words; they only completed 5% of stems corresponding to unrelated filler items. Therefore cued recall on an immediate test seems to provide an even stronger false memory effect in the DRM paradigm than does free recall, as levels of critical item recall on immediate tests is typically in the range of 30–40%.

EXPERIMENT 3: PART-SET CUEING

In Experiment 3 we used yet another method to prime critical lures at test. As described earlier, adding words to the study list increases the likelihood of false memories (Robinson & Roediger, 1997). In this experiment we added items to the test list by means of cues. We employed a part-set cueing experiment (see Roediger, 1973; Slamecka, 1968; Watkins, 1975). After study of 10-item DRM lists, subjects received one of three test conditions. Sometimes the subjects engaged in free recall; other times they were given either five intra-list or five extra-list cues to "help" them remember the studied items. Both intra-list and extra-list cues from the same semantic category actually inhibit rather than aid recall of list items, relative to the no-cue condition (Watkins, 1975). The question was whether providing cues at test (and hence adding related items) would increase false recall in the same way as increasing list length at study does (Robinson & Roediger, 1997).

Watkins (1975) proposed a cue-overload account of part-set cueing in which part-set cues were hypothesised to have an inhibitory effect on recall by effectively increasing the number of tokens that must be remembered. As is well known, the more items that are subsumed under a particular cue (e.g., the more items that are presented in a category, or simply the longer a list), the less effective is a retrieval cue (such as a category name) for provoking recall of a particular list item (e.g., Tulving & Pearlstone, 1966). If Watkins' (1975) hypothesis is correct (and other evidence is consistent with it; see Mueller & Watkins, 1977), it predicts an interesting pattern of data. Provision of part-set cues should decrease recall of list items but increase recall of the critical targets. Robinson and Roediger (1997) showed that adding related items to a study list decreased recall of list items but increased recall of the critical item, and Watkins' (1975) theory proposes that part-set cues work just like adding extra items to the list. On the other hand, other theories of the part-set cueing effect (Roediger, 1978; Rundus, 1973) would predict that recall of both list and critical items should decrease with provision of cues. Of course, other patterns are possible; part-set cues may have an effect on list items but no effect on critical items. This last possibility turns out to be the pattern we obtained.

Method

Participants. A total of 36 Washington University undergraduates participated either to fulfil a course requirement or for monetary compensation.

Materials. A total of 30 lists of words from Stadler et al. (1999) was used in Experiment 3.

Each list consisted of 15 words, which were subdivided into three sets matched for backward associative strength (A, B, C). Each of the sets was randomly designated to be the intra-list cues (A), the extra-list cues (B), or never cued (C). Subjects always studied the words comprising sets A (intra-list cues) and C (never cued); by definition, subjects never studied the words that would later be extra-list cues (Set B). Thus, each subject studied the same list of 10 words for each of the 30 lists.

Across the course of the experiment, each subject received 10 non-cued recall tests, 10 extra-list cued recall tests, and 10 intra-list cued recall tests. Non-cued recall tests required free recall of the 10 just-studied words. Extra-list cued recall tests required recall of the same 10 just-studied words, but the subject was prompted with the 5 related but unstudied cue words. Intra-list recall tests required the recall of 5 list items; subjects were given 5 list items to help them recall the remainder of the list. To keep the writing equivalent across conditions, participants were also instructed to record the cues if they remembered having studied them. Across subjects, the assignment of lists to test cue conditions was counterbalanced so that all lists appeared equally often in each of three testing conditions.

Procedure. At the beginning of the experiment, participants were informed that they would study 30 lists of words, and that each list of words would contain 10 words presented at rate of one per second. They were further informed about the math filler task and the upcoming memory tests. Participants were told about the possibility of receiving cues, and were instructed not to ignore them but rather to read them carefully and write down any that they thought they had studied.

Words appeared on the computer screen in black type against a light grey screen. After the presentation of a list, subjects turned to the next page in their booklet. The first page contained multiplication problems; the subjects worked on these for 30 seconds. They then turned to the next page in the booklet, and took the memory test on the just-studied list. Depending on the experimental condition, participants either saw a blank page on which to write down the to-be-remembered words, or they received five cue words followed by space for recording remembered words. They were given 75 seconds in which to complete this memory task. Participants were instructed to use care in turning the booklet pages so that they would keep pace with the computer program.

Results

A 2 (item-type) $\times$ 3 (test cue condition) ANOVA was conducted on proportion of items recalled. In the case of studied items, this dependent measure reflected the proportion recalled of the five never-cued items per list. Overall, subjects recalled a greater proportion of studied items (.58) than they intruded critical lures (.21), $F(1, 35) = 117.82$, $MSe = .07$. However, this factor interacted with cue condition, $F(2, 70) = 5.61$, $MSe = .01$. As shown in Table 2, a separate ANOVA on studied items revealed the standard part-set cueing effect, $F(2, 70) = 8.77$, $MSe = .01$. A greater proportion of studied items was recalled when no cues were given than when the cues were extra-list, $t(35) = 2.24$, $SEM = .02$, or intra-list, $t(35) = 4.29$, $SEM = .02$. The difference between the intra- and extra-list conditions was only marginally significant, $t(35) = 1.88$, $SEM = .01$. However, a separate ANOVA on proportion of false recall indicated that intrusions did not vary significantly across the three testing conditions, $F(2, 70) = 1.13$, $MSe = .01$. If anything, there was a tendency for false recall to be higher after part-set cueing; there was a trend towards an increase in false recall following intra-list cues, as compared to the no-cue condition, but it was not significant, $t(35) = 1.58$, $SEM = .03$, $p > .1$.

TABLE 2
Experiment 3

	Test cue condition		
	No cues	*Extra-list*	*Intra-list*
Veridicial	.62	.58	.55
False	.19	.20	.23

Veridical and false recall in no cue, extra-list, and intra-list cue conditions for Experiment 3.

Discussion

In Experiment 3, subjects showed the standard part-set cueing effect for veridical but not false recall. When cued with intra- or extra-list cues, subjects recalled fewer of the remaining list items than if they had received no cues at all. However, neither cue set had a significant effect on the frequency with which critical lures were intruded into recall. For the third time, a retrieval manipulation involving presentation of related items prior to recall of the critical target had no effect on creation of false memories.

Given that our manipulation was sensitive enough to detect small part-set cueing effects with veridical memories, we believe the manipulation was an effective one. It is possible, however, that the cues exerted two competing influences on false memories. On the one hand, the cues could have increased activation of the to-be-recalled lures, in the same way that additional list items increase false memories (Robinson & Roediger, 1997). On the other hand, the cues could have had an inhibitory effect on false memories, in the same way that they do for veridical memories (e.g., Roediger, 1973; Slamecka, 1968). These two effects combined may have yielded the null effect that we obtained, but of course such logic for null results is not compelling without evidence for the two compensatory processes. However, reports from other laboratories have revealed standard part-set cueing effects for false memories in the DRM paradigm (e.g., Kimball & Bjork, 2002; Reysen & Nairne, 2002), but in all cases these researchers used a wider range of part-set cues than the five cues employed in Experiment 3. Our manipulation of part-set cues was strong enough to affect list recall, so the null effect we obtained in false recall is probably not due to an ineffective manipulation. It should be noted that the inhibitory effect obtained by others is inconsistent with Watkins' (1975) cue overload theory, which maintains that part-set cues should have the same effect as adding items to the list. Our research and that of others fails to bear out this prediction.

GENERAL DISCUSSION

Across three experiments, we found little support for our prediction that increasing activation at test via presentation or testing of related items would increase the occurrence of false memories in the DRM paradigm. In Experiment 1, subjects falsely recognised critical lures from studied lists with the same probability regardless of whether they were tested after zero, three, or six of the studied words. In Experiment 2, subjects falsely recalled the critical lures with the same probability regardless of whether they were tested in the first, third, seventh, or thirteenth position of a cued recall test. In Experiment 3, subjects falsely recalled the critical lures with the same probability regardless of the presence of extra- or intra-list cues.

Why were we unsuccessful in priming false memories at test? One possibility is that there was a ceiling effect on activation levels. Although there was no ceiling effect on the probability of false memories in any of the studies, it is possible that the relevant ceiling effect is on conceptual priming. That is, for lures related to studied items, they might have received so much priming during the study phase that they would not show further effects at test from exposure to a few additional related items. Such a result would again support the strength of false memories created in the DRM paradigm.

Our results were not as we predicted; how well, then, do they mesh with the rest of the literature? On the one hand, our results join a growing literature on the surprising behaviour of false memories following retrieval manipulations. We deem this behaviour at test "surprising" because unlike other memory illusions, false memories created in the DRM paradigm are relatively unaffected by manipulations at test aimed at reducing the illusion. For example, in the eyewitness post-event information paradigm, post-encoding warnings can reduce suggestibility in some circumstances (e.g., Christiaansen & Ochalek, 1983; Wright, 1993). However, in the DRM paradigm, warnings at test do little to reduce the robustness of the illusion (Gallo et al., 1997, 2001; Neuschatz et al., 2001). Instructions to monitor source represent a second test variable that often reduces memory illusions. Although source monitoring typically reduces eyewitness suggestibility (Lindsay & Johnson, 1989) and erroneous judgements of fame (Multhaup, 1995), source monitoring instructions have either had small effects (e.g., Multhaup & Conner, 2002) or, paradoxically, actually enhance the occurrence of false memories in the DRM paradigm (Hicks & R.L. Marsh, 2001). As a final piece of evidence supporting the robustness of the DRM illusion, we point to its persistence on much-simplified tests—even an immediate one-item recognition test after a warning about the illusion (McDermott & Roediger, 1998).

On the other hand, there are a number of situations in which the illusion is less persistent than in the original Roediger and McDermott (1995) paper. The illusion is reduced following multiple study–test trials (Kensinger & Schacter, 1999; McDermott, 1996), when the studied items are made to be distinctive (Dodson & Schacter, 2001; Israel & Schacter, 1997; Kellogg, 2000; Schacter, Israel, & Racine, 1999), and when subjects are given more time to study the lists (Gallo & Roediger, 2002; McDermott & Watson, 2001). While multiple study trials and increased study time presumably increase the activation of the critical lure, they also allow the subject to encode item-specific information that can later be used to discriminate true from false memories. A similar argument holds for distinctive items; studying pictures or pronouncing aloud the study words leads to the encoding of item-specific information that later aids in the rejection of lures. Thus, subjects' judgements are not based solely on activation levels but also involve the use of strategies at test to discriminate true and false memories (see Gallo, McDermott, Percer, & Roediger, 2001; McDermott & Watson, 2001; Schacter et al., 1999; Smith & Hunt, 1998).

It may be that when subjects can successfully monitor for distinctive information, we will not see effects of activation manipulations at test, as in the current experiments. A recent experiment by Benjamin (2001) supports this hypothesis. His subjects were able to take advantage of repeated presentations of items and reduce false memories, presumably by monitoring for item-specific information. However, when the test discouraged the use of monitoring strategies (i.e., a speeded test), multiple study trials actually led to more false memories. When under time pressure, subjects were forced to rely solely on the activation or familiarity of items. Let us now apply the Benjamin (2001) results to our experiments: we tested young subjects under no time pressure, potentially allowing them to monitor activation well enough that their performance was not affected by our test manipulations. Thus, although increases in activation at test may increase the probability of false memories, this effect can be overridden if the subjects monitor for item-specific information. Further research may show an effect of test manipulations such as ours when subjects are given a response deadline (or have attention divided) during the test or when response times are measured, but under standard conditions of testing in the DRM paradigm, presenting related items before testing of the critical item has remarkably little effect.

Manuscript received 15 October 2001
Manuscript accepted 10 July 2002
PrEview proof published online 25 July 2003

REFERENCES

Balota, D.A., Cortese, M.J., Ducheck, J.M., Adams, D., Roediger, H.L. III, McDermott, K.B., & Yerys, B.E. (1999). Veridical and false memories in healthy older adults and in dementia of the Alzheimer's type. *Cognitive Neuropsychology*, *16*, 361–384.

Benjamin, A.S. (2001). On the dual effects of repetition on false recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *27*, 941–947.

Christiaansen, R.E., & Ochalek, K. (1983). Editing misleading information from memory: Evidence for the coexistence of original and post-event information. *Memory & Cognition*, *11*, 467–475.

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.

Dodson, C.S., & Schacter, D.L. (2001). ''If I had said it I would have remembered it'': Reducing false memories with a distinctiveness heuristic. *Psychonomic Bulletin & Review*, *8*, 155–161.

Fernández, A., Diez, E., Alonso, M.A., & Beato, M.S. (2001). *False recognition induced by implicit associative responses at the time of the test*. Paper presented at the Third International Conference on Memory, Valencia, Spain.

Gallo, D.A., McDermott, K.B., Percer, J.M., & Roediger, H.L. III (2001). Modality effects in false recall and false recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *27*, 339–353.

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 (2002). Variability among word lists in eliciting memory illusions: Evidence for associative activation and monitoring. *Journal of Memory & Language*, *47*, 469–497.

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.

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

Hicks, J.L., & Marsh, R.L. (2001). False recognition occurs more frequently during source identification than during old–new recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *27*, 375–383.

Israel, L., & Schacter, D.L. (1997). Pictorial encoding reduces false recognition of semantic associates. *Psychonomic Bulletin & Review*, *4*, 577–581.

Johnson, M.K., Hashtroudi, S., & Lindsay, D.S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3–28.

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

Kellogg, R.T. (2000). Presentation modality and mode of recall in verbal false memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *27*, 913–919.

Kensinger, E.A., & Schacter, D.L. (1999). When true memories suppress false memories: Effects of ageing. *Cognitive Neuropsychology*, *16*, 399–415.

Kimball, D.R., & Bjork, R.A. (2002). The influence of intentional and unintentional forgetting on false memories. *Journal of Experimental Psychology: General*, *131*, 116–130.

Lindsay, D.S., & Johnson, M.K. (1989). The eyewitness suggestibility effect and memory for source. *Memory & Cognition*, *17*, 349–358.

McDermott, K.B. (1996). The persistence of false memories in list recall. *Journal of Memory & 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. 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.

McDermott, K.B., & Watson, J.M. (2001). The rise and fall of false recall: The impact of presentation duration. *Journal of Memory & Language*, *45*, 160–176.

McEvoy, C.L., Nelson, D.L., & Komatsu, T. (1999). What is the connection between true and false memories? The differential roles of interitem associations in recall and recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *25*, 1177–1194.

McKone, E., & Murphy, B. (2000). Implicit false memory: Effects of modality and multiple study presentations on long-lived semantic priming. *Journal of Memory & Language*, *43*, 89–109.

Mueller, C.W., & Watkins, M.J. (1977). Inhibition from part-set cuing: A cue-overload interpretation. *Journal of Verbal Learning & Verbal Behavior*, *16*, 699–709.

Multhaup, K.S. (1995). Aging, source, and decision criteria: When false fame errors do and do not occur. *Psychology & Aging*, *10*, 492–497.

Multhaup, K.S., & Conner, C.A. (2002). The effects of considering nonlist sources on the Deese-Roediger-McDermott memory illusion. *Journal of Memory and Language*, *47*, 214–228.

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.

Neuschatz, J.S., Payne, D.G., Lampinen, J.M., & Toglia, M.P. (2001). Assessing the effectiveness of warnings and the phenomenological characteristics of false memories. *Memory*, *9*, 53–71.

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 & 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.

Reysen, M.B., & Nairne, J.S. (2002). Part-set cuing of false memories. *Psychonomic Bulletin & Review*, *9*, 389–393.

Robinson, K.J., & Roediger, H.L. III (1997). Associative processes in false recall and false recognition. *Psychological Science*, *8*, 231–237.

Roediger, H.L. III (1973). Inhibition in recall from cueing with recall targets. *Journal of Verbal Learning & Verbal Behavior*, *12*, 644–657.

Roediger, H.L. III (1978). Recall as a self-limiting process. *Memory & Cognition*, *6*, 54–63.

Roediger, H.L. III, Balota, D.A., & Watson, J.M. (2001a). Spreading activation and the arousal of false memories. In H.L. Roediger III, J.S. Nairne, I. Neath, & A.M. 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., & Robinson, K.J. (1998). The role of associative processes in creating false memories. In M.A. Conway, S.E. Gathercole, & C. Cornoldi (Eds.), *Theories of memory II* (pp. 187–245). Hove, UK: Psychology Press.

Roediger, H.L. III, Watson, J.M., McDermott, K.B., & Gallo, D.A. (2001b). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review*, *8*, 385–407.

Rundus, D. (1973). Negative effects of using list items as recall cues. *Journal of Verbal Learning & Verbal Behavior*, *12*, 43–50.

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.

Schacter, D.L., Israel, L., & Racine, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory & Language*, *40*, 1–24.

Slamecka, N.J. (1968). An examination of trace storage in free recall. *Journal of Experimental Psychology*, *76*, 504–513.

Smith, R.E., & Hunt, R.R. (1998). Presentation modality affects false memory. *Psychonomic Bulletin & Review*, *5*, 710–715.

Stadler, M.A., Roediger, H.L. III, & McDermott, K.B. (1999). Norms for word lists that create false memories. *Memory & Cognition*, *27*, 494–500.

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning & Verbal Behavior*, *5*, 381–391.

Watkins, M.J. (1975). Inhibition in recall with extralist ‘‘cues’’. *Journal of Verbal Learning & Verbal Behavior*, *14*, 294–303.

Watson, J.M., Balota, D.A., & Sergent-Marshall, S.D. (2001). Semantic, phonological, and hybrid veridical and false memories in healthy older adults and in individuals with dementia of the Alzheimer's type. *Neuropsychology*, *15*, 254–268.

Wright, D.B. (1993). Misinformation and warnings in eyewitness testimony: A new testing procedure to differentiate explanations. *Memory*, *1*, 153–166.