Memory & Cognition
2006, 34 (5), 1026-1036

The role of test structure in creating false memories

JENNIFER H. COANE and DAWN M. McBRIDE
Illinois State University, Normal, Illinois

In the Deese/Roediger–McDermott (DRM) paradigm, studying lists of semantic associates results in high rates of false recognition of a nonpresented critical word. The present set of experiments was designed to measure the contribution of additional processing of list items at test to this false memory effect. The participants studied sets of lists and then performed a recognition task for each set. In three experiments, using this paradigm, we investigated false recognition when the number of studied list items presented at test (0, 6, or 12) was manipulated. In Experiments 2 and 3, false recognition of critical lures associated to both studied and nonstudied lists increased significantly as the number of list items included in the test increased. These results indicate that processes occurring at retrieval contribute to false memory effects found with the DRM paradigm.

In recent years, the study of experimentally induced false memory phenomena has increased substantially. One method with which false memories are studied involves using word list paradigms that are based on properties of semantic association and that reliably reveal robust false memory effects. Many recent studies have utilized the Deese/Roediger–McDermott (DRM) paradigm (Roediger & McDermott, 1995), in which participants study lists of semantic associates (e.g., *bed*, *rest*, *awake*, *drowsy*, etc.) of a nonpresented critical lure (e.g., *sleep*). During subsequent free recall or recognition tasks, participants tend to remember the critical lure at rates comparable to the studied items (Roediger & McDermott, 1995). Furthermore, participants often report similar phenomenological experiences (e.g., a vivid memory of having studied the item) for both the list items and the critical lures (Roediger & McDermott, 1995). In other words, participants are able to "retrieve" the experience of encoding the critical item, when in fact it was not studied.

The false memory effect appears to be quite robust, and it occurs reliably across a variety of experimental manipulations. Indeed, some factors have been identified that appear to increase the strength of the illusion. Roediger and McDermott (1995) found higher rates of false recognition of critical lures associated to lists that had been previously recalled than of lures associated to lists that had not been recalled, indicating that repeated retrieval attempts might enhance the false memory effect. In addition, several studies that examined the effects of longer retention intervals on accurate and false memory have found that false memories actually increase over time, whereas accurate memories decrease (McDermott, 1996; Thapar & McDermott, 2001; Toglia, Neuschatz, & Goodwin, 1999). Seamon et al. (2002), however, found that false memories did not increase over time; rather, they were more resistant to decay than accurate memories. Thus, it appears that false memory for the critical lure in the DRM paradigm can be enhanced or at least maintained by testing factors such as the number of tests.

Evidence also suggests that retrieval processes may play a significant role in creating false memories for the critical lure. In the original Roediger and McDermott (1995) study, an output serial position analysis of the free recall data indicated that participants tended to falsely recall the critical lure toward the end of the recall session. This may indicate that prior recall of list items served as a cue for the lure, or that recalling the list items increased the probability that participants would recall the lure as an item highly associated to all studied items. In recognition tasks, one or more list items are often presented before the lure, which may have thus contributed to additional priming of the lure (e.g., Roediger & McDermott, 1995).

One explanation of the effectiveness of the DRM paradigm in creating false memories is semantic activation. According to activation theories, words are linked to one another in a network, and the activation of one lexical concept results in the spread of activation to surrounding concepts (Collins & Loftus, 1975). Studying a list of semantically related items will thus result in strong activation of an item (i.e., the critical lure) associated to all list items. Consequently, the critical lure may be falsely remembered

Portions of the present study were conducted in partial fulfillment of the requirements for a master's degree for the first author, who is presently at Washington University in St. Louis. We thank Tara Babich, Evelyn Beasley-Scott, Bascom Raulerson, Lindsey Rhodes, Brian Sutter, and Chris Wahlheim for their assistance with data collection. We also thank Jason Arndt, Dave McCabe, Jeffrey Anastasi, and three anonymous reviewers for comments on earlier versions of this manuscript.
Correspondence should be addressed to D. M. McBride, Department of Psychology, Illinois State University, Campus Box 4620, Normal, IL 61790-4620 (e-mail: dmcbride@ilstu.edu).

Note—This article was accepted by the previous editorial team, when Colin M. MacLeod was Editor.

Copyright 2006 Psychonomic Society, Inc.

due to the heightened activation. It is generally understood that this activation process occurs automatically and is therefore fast acting, obligatory, and not under conscious control (McDermott & Watson, 2001). Automatic processes are assumed to occur with the simple presentation of the appropriate stimulus; hence, the very structure of the lists of semantic associates is conducive to the activation of the critical lure. When lists of 12 or 15 semantic associates are presented in blocked order, the activation converging upon the critical lure is strong enough to elicit very high rates of false recognition. Currently, theorists propose that activation processes that occur primarily at study combine with source monitoring errors that occur at test to elicit high rates of false memory (McDermott & Watson, 2001); monitoring processes may occur as participants attempt to decide whether an item was studied or thought of, and errors in these processes could result in false alarms to nonstudied items such as the lure.

Evidence from the literature suggests that activation from multiple associates summates; exposure to increasing numbers of related concepts results in stronger activation within the network (Robinson & Roediger, 1997; Roediger, Balota, & Watson, 2001). Furthermore, deeper processing can enhance rates of true and false recall, possibly because more meaningful encoding strengthens the associations between items (Toglia et al., 1999). It is generally assumed that most monitoring processes occur during the test, whereas activation is generally attributed to the encoding phase.

Activation probably occurs both during study and during test, when the associates to the lure are either represented in a recognition test or recalled by participants in a free recall test. The majority of studies have investigated how encoding manipulations contributed to false memory, whereas studies of retrieval manipulations are relatively rare. Although it is not explicitly stated in the literature, the assumption is that activation does not occur at test, but to date very little evidence has been presented that supports this assumption. Furthermore, semantic activation tends to be relatively short lived and to decay across intervening items. Thus, the assumption that activation at encoding alone drives the effects may not be the most parsimonious explanation; rather, it makes sense to assume that the semantic networks are reactivated at test and that, therefore, testing contributes significantly to the memory errors observed in the DRM paradigm.

However, empirical attempts to determine the contributions of retrieval processes in the creation of false memories have yet to provide clear answers. Anastasi, Avery, Sinclair, Weitz, and Rhodes (2003) and Marsh, McDermott, and Roediger (2004) attempted to determine whether manipulating the number of list items presented prior to the lure would affect the likelihood of false recognition, under the assumption that the processes that result in false memory for the lure occur both at study and at test. Accordingly, these researchers manipulated the number of list items presented before the lure, expecting to find different rates of false recognition of the lure under different conditions. For example, in Experiment 1 of Marsh et al.'s study, the number of items presented on a recognition test was manipulated in such a way that the critical lure was presented after 0, 3, or 6 list items. No difference in false memory was found with this testing manipulation. Anastasi et al. used a similar methodology in three of their experiments, presenting 2, 4, or 8 studied list items before the critical lure in the recognition test. They also found that the number of studied items presented before the lure made no difference in critical lure recognition.

More recently, Dodd, Sheard, and MacLeod (2006) also attempted to find evidence for activation of the lure at test by manipulating the number of list items tested prior to the lure. In two experiments using auditory presentation of each list followed by immediate testing, they failed to find any differences in false alarm rates with 0 to 5 list items tested before the lure. In their second experiment, Dodd et al. gave participants a response deadline of 750 msec to encourage them to rely on familiarity more than on recollection, hypothesizing that this manipulation would increase the likelihood of detecting an effect. Dodd et al. failed to detect any differences in false recognition rates, and they concluded that retrieval factors did not contribute to the high false alarm rates observed in the DRM paradigm. However, Diez, Fernandez, and Alonso (2004), who also manipulated the number of list items preceding the lure on a recognition test and implemented a speeded response task, did find significant differences in false recognition when participants were instructed to respond after 250 or 750 msec. False alarms were higher under speeded instructions when 4 list items preceded the lure than when no items preceded the lure on the test. However, when subjects were given 1,500 msec to respond, the number of list items preceding the lure had no effect on false recognition rates. Thus, the role of testing in the creation of false memories remains inconclusive.

One explanation for the absence of a detectable effect can be found in the phenomenon itself: False memory has consistently been found to be extremely resistant to extinction, and even giving explicit warnings to participants about the nature of the paradigm fails to eliminate the effect (e.g., McCabe & Smith, 2002). Encoding processes that resulted in participants' mistakenly identifying the lure as a studied item might have been strong enough that additional effects were simply not detectable with the methods used by Marsh et al. (2004) and Anastasi et al. (2003). Indeed, as reported by Gallo, Roediger, and McDermott (2001), warning participants about the paradigm was most effective when the warning was issued *before* study. When subjects were warned *between* study and test, the warning had little effect, thus indicating that the lure may have been encoded similarly to the studied items.

Although converging evidence appears to indicate that additional processing of list items at test has no effect on false memory rates (Anastasi et al., 2003; Dodd et al., 2006; Marsh et al., 2004), there are theoretical reasons to assume that testing should influence memory performance. If false memories in the DRM paradigm are indeed attributable to automatic spreading activation, then it is not unreasonable to assume that activation does occur in the testing phase and that this activation might summate with the activation that occurs in the encoding phase. One possibility is that the activation at test is similar in nature to the activation at encoding, and that each additional phase of activation further increases the probability that the critical lure will be falsely recognized. Furthermore, recent data from a neuroimaging study indicate that foils on a recognition test that are later remembered in a surprise test show the same patterns of activation as do items encoded under intentional encoding instructions (Buckner, Wheeler, & Sheridan, 2001). Thus, it appears that a recognition test can be considered an opportunity for encoding.

Indeed, evidence that activation that occurs in the testing phase contributes to false recognition rates is provided by false alarm rates for critical lures from nonstudied lists in the Marsh et al. (2004) study. Marsh et al. included nonstudied lists and lures with unrelated items as fillers in the recognition test. Critical lures associated to nonstudied lists were falsely recognized 31% of the time when they were not preceded by any list items in the recognition test. However, when 3 or 6 items from nonstudied lists preceded the presentation of the critical lures, false recognition of the nonstudied lures increased to 49%. Thus, it would appear that when study effects are absent, simply presenting as few as 3 semantic associates as filler items in a recognition test can increase false recognition rates of critical lures. The fact that participants, after viewing as few as 3 list items, consistently made false recognitions of critical lures associated to lists they had never studied indicates that some activation resulting in false memory likely occurred at test. Therefore, the high rate of false recognition of critical lures from nonstudied lists may be a good indicator of the amount of activation that occurs at test. If indeed activation does occur at test, source monitoring errors and encoding effects may not be the only factors contributing to the false memory effect at the time of test.

The present studies were designed to investigate the extent to which activation at retrieval might contribute to the creation of false memories in the DRM paradigm. The present experiments were designed using methodology similar to that employed by Marsh et al. (2004) and Anastasi et al. (2003), but an attempt was made to increase the strength of the testing manipulation and to decrease the activation of the critical lure at study to make testing effects more detectable. In three experiments, the number of list items included in the recognition test was manipulated. Participants were presented 0, 6, or 12 list items in the recognition test prior to the lure presentation to increase the effect of the manipulation. The hypothesis was that as more list items were presented in the recognition test prior to presentation of the critical lure, the probability that the lure would be called *old* would also increase.

More specifically, Experiment 1 replicated the Marsh et al. (2004) study with a different manipulation of the number of studied items tested prior to presentation of the critical lure. Rather than presenting 0, 3, or 6 items prior to the lure, we presented 0, 6, or 12, hypothesizing that by increasing the opportunities for activation at test to converge on the lure, a larger effect would be detected. Experiment 2 was designed to address the possibility that effects of the testing manipulation were absent in past studies because of the strength of lure priming at study. To decrease the strength of the study effects and to increase the likelihood of detecting testing effects for studied lists, the lists were presented in random order, three at a time, with items from three lists intermixed. This study manipulation has been found to weaken the rate of false alarms in free recall tasks (McDermott, 1996; Toglia et al., 1999) and should allow effects of retrieval processes to be more easily detected if they occur. Indeed, previous attempts to investigate activation at test (e.g., Anastasi et al., 2003; Marsh et al., 2004) may have failed to detect any significant effects because of the strength of the activation that occurred during encoding. Presenting the lists in random order should have sufficiently weakened the effects that occurred at study to allow detection of those that might have occurred at test.

In addition, since several prior studies have reported high rates of false recognition for critical lures associated to nonstudied lists (e.g., Marsh et al., 2004; McCabe & Smith, 2002; McDermott, 1996; Seamon et al., 2002), 0, 6, or 12 items from nonstudied lists were included on the recognition test to determine whether a significant change in rates of false recognition of the associated lures occurred. Such a finding would provide evidence that participants could be induced to call old an item that had been presented only at test—that is, an item that they had not studied and for which no associates had been studied—which would thus provide very strong support for the hypothesis that false memories can arise solely as a function of testing conditions. Finally, Experiment 3 tested the differences between blocked and random presentation as a between-subjects variable.

Another difference between the present studies and the Marsh et al. (2004) study is worth highlighting. Marsh et al. presented 18 lists of 15 items each and administered a single recognition test. In the present experiments, the lists were studied in blocks of three and participants were tested after each block. This was done to avoid the high noncritical intrusion rates (27%) found by Marsh et al. and to reduce the long delays between study and test phases. It was hypothesized that with shorter study lists and tests, participants would be more accurate and therefore a stronger estimate of any observed effects would be obtained.

In summary, it was expected that different testing conditions would yield different rates of false recognition. It was hypothesized that as more list items were included in a recognition test, the rates of false recognition would increase because list items would serve as cues for critical lures. Because the critical lures were presented only after all the associated list items had been presented, any differences in false alarms among the three testing conditions could be attributed to this manipulation. In Experiment 1, in a recognition test with blocked study presentation, we manipulated the number of list items presented at test prior to presentation of the lure. In Experiment 2, we manipulated the order of items presented at study in hopes of reducing the activation occurring during encoding, and in Experiment 3 we directly compared the effects of study order. Furthermore, in Experiments 2 and 3, we included nonstudied list items and the associated critical lures in the recognition tests to assess whether we could create false memories solely on the basis of processes occurring during retrieval. Finding increases in false alarms to the critical lures associated to nonstudied lists would provide compelling evidence of false recognition due to testing effects.

EXPERIMENT 1

Method

Participants. Forty-two undergraduate students were recruited from Illinois State University; they received extra credit in their classes. All of the participants were native speakers of English.

Design and Materials. A 3 (item type: list, lure, unrelated) × 3 (number of list items presented at test: 0, 6, 12) completely within-subjects factorial design was used. Twenty-four study lists of 12 items each from the original Roediger and McDermott (1995) study were used. The lists consisted of the 12 highest semantic associates to a nonpresented critical lure, presented at study in decreasing order of associative strength. Thus, the strongest associates appeared in the highest serial order positions. The lists were randomly divided into eight blocks of three lists each. For counterbalancing purposes, three different sets of blocks were created, so that each list appeared in each of the three testing conditions (0, 6, or 12 items presented at test) an equal number of times throughout the study. Within each block, the lists were presented in blocked format for study, such that all the items from one list were in order. No cues separated the three lists within a block. At the beginning and end of each study block, two primacy and recency filler items were presented as buffers.

The recognition test for each block consisted of 45 items. Twenty-two of these items were studied items: 12 from one list, 6 randomly selected from a second list, and the 4 primacy and recency items. The remaining 23 items consisted of the 3 critical lures and 20 nonstudied, unrelated filler items. The critical lures were always presented among the last 9 items on the recognition test, after the list items. Other than this constraint, the order in which items appeared at test was randomized across the participants. Both study and test were administered by computer.

Procedure. The participants were tested individually. They were seated in front of a Macintosh computer, and the experiment was administered via PsyScope software (Cohen, MacWhinney, Flatt, & Provost, 1993). The participants were told that they would be presented lists of words, with the words presented one at a time, and that they should pay close attention to each one, because their memory would be tested. The words were presented in white type on a black screen. A 36-point Times New Roman font was used to present the stimuli. Each word remained on the screen for 3,000 msec, and the interstimulus interval was 500 msec. Lists were presented in blocked order in blocks of three lists; however, no cues indicated a separation between lists. Following each block of three lists, the participants completed a filler task for 30 sec, consisting of mathematical problems contained in a booklet provided by the experimenter. After 30 sec, a beep notified the participants to begin the recognition test on the computer. The instructions on the screen directed the participants to decide whether or not each word that appeared on the screen had been included in the list that they had most recently studied. If they were sure that they had studied a word, they were instructed to press "O" on the keyboard. If they were sure that a word was new or if they could not remember it, they were instructed to press "W" on the keyboard. These keys were selected because they are on the same row on the keyboard, which maximizes comfort, but are sufficiently far apart to avoid confusion. A cue card was attached to the monitor to serve as a reminder. Each word remained on the screen until the participant made a decision.[1] At the end of the recognition test, a prompt appeared on the screen to signal the participants to begin the next part of the experiment, and the next block of three lists was presented. At the end of the experiment, the participants were debriefed. The entire experiment lasted approximately 40 min.

Results and Discussion

Alpha was set at .05 for all analyses. Mean rates of true and false recognition are displayed in Table 1. We will first present results for the false alarms, and then a comparison of veridical and false memory. False recognition data were first analyzed in a one-way repeated measures ANOVA with the number of items presented at test as the factor. No significant differences were observed in rates of false recognition, regardless of the number of list items included in the test [$F(2,82) < 1.0$, $MS_e = .02$, $p = .54$].

A separate 2 (item type: list vs. lure) × 2 (number of items presented at test: 6 vs. 12) repeated measures ANOVA was conducted on the recognition data to compare true and false memory. List items were recognized significantly more often than critical lures [$F(1,41) = 7.22$, $MS_e = .04$, $p = .01$]. Neither the main effect of the number of items presented at test [$F(1,41) < 1.0$, $MS_e = .0003$, $p = .88$] nor the interaction between item type and the number of items presented at test [$F(1,41) < 1.0$, $MS_e = .01$, $p = .45$] was found to be significant, indicating that the number of list items included in the recognition test failed to affect significantly both true and false memory. Accurate recognition of studied items was high overall ($M = .79$), regardless of whether 6 or 12 items were included in the test. False recognition rates did not differ according to the number of list items—6 or 12—presented at test [$t(41) < 1.0$, $p = .63$]. These results are consistent with those reported by Marsh et al. (2004) and Anastasi et al. (2003).

Overall rates of false alarms to noncritical intrusions (unrelated fillers) were quite low ($M = .08$, $SE = .01$), indicating that false alarm rates for critical lures were not due to guessing.

The high rates of false recognition observed in Experiment 1 indicate once more the effectiveness of the DRM paradigm in eliciting false memories. The fact that high rates of false recognition were observed even when no list items were included in the test ($M = .67$) underscores the robustness of the paradigm and indicates that processes that occur at study are very strong and may mask any additional

Table 1
Mean Recognition As a Function of Item Type and Number of Items Presented at Test in Experiment 1

No. Items at Test	List Items		Unrelated Fillers		Critical Lures	
	M	*SE*	*M*	*SE*	*M*	*SE*
0	–				.67	.03
6	.80	.02			.70	.04
12	.79	.02			.71	.04
			.08	.01		

effects that take place during retrieval processes. To test this hypothesis, Experiment 2 was designed with a different study presentation intended to lower the rate of false recognition occurring as a function of encoding processes.

EXPERIMENT 2

In Experiment 2, two major modifications to the design were implemented to further examine the contribution of retrieval processes in false recognition. Because the results obtained in Experiment 1 and those obtained by Marsh et al. (2004) and Anastasi et al. (2003) did not show any differences in rates of false recognition as a function of the number of list items presented at test, Experiment 2 included a manipulation known to lower rates of false memory. Presenting the lists in random as opposed to blocked order has been found to reduce rates of false memory in prior studies (e.g., McDermott, 1996; Toglia et al., 1999). Thus in Experiment 2, list items were randomly intermixed within the study lists. The second modification to the design consisted of including nonstudied lists as fillers in the recognition test. Rates of false recognition of critical lures associated to nonstudied lists that differ as a function of the number of list items presented at test would clearly indicate the level at which retrieval processes are operating. Furthermore, such a finding could be attributed only to activation processes occurring at test, since none of the items were included on the study list, and would thus provide the strongest evidence for false memories due to testing effects.

Method

Participants. Sixty-four participants were recruited from Illinois State University; they received extra credit in their classes. All participants were native speakers of English.

Design and Materials. A 3 (item type: list, lure, unrelated) × 2 (list type: studied, nonstudied) × 3 (number of list items presented at test: 0, 6, 12) completely within-subjects factorial design was used. Thirty-six word lists of 12 items each from the Stadler, Roediger, and McDermott (1999) list norms were used. Twenty-four of the lists were the same as those used in Experiment 1; the remaining 12 lists were added so that the more complex design would have sufficient data points per condition. Overall, the added 12 lists do not differ from the original 24 lists (Stadler et al., 1999). The lists were divided into 6 blocks of 6 lists each. Within each block, the words from 3 studied lists were presented in random order. Four random assignments of lists to conditions were created in four separate running programs. Each participant was randomly assigned to one of these programs. Before and after each block, two filler items were presented as buffers.

The recognition test for each block consisted of 50 items presented in a different random order for each participant with the restriction that the critical lures were always presented after any associated list items in the test. Of the associated list items, 18 were studied items: 12 from one studied list and 6 randomly selected items from a second studied list. In addition, 12 words from one nonstudied list and 6 randomly selected items from a second nonstudied list were tested. The mean backward associative strength (BAS) of the list items for the 6 and 12 items at test conditions for both studied and nonstudied lists ranged from .18 to .21, and did not differ reliably for each condition [$F(3,69) < 1$]. Thus, any differences among conditions could not be attributed to differing levels of mean BAS.

All 6 critical lures (3 from studied lists, 3 from nonstudied lists) were tested among the last 14 items for each set, together with the 4 primacy and recency items and 4 unrelated fillers. Thus, the critical lures always appeared after all the associates had been presented.

Procedure. The procedure was the same as that followed in Experiment 1.[2] The only difference was the duration of the experiment, which lasted approximately 30 min.

Results and Discussion

Alpha was set at .05 for all analyses. Mean rates of true and false recognition are presented in Table 2. False recognition of critical lures was first analyzed in a 2 (list type: studied vs. nonstudied) × 3 (number of items presented at test: 0, 6, or 12) repeated measures ANOVA. A main effect of list type was found, indicating that, overall, false recognition rates for lures were significantly higher in studied than in nonstudied lists [$F(1,63) = 215.9$, $MS_e = .08$, $p < .0005$]. This result indicates that participants were more likely to call a critical lure old after the associates had been studied.

A significant main effect of the number of list items presented at test was found [$F(2,62) = 18.03$, $MS_e = .03$, $p < .001$], indicating that for both studied and nonstudied lists, false recognition of critical lures was affected by the number of list items presented at test. Post hoc analyses conducted for studied lists indicated that critical lures were falsely recognized more often when 6 list items were presented at test than when 0 list items were presented [$t(63) = -2.1$, $p = .041$]. When 12 items were presented at test, false recognition rates also increased in comparison with when no items were presented [$t(63) = -3.3$, $p = .002$]. For nonstudied lists, a similar pattern emerged, with higher rates of false recognition in the 6 items at test condition than in the 0 items at test condition [$t(63) = -5.0$, $p < .001$]. Higher rates of false recognition of critical lures were also observed in the 12 items at test condition than in the 0 items at test condition [$t(63) = -4.8$, $p < .001$]. In both studied and nonstudied lists, no differences were found between the 6 and 12 items at test conditions (both ps $> .05$). Thus, it appears that when the study effects were weakened (as for studied lists in Experiment 2) or were absent (as for nonstudied lists), false recognition of critical lures did increase significantly when more list items were presented at test (for both studied and

Table 2
Mean Recognition As a Function of List Type, Item Type, and Number of Items Presented at Test in Experiment 2

		List Items		Unrelated Fillers		Critical Lures	
List Type	No. Items at Test	*M*	*SE*	*M*	*SE*	*M*	*SE*
Studied	0	–				.57	.03
	6	.82	.02			.64	.04
	12	.80	.02			.68	.03
Nonstudied	0	–				.11	.02
	6	.08	.01			.25	.03
	12	.09	.01			.24	.03
				.08	.01		

nonstudied lists), providing evidence for the contribution of activation at test to the false memory effect.

It is possible, however, that over the course of repeated study–test trials, participants might have simply become more liberal in accepting nonstudied items as old, due to fatigue or other factors. To rule out the hypothesis that participants were simply adopting a more liberal criterion over the repeated study–test blocks, a false alarm × block analysis was performed. No significant differences emerged between the number of false alarms occurring in each block [$\chi^2(5) = 2.497$, $p = .777$], indicating that participants were not endorsing more critical lures as old across test blocks. Finally, the interaction between list type and number of items presented at test was not significant [$F(2,126) = 1.12$, $MS_e = .03$, $p = .35$], indicating that false recognition of critical lures associated to studied and nonstudied lists increased, at similar rates, as the number of list items presented at test increased.

Overall, false alarms to unrelated fillers were low [$M = .08$, $SE = .01$], indicating that participants were able to discriminate between old and new items and were not simply guessing when responding to critical lures. Paired samples *t* tests indicated that unrelated fillers were called old significantly less often than critical lures from studied lists [$t(63) = -20.2$, $p < .001$], or than critical lures from nonstudied lists [$t(63) = -7.1$, $p < .001$].

True and false recognition data were analyzed using a 2 (list type: studied vs. nonstudied) × 2 (item type: list vs. lure) × 2 (number of items presented at test: 6 vs. 12) repeated measures ANOVA. Overall, studied items were called old more often than nonstudied items [$F(1,63) = 683.9$, $MS_e = .06$, $p < .001$]. In addition, no significant differences were observed in terms of item type and number of list items included in the test (both *p*s greater than .05). However, a significant interaction between list type and item type was observed [$F(1,63) = 81.14$, $MS_e = .04$, $p < .001$]. For the studied lists, list items were recognized more often than critical lures, but for the nonstudied lists the opposite pattern occurred, with critical lures being falsely recognized more often than the associated list items. The fact that critical lures associated to nonstudied lists were recognized significantly more often [$t(63) = 7.71$, $p < .001$] than the related list items indicates that this increase in false recognition is due to priming of the lure at test and not to participants' guessing. The three-way interaction of list type, item type, and number of items presented at test was not significant. In conclusion, data from Experiment 2 indicate that retrieval processes do contribute to the robust findings of false memory in the DRM paradigm.

EXPERIMENT 3

The methodological differences between Experiments 1 and 2 made it difficult to draw direct comparisons between the effects of blocked and random presentation of the lists during the study session. Although it appeared that random presentation was effective in reducing the activation occurring at test, Experiments 1 and 2 also differed in terms of the type of foils presented at test. In Experiment 1, unrelated words were used as foils, whereas in Experiment 2, nonstudied DRM lists were used. In Experiment 3, participants studied blocks of three lists in either blocked or random order, and the test included items from studied and nonstudied lists.

Method

Participants. One hundred five participants were recruited from Illinois State University's psychology department subject pool. Fifty-two participants were randomly assigned to the blocked study condition and 53 to the random study condition. Data from 1 additional participant were omitted due to equipment failure. All of the participants were native English speakers and received course credit for their participation.

Design. The study was a 2 (order of presentation at study: blocked, random) × 2 (list type: studied, nonstudied) × 3 (item type: list, critical lure, unrelated) × 3 (number of list items presented at test: 0, 6, 12) design. Order of presentation was manipulated between subjects; all other variables were within subjects.

Materials and Procedure. The materials from Experiment 2 were used, except that half of the participants were randomly assigned to the conditions in which lists were blocked at study and the other half studied lists in random order. In order to avoid potential confounds of different lists being in different conditions, the assignment of lists to blocks in the two conditions was the same. In other words, the only difference between the blocked and random conditions was the order in which items were presented at study. The procedure was identical to that followed in Experiment 2.

Results and Discussion

The effects of list presentation type (blocked or random order) were directly tested in Experiment 3. Overall, the same pattern of results as that observed in Experiment 2 was found. Significant main effects of the number of items presented at test were found in both blocked and random study conditions. Means and standard deviations are presented in Table 3. We present false alarm data first, followed by a comparison of veridical and false recognition. An ANOVA of the critical lure data with presentation order as a between-subjects factor revealed a significant effect of list type, such that lures associated to studied lists were more likely to be falsely remembered than lures associated to nonstudied lists [$F(1,103) = 650.99$, $MS_e = .06$, $p < .001$]. A significant effect of number indicated that, overall, false alarms increased as a function of the number of items presented at test [$F(2,206) = 18.68$, $MS_e = .02$, $p < .001$]. The increase in false alarm rates between the 0 and 6 items at test conditions was significant for both studied and nonstudied lists [$t(104) = 3.62$, $p < .001$, and $t(104) = 4.53$, $p < .001$, respectively], as was the increase between the 0 and 12 items at test conditions [$t(104) = 3.23$, $p = .002$, and $t(104) = 3.68$, $p < .001$, for studied and nonstudied lists, respectively]. False alarms to critical lures did not differ between 6 items at test and 12 items at test conditions ($p = .84$ and $p = .50$ for studied and nonstudied lists, respectively). The effect of order was marginally significant, such that overall false recognition was higher when lists were blocked at study [$F(1,103) = 3.16$, $MS_e = .11$, $p = .078$].

Finally, a significant list type × order interaction was found [$F(1,103) = 5.49$, $MS_e = .06$, $p = .021$], indicat-

Table 3
Means and Standard Deviations As a Function of Study Condition, List Type, Item Type, and Number of Items Presented at Test in Experiment 3

List Type	No. Items at Test	List Items *M*	List Items *SD*	Unrelated Fillers *M*	Unrelated Fillers *SD*	Critical Lures *M*	Critical Lures *SD*
		Blocked Presentation					
Studied	0	–				.62	.24
	6	.83	.12			.74	.22
	12	.83	.15			.73	.23
Nonstudied	0	–				.10	.12
	6	.04	.05			.17	.19
	12	.06	.06			.16	.20
				.05	.01		
		Random Presentation					
Studied	0	–				.56	.26
	6	.80	.11			.63	.26
	12	.78	.12			.63	.27
Nonstudied	0	–				.08	.14
	6	.06	.20			.18	.20
	12	.08	.10			.17	.19
				.07	.01		

ing that, when lists were studied, overall false alarm rates were higher following blocked presentation than following random presentation [$t(103) = 2.42$, $p = .02$]; however, when lists were not studied, rates of false recognition were similar regardless of order of presentation during study [$t(103) = -0.02$, $p = .99$]. Thus, it appears that for studied lists, random presentation was effective in reducing false alarm rates (a finding consistent with the combined results from Experiments 1 and 2). No other effects were significant. Once again, the possibility that participants were adopting a more liberal response criterion across repeated study–test trials was tested. No differences were found as a function of block [$\chi^2(5) = 7.327$, $p = .197$].

A 2 (item type) × 2 (number of items presented at test) × 2 (list type) ANOVA was conducted to compare memory for list items and for critical lures. A significant effect of list type indicated that overall recognition (false and veridical) was higher when lists were studied [$F(1,103) = 1{,}933$, $MS_e = .04$, $p < .001$]. The list type × order of presentation interaction was also significant [$F(1,103) = 10.2$, $MS_e = .04$, $p = .002$]. When lists were studied, overall recognition rates were higher in the blocked order condition than in the random order condition; this difference, however, was driven primarily by the difference in false recognition rates between the two order conditions. For nonstudied lists, no differences emerged between blocked and random presentation of study items. Finally, a significant list type × item type interaction was found [$F(1,103) = 100.8$, $MS_e = .04$, $p < .001$]. For studied lists, veridical recognition was higher than false recognition; however, the opposite pattern occurred when the lists were not studied. No other effects were significant.

To conclude, the effects of blocked and random presentation were tested between subjects while other aspects of the design from Experiment 2 (e.g., the type of foils included on the recognition test, the number of blocks, etc.) remained constant. The main results can be summarized as follows: Overall, false recognition was higher when lists were studied in blocked order; rates of false recognition were affected by the number of items included on a test that preceded the critical lure, and activation at test occurred for both studied and nonstudied lists. The results indicate that false alarms to critical lures did indeed increase significantly as a function of the number of items presented at test following both blocked and random presentation of study lists. The results of Experiment 3 further indicated that testing effects based on the number of items presented prior to the critical lure could occur even after blocked presentation at study. Thus it appears that reducing the activation at test by presenting the lists in random order (as we did in Experiment 2) is not always necessary and that testing effects can occur under different conditions. The null results in Experiment 1, therefore, may have been due to factors other than the high rates of false recognition attributable to processes occurring at encoding. Additional explanations are presented in the General Discussion section below.

Furthermore, the results of Experiment 3 allowed us to rule out the hypothesis that the significant effects observed in Experiment 2 were simply due to criterion changes following random presentation. Mather, Henkel, and Johnson (1997) found higher rates of noncritical intrusions following random presentation of study lists; in the present study, no such differences were found. In fact, false alarms to unrelated fillers were very low in both blocked and random study conditions (*M*s = .05 and .07, respectively) and were not significantly different ($p = .32$).[3] The finding of significant increases in false alarm rates to lures associated to nonstudied lists observed in Experiment 2 was also replicated. In both study conditions, the critical lures from nonstudied lists were more likely to be falsely recognized after 6 or 12 items had preceded their appearance on the test.

GENERAL DISCUSSION

As in many past studies (e.g., Roediger & McDermott, 1995; Stadler et al., 1999), the patterns of results in three experiments provided evidence for high rates of false recognition of critical lures following study of lists of semantic associates. In the present study, however, the experiments were specifically designed to measure the contributions of retrieval processes to the false memory effect. The number of list items included in the recognition test was the main experimental manipulation in all experiments. In Experiment 1, no differences in rates of false recognition as a function of the number of list items included on the test were found. However, in Experiments 2 and 3, significant increases in false alarm rates were observed when 6 or 12 items preceded the critical lure. When 6 or 12 list items were included in the test, the critical lures associated to those lists were more often falsely recognized than when no list items were presented at test. Experiment 3 provided a direct test of whether order of presentation at study was a contributing factor in the observed differences. Effects of the number of list items presented at test were found after both blocked and random study. Furthermore, in Experiments 2 and 3, the observed increase in rates of false recognition of critical lures associated to nonstudied lists provided a good measure of testing effects. The high rates of veridical recognition in all experiments, combined with low rates of noncritical intrusions, indicated that participants' responses were not simply guesses.

Presentation of list items in random order in Experiment 2, in comparison with Experiment 1, appears to have been successful in reducing the overall rates of false recognition for studied lists. The manipulation appears to have been most effective in reducing false alarms to critical lures when no list items were included in the recognition test (.67 in Experiment 1 vs. .57 in Experiment 2). However, in Experiment 3, differences in false alarm rates as a function of the number of items presented at test were found following both blocked and random presentation. Therefore, the null results in Experiment 1 may have been due to factors other than order of presentation. One possible explanation of these results is that fewer participants (42) were tested in Experiment 1 (64 were tested in Experiment 2, and 52 in the blocked condition in Experiment 3), thus giving lower power to detect an effect. Observed power estimated from the data in Experiment 1 was .15. When estimated from the blocked condition of Experiment 3, power was greater than .80 for the difference between 0 and 6 items and between 0 and 12 items preceding the lure, indicating that low power may not be a sufficient explanation for the null results in Experiment 1. However, an important difference between Experiments 1 and 3 concerns the nature of the filler items in the recognition test. In Experiment 1, the fillers were all unrelated items, whereas in Experiment 3 the majority of the fillers were nonstudied DRM lists. It is possible that the difference in type of fillers could have affected participants' response bias. In fact, in Experiment 1, participants could endorse as old any item that was thematically related to the studied items and minimize their error rate (as the only error would be falsely recognizing the critical lure), whereas in Experiment 3, the relatedness of the fillers might have made the discrimination more difficult, especially as the test progressed (see Tun, Wingfield, Rosen, & Blanchard, 1998, for a related argument). In addition, a nonsignificant trend in the predicted direction was present in Experiment 1, indicating that the effect, although smaller, was possibly present. In conclusion, it is unclear why the number of list items preceding the lure affected only false recognition rates in Experiments 2 and 3. Future studies should investigate this issue further.

The number of list items included in the test, therefore, does appear to affect whether or not participants would respond "old" to a nonpresented critical lure; this effect appears to be relatively independent of the order in which items are presented at study. This conclusion is also supported by the significant increase in rates of false recognition of the critical lures associated to nonstudied lists after 6 or 12 list items were included on the test. This increase, indeed, can be due only to processes occurring at test, because participants had no study exposure to the items in question.

Semantic activation theories can explain how studying one word can prime concepts that are related in meaning, increasing their accessibility and possibly the likelihood that they will be falsely remembered (McDermott & Watson, 2001). Robinson and Roediger (1997) manipulated the number of items included in the lists, presenting 3, 6, 9, 12, or 15 words. They found that the probability of falsely remembering the critical lure was dependent on the number of words in the list: The more associated items participants studied, the higher the probability that they falsely recalled the critical lure. Data from Experiments 2 and 3 are consistent, at least in part, with this finding. One prediction of the present studies, however, failed to be completely supported. It was hypothesized that the rates of false recognition would increase as the number of list items presented at test increased. This hypothesis was only partially supported; false recognition increased significantly from the 0 items at test to the 6 items at test conditions for both studied and nonstudied lists, but no further increases occurred when 12 items were presented at test. Thus it appears that the additional items included in the latter condition failed to affect the rate of false recognition. This finding is inconsistent not only with the predictions of the present study, but also with prior studies indicating that longer lists resulted in higher rates of false memory (e.g., Robinson & Roediger, 1997). It is possible that the contribution of activation at test approaches ceiling levels after approximately 6 items are presented, and therefore that no further increases occur. Marsh et al. (2004), however, found no difference in rates of false alarms to lures associated to nonstudied lists between the 3 and the 6 items presented at test conditions.

Two possible explanations could account for the null difference between the 6 and 12 items at test conditions. First, it is possible that activation at test reaches a threshold at or around 6 items (or 3 items, as found by Marsh et al.). Although compelling at first, this explanation is not consistent with the evidence from Robinson and Roediger (1997), which showed a linear increase as a function of list length. A second explanation is that participants started endorsing nonstudied list items as a result of interitem associations and, in the 6 items at test condition, also endorsed the critical lure. In the 12 items at test condition, however, it is possible that because more items were presented, additional monitoring processes were called online to discriminate, and thus the tendency to endorse items as old failed to show any additional increases. The experiments reported here, however, cannot truly distinguish which explanation (if either of those proposed) is more likely to be correct; it therefore remains an empirical question. However, because no increase in false recognition rates occurred between the 6- and the 12-item conditions for both studied and nonstudied lists, these data seem to support the threshold explanation as more likely. Further studies might attempt to investigate whether any differences can be detected at other levels (i.e., by presenting 3 or 9 items at test as well as 0, 6, or 12), and why manipulating the number of list items at study results in monotonic increases in false memory (Robinson & Roediger, 1997), whereas the same manipulation at test does not.

One issue that remains to be addressed is how repeated presentation of list items at study or at test results in different effects. McDermott (1996) and Benjamin (2001) provided evidence that repeated exposure to DRM lists at study reduced overall false memory. However, it appears that in the present series of experiments, repeated presentation of list items at test increased false recognition. Although it is logical to assume that the automatic components of activation are quite similar during encoding and retrieval, it is also likely that participants adopt different strategies for learning items for an upcoming memory test and for processing the same items on a recognition test. Intentional encoding strategies are more likely to involve elaborative or semantic processing, thus resulting in the robust false memory rates observed in these and other studies. At test, however, participants simply have to decide whether an item was studied or not, and are thus less likely to try to encode it. Given the results reported here, it seems that, under certain conditions, the activation occurring at both study and test can summate; this is indicated by the higher rates of false alarms in the 6 and the 12 items at test conditions, when additional activation at test yielded more false memories than in the 0 items at test condition. Future studies might attempt to investigate whether the differences between activation at study and at test may be due to effortful processing occurring during encoding (e.g., by presenting lists under incidental learning conditions at study).

Several other studies have attempted to find evidence for activation of the critical lure at test (Anastasi et al., 2003; Dodd et al., 2006; Marsh et al., 2004). As in the present studies, these researchers also manipulated the number of list items that preceded the critical lure on recognition tests. Several methodological differences between the present studies and the previous ones might explain the inconsistent results. For example, Marsh et al. (2004) presented all 18 DRM lists at once, followed by a single recognition test. In their study, false alarms to unrelated fillers were quite high (.27). Thus, the relatively liberal response criterion adopted by participants in the Marsh et al. study might have obscured what appears to be a relatively small contribution from testing to the overall false memory effect.

Results of Experiments 2 and 3 in the present study indicate that some activation of the critical lure does occur during testing, but that the effects of such activation are weaker than those occurring during study. The most compelling evidence for the contribution of testing to the creation of false memories comes from the significant rates of false alarms to lures from nonstudied lists. The fact that participants identified as old more lures associated to nonstudied lists after being presented 6 or 12 list items than after no related items were presented in the recognition test indicates that activation can occur at test, since the list items and lures had not been presented at all during study. Although false alarms to lures from nonstudied lists were much lower than false alarms to lures from studied lists, they were still significantly greater than zero, and they thus provided strong support for the hypothesis that testing can and does contribute to the activation and subsequent false recognition of critical lures in the DRM paradigm.

In Experiment 2, even after correction for false alarms to noncritical intrusions ($M = .08$), false recognition of critical lures was still significantly greater than zero after participants processed 6 list items (corrected $M = .17$) or 12 list items (corrected $M = .16$), both $ps < .001$. In Experiment 3, although false alarms to lures from nonstudied lists were lower overall than in Experiment 2, corrected false recognition rates were still significantly greater than zero both in the 6 items at test condition ($M = .11$) and in the 12 items at test condition ($M = .10$), both $ps < .001$. Corrected false recognition of critical lures in the 0 items at test condition in both experiments was not significantly greater than zero following Bonferroni corrections for multiple tests. The slightly higher rate of false recognition of critical lures for which no associates were presented than of false recognition of unrelated fillers can be explained by the fact that the majority of lures used in the DRM paradigm are high-frequency words and are thus potentially more likely to appear familiar. False recognition of nonstudied list items, however, was at baseline; it did not differ from false alarms to unrelated fillers.

In the DRM paradigm, many list items not only are associates of the critical lure but also are associated in meaning to other list items, which thus increases the probability that interitem activation might also occur. Although the rates of false alarms to unrelated fillers and nonstudied list items did not differ, an examination of the serial position of the nonstudied list items mistakenly identified as old indicated that participants seemed to be more likely to call old nonstudied list items that appeared after approximately 4 items in the 6 items at test condition [$\chi^2(5) = 19.8, p = .001$] or 7 items in the 12 items at test condition [$\chi^2(5) = 27.6, p = .004$]. The fact that, on average, false recognition of nonstudied list items was found after presentation of at least 4 related items provides further support for activation theories of false memory. Although it is also possible that the increase in false alarms to list items was due to participants' adopting a more liberal response criterion as each test progressed or across the six tests, we believe that this is an unlikely explanation. First, false alarms to unrelated fillers remained low, and all of the unrelated fillers appeared among the last 14 items on the test. Furthermore, as indicated by the false alarm by block analyses in Experiments 2 and 3, there was no difference in false alarm rates to critical lures as a function of block. Taken together, these two points seem to run counter to the idea that changes in response criterion accounted for the observed effects.

False memory for the lure could, however, be due both to automatic spread of activation and to a second process dependent on participants' conscious attention to the list items (Roediger, Watson, McDermott, & Gallo, 2001). If participants become aware of the relatedness of the items in the list, they might expect the lure to appear on the list, and therefore might actively and consciously process it. McDermott (1997) argued that the critical lure is consciously thought of during study; she found evidence of perceptual priming in a stem completion task which, she posited, would not have occurred had the participants not accessed the lexical representation of the item during encoding. However, the fact that very fast presentations (20 or 80 msec—in which conscious processes are unlikely to be involved) can still result in false recognition of the lure provides empirical support for the fact that automatic processes play a significant role in the creation of false memories (Seamon, Luo, & Gallo, 1998; but see Zeelenberg, Plomp, & Raaijmakers, 2003).

A source-monitoring process in which participants actively decide the origin of an item is theorized to occur during the retrieval phase, in addition to activation processes (Gallo & Roediger, 2002; McDermott & Watson, 2001). Several manipulations that enable participants to monitor memory accuracy better have been found effective in reducing rates of false alarms to nonstudied items. These include warning subjects prior to the study (e.g., Multhaup & Conner, 2002; Neuschatz, Benoit, & Payne, 2003); presenting the list items in a more distinctive format, such as pictures (e.g., Israel & Schacter, 1997); giving participants repeated study trials (e.g., Benjamin, 2001; McDermott, 1996; McKone & Murphy, 2000); and varying the exposure time of studied items to allow deeper encoding (e.g., McDermott & Watson, 2001). Thus, it appears that when subjects are able to distinguish more reliably between old and new items through effective monitoring of the origin of an item, they can reduce their susceptibility to the illusion. The results for the nonstudied lists in Experiments 2 and 3 in the present study might be explained in terms of errors in source monitoring processes, such as participants' mistakenly deciding that the associated lures had been studied. As more semantically related list items were presented, the activation converged upon the critical lure, increasing its familiarity to the point that it might have been called old at greater than chance rates.

Fuzzy-trace theory has also been proposed to account for observed findings in false memory studies. This theory posits that items in a list of words are processed both at a surface form level and at a meaning content level, and that dissociated representations of both the former (verbatim traces) and the latter (gist traces) are stored (Brainerd & Reyna, 2002). Memory for the verbatim form is presumed to decline more rapidly than memory for the overall gist or theme of the list. During a memory task such as recognition, participants make a decision on the basis of the overall familiarity of the item as derived from the gist of the entire list, which results in memory errors. Gist and verbatim retrieval are proposed to have opposite effects on false memory: The former can result in increased errors if the judgments are made on the basis of familiarity, whereas the latter can reduce false memory by neutralizing the familiarity of meaning (Brainerd & Reyna, 2002). As the verbatim traces of individual items are strengthened—for example, by deeper processing during study—participants rely less on the gist traces to make a decision on a particular item. Results such as those observed with increased exposure time and repeated presentation of the lists (which are both manipulations that can strengthen the verbatim representation) are consistent with fuzzy-trace theory. Fuzzy-trace theory also implies that testing conditions might result in an increased level of false memory because semantically related items, presented in a recognition task or generated by participants in free recall, serve as cues for the gist traces of the lure (Toglia et al., 1999). This idea is consistent with Robinson and Roediger's (1997) finding that false recognition of the lure increases with the number of studied items, as well as with the results of Experiment 2 in the present study. The results of the present study are also consistent with another prediction of fuzzy-trace theory, in that the random presentation of list items in the study sessions of Experiment 2 might have contributed to the weakening of the gist traces and thus resulted in lower overall rates of false recognition. Therefore, although fuzzy-trace theory does not directly predict that retrieval processes may contribute to the creation of false memories, it is not inconsistent with the results of the present study.

In addition to activation-monitoring theory and fuzzy-trace theory, other models have been proposed to explain the occurrence of false memory in the DRM paradigm. According to models such as the one proposed by Anderson (1983), presentation of a cue during retrieval serves to reactivate associative pathways either present in semantic memory or laid down during encoding. In the DRM paradigm, encoding processes strengthen the traces connecting list items and nonpresented associates along semantic activation networks. At retrieval, the associative pathways cause repeated convergence of activation on the lure, thereby increasing its familiarity and the probability of an old response. This model can easily account for the effects reported here, since presentation of list items on a test can strengthen pathways laid down at encoding (for studied lists) or strengthen pathways in semantic memory during retrieval (for nonstudied lists).[4]

In conclusion, contrary to results reported by Marsh et al. (2004), Dodd et al. (2006), and Anastasi et al. (2003), results in the present study provided evidence that processes at both study and test contributed to false memory effects in the DRM paradigm. It appears that in this paradigm, the false memory phenomenon is driven largely by the processes that occur during study, as indicated by recent findings that participants were better able to reduce their susceptibility to the illusion when they were warned before study rather than before retrieval (McCabe & Smith, 2002). However, processes occurring at test can and do contribute to the size of the effect, and these testing effects should be considered in future studies of false memory effects using the DRM paradigm.

REFERENCES

Anastasi, J. S., Avery, J., Sinclair, S. M., Weitz, J. R., & Rhodes, M. G. (2003, June). *The case against critical lure activation during retrieval using the DRM paradigm*. Poster presented at the 15th annual convention of the American Psychological Society, Atlanta.

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning & Verbal Behavior*, **22**, 261-295.
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. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science*, **11**, 164-169.
Buckner, R. L., Wheeler, M. E., & Sheridan, M. A. (2001). Encoding processes during retrieval tasks. *Journal of Cognitive Neuroscience*, **13**, 406-415.
Cohen, J., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments, & Computers*, **25**, 257-271.
Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, **82**, 407-428.
Diez, E., Fernandez, A., & Alonso, M. A. (2004, November). *Activation at test contributes to false recognition in fast response conditions*. Poster presented at the 45th Annual Meeting of the Psychonomic Society, Minneapolis.
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.
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.
Israel, L., & Schacter, D. L. (1997). Pictorial encoding reduces false recognition of semantic associates. *Psychonomic Bulletin & Review*, **4**, 577-581.
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.
Mather, M., Henkel, L. A., & Johnson, M. K. (1997). Evaluating characteristics of false memories: Remember/know judgments and memory characteristics questionnaire compared. *Memory & Cognition*, **25**, 826-837.
McCabe, D. P., & Smith, A. D. (2002). The effect of warnings on false memories in young and older adults. *Memory & Cognition*, **30**, 1065-1077.
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., & Watson, J. M. (2001). The rise and fall of false recall: The impact of presentation duration. *Journal of Memory & Language*, **45**, 160-176.
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.
Multhaup, K. S., & Conner, C. A. (2002). The effects of considering nonlist sources on the Deese-Roediger-McDermott memory illusion. *Journal of Memory & Language*, **47**, 214-228.
Neuschatz, J. S., Benoit, G. E., & Payne, D. G. (2003). Effective warnings in the Deese–Roediger–McDermott false-memory paradigm: The role of identifiability. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 35-41.
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, Balota, D. A., & Watson, J. M. (2001). Spreading activation and the arousal of false memories. In H. L. Roediger III, J. S. Nairne, I. Neath, & A. 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, Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review*, **8**, 385-407.
Seamon, J. G., Luo, C. R., & Gallo, D. A. (1998). Creating false memories of words with or without recognition of list items: Evidence for nonconscious processes. *Psychological Science*, **9**, 20-26.
Seamon, J. G., Luo, C. R., Kopecky, J. J., Price, C. A., Rothschild, L., Fung, N. S., & Schwartz, M. A. (2002). Are false memories more difficult to forget than accurate memories? The effect of retention interval on recall and recognition. *Memory & Cognition*, **30**, 1054-1064.
Stadler, M. A., Roediger, H. L., III, & McDermott, K. B. (1999). Norms for word lists that create false memories. *Memory & Cognition*, **27**, 494-500.
Thapar, A., & McDermott, K. B. (2001). False recall and false recognition induced by presentation of associated words: Effects of retention interval and level of processing. *Memory & Cognition*, **29**, 424-432.
Toglia, M. P., Neuschatz, J. S., & Goodwin, K. A. (1999). Recall accuracy and illusory memories: When more is less. *Memory*, **7**, 233-256.
Tun, P. A., Wingfield, A., Rosen, M. J., & Blanchard, L. (1998). Response latencies for false memories: Gist-based processes in normal aging. *Psychology & Aging*, **13**, 230-240.
Zeelenberg, R., Plomp, G., & Raaijmakers, J. G. W. (2003). Can false memories be created through nonconscious processes? *Consciousness & Cognition*, **12**, 403-412.

NOTES

1. In addition to recording old/new responses, we also analyzed median reaction times (RTs) to old responses. Overall, RTs were significantly shorter for list items (median average = 821 msec) than for critical lures (median average = 895 msec) [$F(1,40) = 12.46, p = .001$], but RTs did not differ for number of list items presented at test [$F(1,40) < 1.0, p = .682$].

2. Median RTs were also analyzed in Experiment 2. A 2 (list type: studied, nonstudied) × 3 (number of items presented at test: 0, 6, 12) ANOVA revealed a main effect of list type, with faster responses to critical lures associated to studied lists (median average = 764 msec) than to lures associated to nonstudied lists (median average = 1,029 msec) [$F(1,27) = 15.8, p = .001$], but, as in Experiment 1, no significant effect of number of items at test was found [$F(2,42) < 1.0, p = .746$].

3. We thank Dave McCabe for bringing this point to our attention.

4. We thank Dave Balota for suggesting this alternative theoretical approach.

(Manuscript received October 16, 2003;
revision accepted for publication June 15, 2005.)