Running head: TESTING CONDITIONS INFLUENCE FALSE MEMORY

MS # 03-286s2

Production # C626

The Role of Test Structure in Creating False Memories

Jennifer H. Coane

and

Dawn M. McBride

Illinois State University

Address Correspondence to:
Dawn M. McBride
Department of Psychology
Illinois State University
Campus Box 4620
Normal, IL 61790-4620
e-mail: dmcbride@ilstu.edu
(309) 438-7146
Fax: (309) 438-5789

Abstract

In the Deese-Roediger-McDermott (DRM) paradigm, studying lists of semantic associates results in high rates of false recognition of a non-presented critical word. The current set of studies was designed to measure the contribution of additional processing of list items at test to this false memory effect. Participants studied sets of lists and then performed a recognition task for each set. Using this paradigm, three experiments 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 non-studied 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.

The Role of Test Structure in Creating False Memories

In recent years, the study of experimentally induced false memory phenomena has increased substantially. One method by which false memories are studied involves using word list paradigms that are based on properties of semantic association and 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 non-presented 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 for having studied the item) for the list items and critical lures (Roediger & McDermott, 1995). In other words, participants are able to “retrieve” the actual experience of encoding the critical item, when in fact it was not studied.

The false memory effect appears quite robust and 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 for lists that had been previously recalled than for 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, while accurate memories decrease (McDermott, 1996; Thapar & McDermott, 2001; Toglia, Neuschatz, & Goodwin, 1999). Seamon, Luo, Kopecky, Price, Rothschild, Fung, and Schwartz (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.

There is also evidence that retrieval processes might 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 might indicate that prior recall of list items serves as a cue for the lure, or that recalling the list items increases the probability that participants will recall the lure as a highly associated item to all studied items. In recognition tasks, one or more list items are often presented before the lure, and might thus contribute to additional priming of the lure (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 associated to all list items (i.e., the critical lure). Consequently, the critical lure may be falsely remembered due to the heightened activation. It is generally understood that this activation process occurs automatically, and it 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, so 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 occurring primarily at study combine with source monitoring errors at test to elicit high rates of false memory (McDermott & Watson, 2001). Monitoring processes might occur as participants attempt to decide whether an item was studied or thought of, and errors in these processes can result in false alarms to non-studied items such as the lure.

There is evidence from the literature that activation from multiple associates summates, so exposure to increasing numbers of related concepts will result 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, and 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 re-presented in a recognition test or recalled by participants in a free recall test. The majority of studies have investigated how encoding manipulations contribute to false memory, whereas retrieval manipulations are relatively rare. Although it is not explicitly stated in the literature that activation does not occur at test, to date very little evidence has been presented that supports this possibility. 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 re-activated 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 the number of list items presented prior to the lure might 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, such 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 no difference in false lure recognition due to the number of studied items presented before the lure.

More recently, Dodd, Sheard, and MacLeod (in press) 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 and immediate testing following each list, Dodd et al. failed to find any differences in false alarm rates with zero to five list items tested before the lure. In their second experiment, they gave participants a response deadline of 750 ms to encourage participants to rely more on familiarity 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, concluding that retrieval factors do 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 ms and 750 ms. False alarms were higher under speeded instructions when four list items preceded the lure than when no items preceded the lure on the test. However, when subjects were given 1500 ms to respond, the number of list items preceding the lure had no effect on false recognition rates. Thus, the evidence for 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: It 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 result in participants mistakenly identifying the lure as a studied item might be strong enough that additional effects are simply not detectable with the method 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. However, when subjects were warned between study and test, the warning had little effect, indicating that the lure might have been encoded similarly to the studied items.

Although there appears to be converging evidence indicating that additional processing of list items at test has no effect on false memory rates (Anastasi et al., 2003; Dodd et al., 2004; 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 there is no reason to assume activation should not occur in the testing phase and that this activation might summate with the activation due to the encoding phase. One possibility is that the activation is similar in nature to the activation that occurs at encoding, and that each additional phase of activation further increases the probability the critical lure will be falsely recognized. Furthermore, recent data from a neuro-imaging study indicates that foils on a recognition test that are later remembered in a surprise test show the same patterns of activation as 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 of contributions to false recognition rates due to activation occurring in the testing phase is provided by false alarm rates for critical lures from non-studied lists in the Marsh et al. (2004) study. Marsh et al. included non-studied lists and lures with unrelated items as fillers in the recognition test. Critical lures associated to non-studied lists were falsely recognized 31% of the time when no list items were presented before them in the recognition test. After test trials of three or six list items from non-studied lists, however, false recognition of non-studied lures increased to 49%. Thus, it would appear that when study effects are absent, simply presenting as few as three semantic associates as filler items in a recognition test can increase false recognition rates of critical lures. The fact that participants consistently falsely recognized critical lures from lists they never studied after viewing as few as three list items indicates that some activation resulting in false memory likely occurs at test. Therefore, the high rate of false recognition of critical lures from non-studied lists might be a good indicator of the amount of activation that occurs during test. If indeed activation does occur at test, source monitoring errors and encoding effects may not be the only factors contributing to the 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 current studies were designed with methodology similar to that used 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 size 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 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 the critical lure. Rather than presenting zero, three, or six items prior to the lure, we presented zero, six, or 12, hypothesizing that a larger effect would be detected by increasing the opportunities for activation at test to converge on the lure. Experiment 2 was designed to address the possibility that effects of the testing manipulation were absent in past studies due to 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 due to strength of activation occurring during encoding. Presenting the lists in random order should sufficiently weaken the effects that occur at study to allow detection of those that might occur at test. In addition, as several prior studies have reported high rates of false recognition for critical lures associated to non-studied lists (e.g., Marsh et al., 2004; McDermott, 1996; McCabe & Smith, 2002; Seamon et al., 2002), 0, 6, or 12 items from non-studied 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 can be induced to call *old* an item they had not studied and for which no associates had been studied but only presented at test and 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 worthy of highlighting. In their study, Marsh et al. presented 18 lists of 15 items each and administered a single recognition test. In the current studies, the lists were studied in blocks of three and tested after each block. This was done to avoid the high non-critical 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, thus providing a stronger estimate of any observed effects.

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 between the three testing conditions could be attributed to this manipulation. In Experiment 1, we manipulated the number of list items presented at test prior to the lure in a recognition test with blocked study presentation. 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 in a single experiment. Furthermore, in Experiments 2 and 3, we included non-studied list items and the associated critical lures in the recognition test to assess whether we could create false memories based solely on processes occurring during retrieval. Finding increases in false alarms to the critical lures associated to non-studied 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 in exchange for extra credit for their classes. All participants were native speakers of English.

Design and Materials. A 3 (item type: list, lure, unrelated) x 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 non-presented 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 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 session, two 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, six randomly selected from a second list, and the four primacy and recency items. The remaining 23 items consisted of the three critical lures and 20 non-studied, unrelated filler items. The critical lures were always presented after the list items among the last nine items on the recognition test. Other than this constraint, the order in which items appeared at test was randomized across participants. Both study and test were administered by computer.

Procedure. 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). Participants were told that they would be presented lists of words and their memory would be tested. They were told that the words would be presented one at a time, and they should pay close attention to each one, as 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 3000 ms and the inter-stimulus interval was 500 ms. 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, participants completed a filler task for 30 s, consisting of mathematical problems contained in a booklet provided by the experimenter. After 30 s, a beep notified participants to begin the recognition test on the computer. The instructions on the screen directed participants to decide whether or not each word that appeared on the screen had been included in the list they most recently studied. If they were sure they had studied a word, they were instructed to press “O” on the keyboard. If they were sure the 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 were determined to be sufficiently distant to avoid confusion but on the same row on the keyboard to maximize comfort. 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 begin the next part of the experiment and the next block of three lists was presented. At the end of the experiment, participants were debriefed. The entire experiment lasted approximately 40 min.

Results & 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, $F(2,82) < 1.0$, $p = .54$, $MSe = .02$.

A separate 2 (item type: list vs. lure) x 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, $F(1,41) = 7.22$, $p = .01$, $MSe = .04$. Neither the main effect of the number of items presented at test, $F(1,41) < 1.0$, $p = .88$, $MSe = .0003$, nor the interaction between item type and the number of items presented at test, $F(1,41) < 1.0$, $p = .45$, $MSe = .01$, were found to be significant, indicating that the number of list items included in the recognition test failed to significantly affect both true and false memory. Accurate recognition of studied items was high overall ($M = .79$), regardless of whether six or 12 items were included in the test. False recognition rates did not differ when six or 12 list items were 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 non-critical 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 be masking any additional effects that might 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. Based upon the results obtained in Experiment 1 and upon those obtained by Marsh et al. (2004) and Anastasi et al. (2003) that failed to detect any differences in rates of false recognition as a function of the number of list items presented at test, Experiment 2 was designed to include 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 non-studied lists as fillers in the recognition test. Rates of false recognition of critical lures associated to non-studied 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 only be attributed to activation processes occurring at test as 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 in exchange for extra credit for their classes. All participants were native speakers of English.

Design & Materials. A 3 (item type: list, lure, unrelated) x 2 (list type: studied, non-studied) x 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 because the more complex design required more lists to 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 six blocks of six lists each. Within each block, the words from three 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 these, 18 were studied items, 12 from one studied list and six randomly selected items from a second studied list. In addition, 12 words from one non-studied list and six randomly selected items from a second non-studied list were tested. The items selected for the six items at test conditions were randomly chosen. The mean backward associative strength (BAS) of the list items of the six and 12 items at test conditions for both studied and non-studied 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 six critical lures (three from studied lists, three from non-studied lists) were tested among the last 14 items for each set, together with the four primacy and recency items and four 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 & 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. non-studied) x 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 non-studied lists, $F(1,63) = 215.9$, $p < .0005$, $MSe = .08$. 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 present, $F(2,62) = 18.03$, $p < .001$, $MSe = .03$, indicating that for both studied and non-studied 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 six list items were presented at test than when zero list items were presented, $t(63) = - 2.1$, $p = .041$. When 12 items were presented at test false recognition rates also increased compared to when no items were presented, *t*(63) = - 3.3, *p* = .002. For non-studied lists, a similar pattern emerged, with higher rates of false recognition in the six items at test condition than in the zero items at test condition, *t*(63) = - 5.0, *p* < .001. Higher rates of false recognition of critical lures was also observed in the 12 items at test conditions compared to the zero items at test condition, *t*(63) = - 4.8, *p* < .001. In both studied and non-studied lists no differences were found between the six and 12 items at test conditions, both *p*s > .05. Thus, it appears that when the study effects were weakened (as for studied lists in Experiment 2) or were absent (as for non-studied lists), false recognition of critical lures did increase significantly when more list items were presented at test (for both studied and non-studied 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, due to fatigue or other factors, participants might have simply become more liberal in accepting non-studied items as *old*. To rule out the hypothesis that participants were simply adopting a more liberal criterion over the repeated study-test blocks, a false alarm by block analysis was performed. No significant differences emerged between the number of false alarms occurring in each block, χ^2 (5) = 2.497, *p* = .777, indicating 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, *p* = .35, *MSe* = .03, indicating that false recognition of critical lures associated to studied and non-studied 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 non-studied lists, $t(63) = -7.1$, $p < .001$.

True and false recognition data were analyzed using a 2 (list type: studied vs. non-studied) x 2 (item type: list vs. lure) x 2 (number of items presented at test: 6 vs. 12) repeated measures ANOVA. Overall, studied items were called *old* more often than non-studied items, $F(1,63) = 683.9$, $p < .001$, $MSe = .06$. 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, $F(1,63) = 81.14$, $p < .001$, $MSe = .04$. For the studied lists, list items were recognized more often than critical lures, but for the non-studied 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 non-studied 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 between list type, item type, and number of items presented at test was not significant. In conclusion, data from Experiment 2 indicated 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, non-studied 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 non-studied lists.

Method

Participants. One hundred and 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 one additional participant were omitted due to equipment failure. All participants were native English speakers and received course credit for their participation.

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

Materials & Procedure. The same materials from Experiment 2 were used, except that half the participants were randomly assigned to the conditions in which lists were blocked at study and half studied lists in random order. The assignment of lists to blocks in the two conditions was the same, in order to avoid potential confounds of different lists being in different conditions. 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 & Discussion

The effects of list presentation type (blocked or random order) were directly tested in Experiment 3. Overall, the same pattern of results 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 non-studied lists, $F(1,103) = 650.99$, $p < .001$, $MSe = .06$; a significant effect of number, indicating that, overall, false alarms increased as a function of the number of items presented at test, $F(2,206) = 18.68$, $p < .001$, $MSe = .02$. The increase in false alarm rates between the zero and six items at test conditions was significant for both studied and non-studied lists, $t(104) = 3.62$, $p < .001$ and $t(104) = 4.53$, $p < .001$ respectively, as was the increase between the zero and 12 items at test conditions, $t(104) = 3.23$, $p = .002$ and $t(104) = 3.68$, $p < .001$ for studied and non-studied lists. False alarm rates did not differ when six or 12 items preceded the lure ($p = .84$ and $p = .50$ for studied and non-studied lists respectively). The effect of order was marginally significant, such that overall false recognition was higher when lists were blocked at study, $F(1,103) = 3.16$, $p = .078$, $MSe = .11$. Finally, a significant list type by order interaction was found, $F(1,103) = 5.49$, $p = .021$, $MSe = .06$, indicating 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, $t(103) = -.02$, $p = .99$. 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, χ^2 (5) = 7.327, $p = .197$.

A 2 (item type) x 2 (number of items presented at test) x 2 (list type) ANOVA was conducted to compare memory for list items and critical lures. A significant effect of list type indicated overall recognition (false and veridical) was higher when lists were studied, $F(1,103) = 1933$, $p < .001$, $MSe = .04$. The list type by order of presentation interaction was significant, $F(1,103) = 10.2$, $p = .002$, $MSe = .04$. When lists were studied, overall recognition rates were higher in the blocked order condition than in the random order condition; this difference, however, was primarily driven by the difference in false recognition rates between the two order conditions. For non-studied lists, no differences emerged between blocked and random presentation of study items. Finally, a significant list type by item type interaction was found, $F(1,103) = 100.8$, $p < .001$, $MSe = .04$. 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 maintaining other aspects of the design constant from Experiment 2 (i.e., the type of foils included on the recognition test, the number of blocks, etc.). 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 non-studied 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 indicate that testing effects based on the number of items presented prior to the critical lure can 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 below.

Furthermore, the results of Experiment 3 allow 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 non-critical 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 ($M = .05$ and .07, respectively) and were not significantly different ($p = .32$).[3] The finding of significant increases in false alarm rates to lures associated to non-studied lists observed in Experiment 2 was also replicated. In both study conditions, the critical lures from non-studied lists were more likely to be falsely recognized after six 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 the critical lure following study of lists of semantic associates. In the current study, however, the experiments were designed specifically 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 six or 12 items preceded the critical lure. When six 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 non-studied lists provided a good measure of testing effects. The high rates of veridical recognition in all experiments, combined with low rates of non-critical intrusions indicated that participants were not simply guessing in their responses.

In comparison to Experiment 1, presentation of list items in random order in Experiment 2 appeared to have been successful in reducing the overall rates of false recognition for studied lists. The manipulation appeared 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 versus .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 were tested in Experiment 1 (42, compared to 64 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 zero and six items preceding the lure and zero and 12 items preceding the lure, indicating low power may not be a sufficient explanation for the null results in Experiment 1. However, an important difference between Experiments 1 and 3 concerned 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 non-studied 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 non-significant 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 only affected 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 the likelihood with which participants respond *old* to a non-presented critical lure and 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 non-studied lists after six or 12 list items were included on the test. This increase, indeed, can only be due to processes occurring at test, as 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 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 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, as the number of list items presented at test increased, so would the rates of false recognition. This hypothesis was only partially supported, as false recognition increased significantly for both studied and non-studied lists from the zero items at test to the six items at test conditions, 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 not only inconsistent with the predictions of the present study, but also with prior studies indicating longer lists result 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 six items are presented, and no further increases occur. Marsh et al. (2004), however, found no difference in rates of false alarms to lures associated to non-studied lists between the three and six items presented at test conditions. There are two possible explanations for the null difference between the six and 12 items at test conditions. First, it is possible that activation at test reaches a threshold at or around six items (or three items, as found by Marsh et al.). Although compelling at first, this explanation is not consistent with the evidence from Robinson and Roediger (1997), who found a linear increase as a function of list length. A second explanation is that participants started endorsing non-studied list items as a result of inter-item associations and, in the six items at test condition also endorsed the critical lure. In the 12 items at test condition, however, it is possible that as 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 studies 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 12 item conditions for both studied and non-studied 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 three or nine items at test as well as zero, six, or 12) and why manipulating the number of list items at study results in monotonic increases in false memory (Robinson & Roediger, 1997) and 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 when attempting to learn items for an upcoming memory test and when they are 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 have to decide whether an item was studied or not, and are thus less likely to try to encode it. Based on the results reported here, it seems that the activation occurring at both study and test, under certain conditions, can summate, as indicated by the higher rates of false alarms in the six and 12 items at test conditions, when additional activation at test yielded more false memories than in the zero items at test condition. Future studies might attempt to investigate whether the differences between activation at study and at test are 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., in press; 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 current 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). It is possible that the relatively small effects detected here simply were masked by the fact that participants seemed to be adopting a relatively liberal response criterion.

Results of Experiments 2 and 3 in the current 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 non-studied lists. The fact that participants identified as *old* more lures associated to non-studied lists after being presented six or 12 list items in the recognition test indicates that activation can occur at test, as the list items and lures had not been presented at all during study. Although false alarms to lures from non-studied lists were much lower than false alarms to lures from studied lists, they were still significantly greater than zero, and thus provide 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 correcting for false alarms to non-critical intrusions (*M* = .08), false recognition of critical lures was still significantly greater than zero after processing six list items (corrected *M* = .17) or 12 list items (corrected *M* = .16), both *p*s < .001. In Experiment 3, although false alarms to lures from non-studied lists were overall lower than in Experiment 2, corrected false recognition rates were still significantly greater than zero in both the six items at test condition (*M* = .11) and in the 12 items at test condition (*M* = .10), both *p*s < .001. Corrected false recognition of critical lures in the zero 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 compared to 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 thus potentially more likely to appear familiar. False recognition of non-studied list items, however, was at baseline, as it did not differ from false alarms to unrelated fillers.

In the DRM paradigm, many list items are not only associates of the critical lure but are also associated in meaning to other list items, thereby increasing the probability that inter-item activation might also occur. Although the rates of false alarms to unrelated fillers and non-studied list items did not differ, an examination of the serial position of the non-studied list items mistakenly identified as *old* indicated that participants seemed to be more likely to call *old* non-studied list items that appeared after approximately four items (in the six items at test condition, $\chi^2[5] = 19.8$, $p = .001$) or seven items (in the 12 items at test condition, $\chi^2[5] = 27.6$, $p = .004$). The fact that, on average, false recognition of non-studied list items was found after presentation of at least four 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 this is an unlikely explanation. First of all, false alarms to unrelated fillers remained low, and all 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 were behind the observed effects.

False memory for the lure could, however, be due to both automatic spread of activation and to a second process dependent on conscious attention to the list items by participants (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, thus they actively and consciously process it. McDermott (1997) argued that the critical lure is consciously thought of during study, as she found evidence of perceptual priming in a stem completion task which, she posited, should not occur had the participants not accessed the lexical representation of the item during encoding. However, the fact that very fast presentations (20 or 80 ms) can still result in false recognition of the lure, when conscious processes are unlikely to be involved, 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, & Raajmakers, 2003).

In addition to activation processes, a source-monitoring process in which participants actively decide the origin of an item is theorized to occur during the retrieval phase (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 non-studied 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. Results for non-studied lists in Experiments 2 and 3 in the present study might be explained in terms of errors in source monitoring processes as subjects mistakenly decided 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 it was be 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 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 will make a decision on the basis of the overall familiarity of the item, as provided by the gist of the entire list, thus resulting 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, while 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 (both manipulations that can strengthen the verbatim representation) are consistent with fuzzy-trace theory. Fuzzy-trace theory, on the other hand, 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 further consistent with one 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 resulted in lower overall rates of false recognition. Therefore, although fuzzy-trace theory does not directly predict that retrieval processes might contribute to the creation of false memories, it is not inconsistent with the results found in the current 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), during retrieval presentation of a cue serves to re-activate 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 non-presented 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, as 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 non-studied lists).[4]

In conclusion, contrary to results reported by Marsh et al. (2004), Dodd et al. (in press), and Anastasi et al. (2003), results in the current study provided evidence that processes at both study and test contribute to false memory effects in the DRM paradigm. It appears that, in the DRM paradigm, the false memory phenomenon is driven largely by the processes that occur during study, as indicated by recent findings that participants are better able to reduce their susceptibility to the illusion when warned before study than when the warning is given 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 explanations of false memory effects using the DRM paradigm.

Author Note

Portions of the present study were conducted as partial fulfillment of the requirements for a Master's degree for the first author, who is presently at Washington University in Saint 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.

Footnotes

[1] In addition to recording *old/new* responses, median reaction times (RTs) to *old* responses were also analyzed. Overall, RTs were significantly faster for list items (*Median* average = 821 ms) than for critical lures (*Median* average = 895), $F(1,40) = 12.46$, $p = .001$, but RTs did not differ based on 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 vs. non-studied) x 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 with studied lists (*Median* average = 764) than to lures associated to non-studied lists (*Median* average = 1029), $F(1,27) = 15.8$, $p = .001$, 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.

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, GA.

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior, 22,* 261-295.

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

Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science, 4*, 577-581.

Buckner, R. L., Wheeler, M. E., & Sheridan, M. A. (2001). Encoding processes during retrieval tasks. *Journal of Cognitive Neuroscience, 13,* 406-415.

Cohen J. D., MacWhinney B., Flatt M., & Provost J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and 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*). Activation at test contributes to false recognition in fast response conditions*. Poster presented at the 45th Annual Convention of the Psychonomic Society, Minneapolis, MN.

Dodd, M. D., Sheard, E. D., & MacLeod, C. M. (in press). Re-exposure to studied items at test does not influence false recognition. *Memory.*

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 and 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 and Review, 8*, 579-586.

Israel, L, & Schacter, D. L. (1997). Pictorial encoding reduces false recognition of semantic associates. *Psychonomic Bulletin and Review, 4*, 582-586.

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 compared. *Memory and 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 and Language, 35*, 212-230.

McDermott, K. B. (1997). Priming on perceptual implicit memory tests can be achieved through presentation of associates. *Psychonomic Bulletin and 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 and 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 and 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 and Language, 47*, 214-228.

Neuschatz, J. S., Beniot, 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, and 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). Analysis of the serial position curve. In H. L. Roediger & J. S. Nairne. (Eds). *The nature of remembering: Essays in honor of Robert G. Crowder,* (pp. 95-115). Washington, D.C.: 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, and 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 words 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 level of processing and retention interval. *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 and Aging, 13,* 230–240.

Zeelenberg, R., Plomp, G., & Raajmakers, G. W. (2003). Can false memories be created through nonconscious processes? *Consciousness and Cognition, 12,* 403-412.

Table 1

Mean Recognition Rates as a Function of Item Type and Number of Items Presented at Test in Experiment 1 (Standard Errors in Parentheses)

	Item type	
Number of items at test	List	Critical Lure
Zero	------	.67 (.03)
Six	.80 (.02)	.70 (.04)
Twelve	.79 (.02)	.71 (.04)
Unrelated Fillers	.08 (.01)	

Table 2

Mean Recognition Rates as a Function of List Type, Item Type and Number of Items Presented at Test in Experiment 2 (Standard Errors in Parentheses)

	Item type	
Number of items at test	List	Critical Lure
Studied Lists		
Zero	------	.57 (.03)
Six	.82 (.02)	.64 (.04)
Twelve	.80 (.02)	.68 (.03)
Non-studied Lists		
Zero	------	.11 (.02)
Six	.08 (.01)	.25 (.03)
Twelve	.09 (.01)	.24 (.03)
Unrelated Fillers	.08 (.01)	

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

	Item Type	
Number of Items at Test	List Items	Critical Lures
Blocked Presentation		
Studied Lists		
Zero	---	.62 (.24)
Six	.83 (.12)	.74 (.22)
Twelve	.83 (.15)	.73 (.23)
Non-studied Lists		
Zero	---	.10 (.12)
Six	.04 (.05)	.17 (.19)
Twelve	.06 (.06)	.16 (.20)
Unrelated Fillers	.05 (.01)	
Random Presentation		
Studied Lists		
Zero	---	.56 (.26)
Six	.80 (.11)	.63 (.26)
Twelve	.78 (.12)	.63 (.27)

Non-studied Lists		
Zero	---	.08 (.14)
Six	.06 (.20)	.18 (.20)
Twelve	.08 (.10)	.17 (.19)
Unrelated Fillers	.07 (.01)	