False Memory in a Short-Term Memory Task

Jennifer H. Coane[1], Dawn M. McBride[2], Bascom A. Raulerson III[2], and J. Scott Jordan[2]

[1]Washington University in Saint Louis, MO, [2]Illinois State University, Normal, IL, both USA

Abstract. The Deese/Roediger-McDermott (DRM; Roediger & McDermott, 1995) paradigm reliably elicits false memories for critical nonpresented words in recognition tasks. The present studies used a Sternberg (1966) task with DRM lists to determine whether false memories occur in short-term memory tasks and to assess the contribution of latency data in the measurement of false memories. Subjects studied three, five, or seven items from DRM lists and responded to a single probe (studied or nonstudied). In both experiments, critical lures were falsely recognized more often than nonpresented weak associates. Latency data indicated that correct rejections of critical lures were slower than correct rejections of weakly related items at all set sizes. False alarms to critical lures were slower than hits to list items. Latency data can distinguish veridical and false memories in a short-term memory task. Results are discussed in terms of activation-monitoring models of false memory.

Keywords: false memory, short-term memory, reaction time

The Deese/Roediger-McDermott (DRM; Roediger & McDermott, 1995) paradigm was developed to investigate false memory. In this paradigm, participants study lists of words (e.g., bed, rest, awake, etc.) related to an associated, nonpresented lure (e.g., sleep). On later memory tests, participants reliably intrude or falsely recognize this critical lure (CL).

Research comparing accurate memories for list items and false memories for CLs has focused on long-term memory tasks. The purpose of the current study was to extend these findings to a short-term recognition task. If subjects falsely recognize the CL in a short-term memory task after studying very short lists of items, this implies that (1) the processes underlying the creation of false memories occur rapidly, and (2) little cumulative association is needed to produce false recognition. Furthermore, a short-term memory task provides the opportunity to examine response times (RTs) in combination with accuracy data, thus, offering further insight into the cognitive processes underlying memory performance.

We used a Sternberg (1966) task with DRM lists. In this paradigm, subjects study sets of one to seven items. Immediately after the last item in the set, a probe is presented, and participants indicate whether the probe was included in the preceding set. In general, RT increases as set size increases, with similar functions for positive and negative probes.

Prior studies have examined the effects of relatedness on short-term memory. Using lists of semantically related items, Jones and Anderson (1982) found that when a recognition probe was a member of the same semantic category as an item from the study set, RTs and false-alarm rates increased. Bartha, Martin, and Jensen (1998) found that correct rejections of probes related to one item from the memory set were significantly slower than rejections of unrelated probes and that related probes yielded higher error rates than unrelated probes. Thus, it appears that semantic information plays a role in short-term memory tasks. However, past studies did not focus on false memories or the processes that cause them.

Applying the DRM paradigm to a short-term recognition task provides a stronger test of relatedness effects in short-term memory. The DRM lists were designed such that the CL for a list would be the strongest associate of all the list items. Therefore, if associative false memory effects occur in short-term memory tasks, the DRM paradigm should detect them. Further, if these effects are found in a short-term memory task (with few study items and minimal delay), this suggests that the processes underlying these effects occur simultaneously with or immediately after study.

In the DRM paradigm, false memory can be explained by the activation-monitoring account (Roediger, Balota, & Watson, 2001), according to which the CL is activated through spreading activation processes as participants encode related items. The activation converging on the CL increases its familiarity, leading to memory errors at the time of test, when subjects mistakenly attribute the item's familiarity to a study event. The CL's high activation should encourage subjects to respond *old* to these items, but the small memory load in a short-term memory task (seven items at most) should allow them to monitor the source of the CL's familiarity. Monitoring in this case should be easier compared to DRM studies that use longer study lists and subjects should be better able to counteract the effects of activation and correctly reject the CL. The monitoring process, however, should be slow, according to two-process models of recognition (e.g., Atkinson & Juola, 1974), which assume separate criteria for *old* and

DOI 10.1027/1618-3169.54.1.62

new responses. If an item's familiarity falls above the higher criterion, subjects can make a rapid *old* response. If familiarity falls below the lower criterion, subjects will make a rapid *new* response. Items whose familiarity falls between the two criteria undergo an additional checking process, yielding longer latencies. The heightened familiarity of the CL should lead participants toward endorsing the CL as *old,* but also engage the slower checking process. Thus, responses to CLs should be slower than responses to unrelated or weakly related items, which should be less likely to engage the checking process. Specifically, we expected slower correct rejections and false alarms for CLs, as these items should require additional checking because of the heightened activation. Consistent with prior findings showing that false alarms to CLs increase as list length increases (Robinson & Roediger, 1997), we also expected more false alarms as list length increased, reflecting additional activation processes. Thus, effects on memory performance should emerge in both latency and accuracy data.

Two experiments were designed to compare RTs and accuracy in a short-term memory task. DRM lists of set size three, five, and seven items were presented with a single-item recognition test immediately following each study set. Set size was manipulated to examine the effect of the number of related list items on short-term false memory (cf. Robinson & Roediger, 1997) and to determine whether set size effects might dissociate RTs to list and CL items. Using short lists allowed us to emphasize speed and accuracy instructions. The majority of studies using the DRM paradigm have emphasized accuracy over speed, and latency data have rarely been reported. The few studies reporting RT data in the DRM (e.g., Jou, Matus, Aldridge, Rogers, & Zimmerman, 2004; Tun, Wingfield, Rosen, & Blanchard, 1998) have not clarified the role RTs play in discriminating veridical and false memories, as they reported different results. Jou et al. reported that false alarms were slower than hits, but Tun et al. reported no difference between hit and false alarm RTs. Further, both studies used long lists and retention intervals and, thus, may not offer a stringent test of the predictions of an activation account of false memories in the DRM paradigm. The short lists used in the present studies were designed to provide a measure of activation converging on the CL while reducing participants' memory load and allowing for effective monitoring processes. Consistent with the activation-monitoring account, we expected slower latencies for correct rejections of CLs, as activation and monitoring were placed in opposition.

To summarize, we predicted that responses to CLs (hits and false alarms) would be slower than responses to the other probes because of the heightened activation or because these items would be more likely to engage additional monitoring processes. We also predicted that errors and latencies would increase as a function of set size, but more so for CLs as the increased activation of these items would elicit more errors and longer RTs.

General Method

The two experiments have similar methodologies which are described below.

Materials and Design

Probe type and set size were manipulated within subjects. The studied list item probe was randomly selected from the study set. The weakly related list item probe was selected at random from the items in that list not included in that study set. The CL probe was the CL associated with that list. The unrelated probe was an item unrelated to any item in all DRM lists.

Thirty-six lists from Stadler, Roediger, and McDermott (1999) were used. The related memory sets contained three, five, or seven items selected at random from each list. Mean backward associative strength (BAS, the probability that a list item will elicit the CL on a free association task; Roediger, Watson, McDermott, & Gallo, 2001) was equivalent across sets. Initial selection of items for each set was random and each set was checked to ensure that mean BAS was equivalent across all three set sizes. When necessary, items with higher or lower BAS were replaced. The mean BAS across all sets was .23.

Within a block, each DRM list appeared once. Participants studied a memory set from each DRM list four times throughout Experiment 1 and three times throughout Experiment 2. Sets of different size were presented randomly within each block of 36 lists so subjects could not anticipate the length of each set on a trial-by-trial basis.

To discourage participants from adopting a liberal response criterion (i.e., guessing "yes"), the correct response was "no" on 75% of the trials in Experiment 1 and 66% in Experiment 2. This balance of *old* to *new* responses was selected for three reasons. First, pilot data showed that participants adopted a liberal criterion (i.e., false alarms to unrelated probes and CLs were equally high). Second, we were interested in whether differences in latencies would emerge under conditions contributing to easy rejection of the CL. Third, given the high false-alarm rates observed in DRM studies, we predicted that many CLs would be endorsed as *old,* thus, increasing the ratio of positive to negative trials in terms of participants' response patterns.

Procedure

Subjects were tested individually. The experiment was administered on a Macintosh computer. Instructions emphasized the importance of speed and accuracy equally. Responses were recorded using a button box. Subjects were told to respond "yes" if the probe had been in the preceding set and "no" if the probe had not been in the set. The "yes" button and the "no" button were counterbalanced across subjects.

Items appeared one at a time in the center of the screen, in white on a black background, for 750 ms with a 250 ms interstimulus interval. After the last item in each set, a fixation cross appeared for 750 ms, followed by the probe, which remained visible until the subject made a response. Participants pressed the space bar to initiate the next set. At the end of each block, participants took a 1 min break and initiated the next block with a key press.

Experiment 1

In Experiment 1, all four probe types were included. The effects of degree of relatedness were tested by using probes that vary in associative strength to the list items (e.g., CLs were most related, weakly related items moderately related, and unrelated items least related) on false recognition judgments.

Participants

Thirty-eight undergraduates from Illinois State University participated for course credit.

Materials

Each participant completed 144 trials, divided into four blocks. Each block contained 12 memory sets for each set size, and each probe type was presented nine times.

Results and Discussion

RT Data

Data points with RTs that fell above or below 2.5 *SD*s for each subject were omitted from the analyses. This procedure eliminated 3.2% of responses. Alpha was set at .05 for all analyses in both experiments. Where multiple comparisons were conducted, a Bonferroni correction was applied.

Mean RTs[1] for correct responses are presented in Figure 1. A 3 × 4 repeated measures ANOVA indicated significant main effects of probe type, $F(2, 90) = 24.13$, $MSE = 19{,}931$, and set size, $F(2, 74) = 32.26$, $MSE = 18{,}808$.[2] The interaction approached significance, $F(3, 117) = 2.5$, $p = .058$, $MSE = 12{,}076$. The increase in RTs was steeper for CLs and studied-list items than for nonstudied-list items and unrelated probes. For the sake of brevity, we focus on the main comparisons of interest. Consistent with our predictions, correct rejections of CLs were slower than correct rejections of weakly associated items at all three set sizes, all p values < .008, indicating that the strength of association modulated latencies for correct responses. Correct rejections of weakly associated items and unrelated items did not differ at any set size, all p values > .05. These findings indicate that CLs are more likely than other nonstudied items to undergo the additional checking process and yield longer latencies.

In separate analyses, we compared RTs for list item hits and CL false alarms (see Figure 2[3]). As there were only four data points at set size three, the analysis was conducted only on set sizes five and seven. Data for 16 subjects were available for analysis, as a result of empty cells across subjects. False alarms were slower than hits, $F(1, 15) = 12.47$, $MSE = 85{,}398$. The effect of set size was significant, $F(1, 15) = 4.94$, $MSE = 67{,}663.94$, but the interaction was not, $F(1, 15) = 1.6$. Paired samples *t* tests, conducted with the full data set ($N = 38$, with subjects included when the set size cell was not empty), indicated that false alarms were slower than hits at set sizes five and seven, $t(20) = 4.12$ and $t(24) = 3.21$, respectively. At set size three, false alarms were marginally slower than hits, $t(3) = 2.96$, $p = .059$. As predicted by the activation-monitoring account, nonstudied items with high familiarity (e.g., the CL) were more likely to undergo the checking process than studied items, which presumably had a level of activation that was high enough to exceed the criterion for a faster *old* response. Thus, for both correct rejections and false alarms, the pattern of RTs predicted by two-process models was found.

Accuracy Data

Mean accuracy data (proportion of correct responses out of the total possible for each condition) are shown in Table 1. Correct responses were submitted to the same ANOVA as the RT data. All effects were significant: $F(2, 81) = 16.1$, $MSE = .02$, for probe type; $F(2, 55) = 29.3$, $MSE = .01$, for set size; and $F(2, 89) = 9.6$, $MSE = .02$, for the interaction. Accuracy decreased as a function of set size, but the decrease was more marked for CLs and studied-list items than for weakly related items and unrelated probes (which showed no effect of set size). Again, we focus on the comparisons of interest. False alarms to CLs occurred more often than false alarms to weak-

1 Reported degrees of freedom are corrected for sphericity (Greenhouse-Geisser).

2 Median RTs in both experiments were submitted to the same analyses with the same patterns of results found. Median RTs were on average faster than mean RTs, indicating that the differences in RTs were largely in the tail of the distribution. Analyses of RT distributions indicated that for hit and CL false-alarm distributions, approximately 15% of the responses were nonoverlapping. For CL and other nonstudied-item correct rejections, approximately 7% of the responses were nonoverlapping. These results indicate that a small, but significant, proportion of the RT processes for CL items are slower than processes for other items. Although the degree of overlap is substantial, we believe that, combined with the fact that median RTs analyses yielded the same results as mean RT analyses and that outliers had been trimmed, the additional processing of CLs does reflect the cognitive processes involved.

3 The figure displays data from all subjects, not just the 16 included in the analyses.

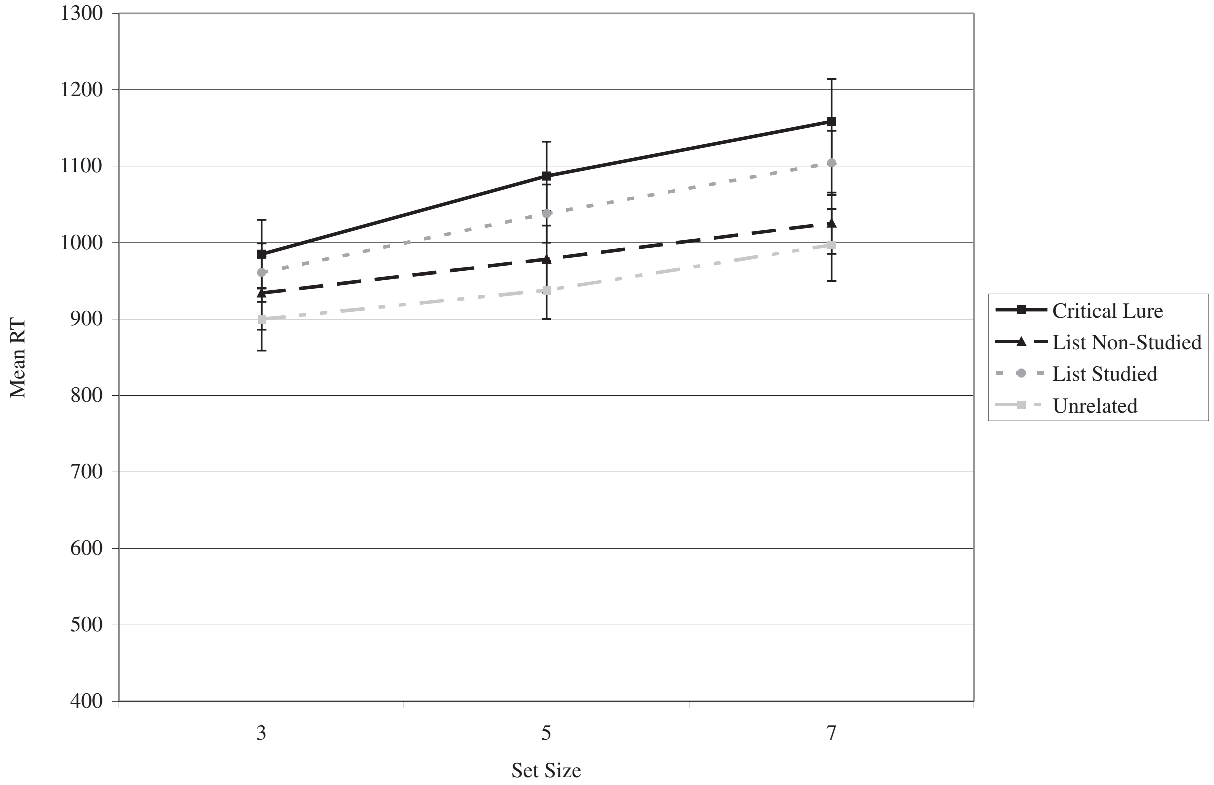

Figure 1. Mean correct response latencies as a function of probe type and set size in Experiment 1 (related lists).

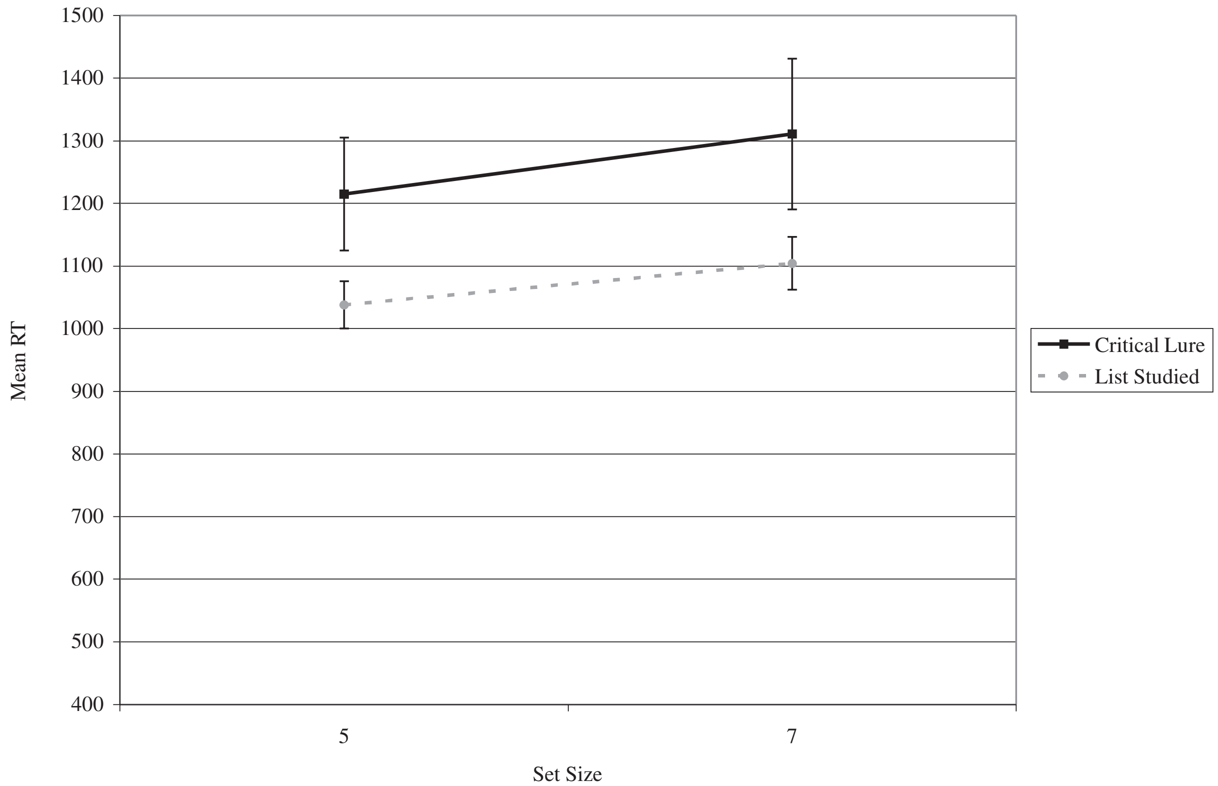

Figure 2. Mean hit and false alarm response latencies at set sizes five and seven in Experiment 1 (related lists).

Table 1. Accuracy as a function of probe type and set size in Experiment 1 (standard errors)

Probe type	Set size		
	3	5	7
Critical lure	.99 (.006)	.92 (.013)	.78 (.042)
Nonstudied list item	1.00 (.000)	.94 (.025)	.94 (.013)
Studied list item	.97 (.010)	.92 (.014)	.89 (.018)
Unrelated item	1.00 (.002)	.99 (.004)	1.00 (.002)

ly related items at set sizes five and seven, both p values ≤ .001, providing further support for the activation strength hypothesis. This difference approached significance at set size three, ($p = .058$). More false alarms were made to weakly-associated items than to unrelated items at set sizes five ($p = .068$) and seven ($p < .001$).

Summary

Experiment 1 indicates that false memories can occur in short-term memory tasks. As predicted, correct rejections of CLs were slower than correct rejections of weak associates and false alarms to CLs were consistently slower than hits to list items, indicating that additional processing was required for CLs because of their heightened activation.

Experiment 2

Experiment 2 was conducted to replicate the results of Experiment 1 and to test whether the results were the result of item differences. The CLs in the current study were high frequency items ($M = 128$ per million; Kucera & Francis, 1967), with numerous associates ($M = 66$, based on the free association norms of Nelson, McEvoy, & Schreiber, 1998) and rich connectivity networks. The average associative strength of list items to the CLs (connectivity) was 1.64 (obtained from Roediger, Watson et al., 2001). Thus, the results of Experiment 1 might have been due to item characteristics, as high frequency items tend to yield more false alarms than lower frequency items. To test this hypothesis, CLs were presented as probes after sets of related and unrelated items. We expected to replicate the findings of Experiment 1 for related lists, but not for unrelated lists, as responses to CLs in the latter should be no different than responses to other nonstudied probes.

Method

Participants

Seventy-three undergraduates participated for course credit. One participant's data were omitted because of a failure to follow instructions. Each participant provided six data points for each cell of the design, whereas each subject in Experiment 1 provided 12 data points per cell. Thus, twice as many subjects participated in Experiment 2.

Design and Materials

The 36 lists used in Experiment 1 were divided into two groups of 18 lists. One group was assigned to the related-lists condition, and the other to the unrelated-lists condition. The related-lists condition was the same as in Experiment 1, with the exception that the unrelated probe was not included. For the unrelated-lists condition, items were selected such that no two items within a set were from the same list. The studied probe was randomly selected from the set, the nonstudied probe was an unrelated word, and the CL was one of the 18 lures in that group with the constraint that none of its associates appeared in the set. The sets were counterbalanced across conditions, such that each list appeared in the related-lists and unrelated-lists conditions. In each block, half the sets were related, half unrelated.

Results and Discussion

RT Data

The trimming procedure used in Experiment 1 eliminated 3% of the data. Mean RTs for correct responses are shown in Figure 3. In a three-way ANOVA with probe type, set size, and list type as factors, all main effects were significant, $F(2, 110) = 16.5$, $MSE = 28{,}045$, for probe type; $F(2, 127) = 116$, $MSE = 15{,}500$, for set size; and $F(1, 70) = 83.7$, $MSE = 13{,}752$, for list type. The probe type by list type interaction was significant, $F(2, 107) = 59.4$, $MSE = 12{,}888$, as was the set size by list type interaction, $F(2, 140) = 15.5$, $MSE = 8{,}580$. Importantly, the three-way interaction was significant, $F(3, 227) = 5.2$, $MSE = 8{,}529$. We focus on the main comparisons of interest. For related lists, correct rejections of CLs were significantly slower than correct rejections of weakly related items at all set sizes, all p values < .001. However, for unrelated lists, correct rejections of CLs and correct rejections of unrelated items did not differ, all p values > .5. These findings indicate that the degree to which the probe is related to the items in the memory set influences latencies and that the effects are not the result of item-specific characteristics of CLs.

We also compared hit and false alarm RTs in the related-list condition (see Figure 4[4]). Because of empty cells at set size three, the analyses include data from set sizes five and seven ($N = 24$). Hits were faster than false alarms, $F(1, 23) = 5.98$, $MSE = 55{,}640$. Neither the effect of set size nor the interaction was significant. Hit and false alarm RTs were similarly affected by set size increases. Although

4 The figure displays data from all subjects, not just the 24 included in the analyses.

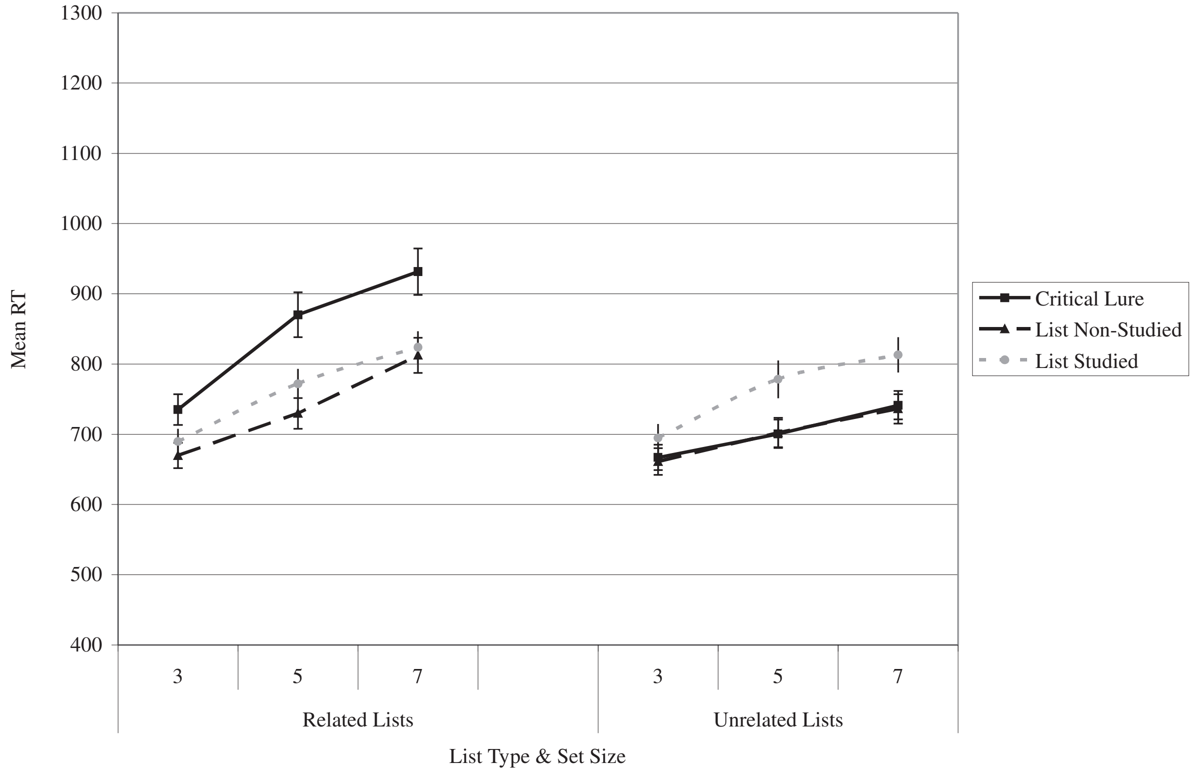

Figure 3. Mean correct response latencies as a function of probe type, set size, and list type in Experiment 2 (related and unrelated lists).

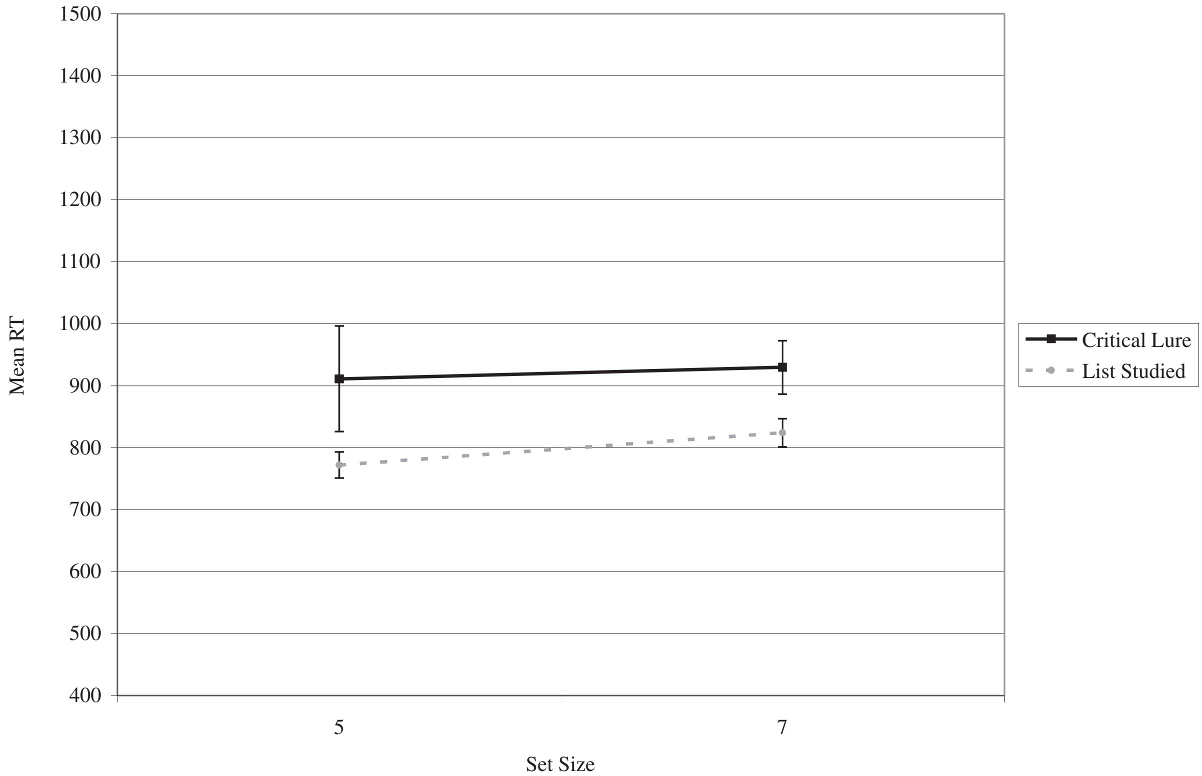

Figure 4. Mean hit and false alarm response latencies at set sizes five and seven in Experiment 2 (related lists).

no set size effects emerged in the subset of the data described above, *t* tests comparing set size effects for all 72 subjects indicated significant set-size effects for both list and CL items, $t(71) = -3.89$, $p < .001$ and $t(70) = -2.2$, $p = .031$, respectively.

Accuracy Data

Accuracy data (see Table 2) were submitted to the same ANOVA as latency data. All main effects and two-way interactions were significant, all *p* values ≤ .006. Importantly, the three-way interaction between probe type, set size, and list type was significant, $F(4, 284) = 6.0$, $MSE = .009$.

For related lists, planned comparisons revealed more false alarms to CLs than to weakly related items at set sizes five and seven, *p* values ≤ .005. At set size three, the difference approached significance, $p = .096$. The difference in false alarm rates as a function of degree of association to items in the memory set increased as set size increased, as in Experiment 1. However, when CLs were presented as probes after sets of unrelated items, they did not differ from other nonstudied unrelated probes. Overall, accuracy was greater for unrelated lists than for related lists. For unrelated lists, accuracy was significantly greater for CLs and nonstudied-items than for studied items, $p < .001$.

Table 2. Accuracy as a function of probe type, set size, and list type in Experiment 2 (standard errors in parentheses)

Probe type	Set size		
	3	5	7
Related lists			
Critical lure	.98 (.006)	.90 (.019)	.80 (.024)
Nonstudied list item	.99 (.003)	.96 (.009)	.93 (.010)
Studied list item	.98 (.008)	.94 (.013)	.90 (.016)
Unrelated lists			
Critical lure	.99 (.006)	.99 (.006)	.98 (.006)
Nonstudied list item	1.0 (.002)	.99 (.004)	.98 (.007)
Studied list item	.95 (.014)	.93 (.013)	.90 (.015)

Summary

The longer latencies and higher error rates for CLs do not appear to be caused by item characteristics. In the unrelated-lists condition, CLs did not differ from nonstudied items in either accuracy or latency data, whereas CL accuracy was lower and RTs longer in the related-lists condition. Consistent results for related lists in Experiments 1 and 2 also reduces the probability that participants were adopting particular response strategies, such as relying on the strong associative relationship between the CL probe and the set to facilitate their decision or guessing "yes" if unsure. If subjects had used relatedness in making a decision, it would result in one error (a false alarm to the CL), while making correct rejections of other probes easier (see Tun et al., 1998, for a similar argument). This strategy, however, might also yield more false alarms to weakly related probes. This did not occur, indicating either that subjects did not rely on relatedness of the probe to the memory set or that they were sensitive to differences in the degree of associative strength. Further, given the conservative balance of *old* to *new* responses implemented in both experiments, it is unlikely that the higher false alarm rate to CLs was the result of liberal response bias.

General Discussion

While previous studies of false memory have focused on long-term memory tasks, the current study applied the DRM paradigm to a short-term task. Results confirmed that RTs can be used to investigate processing of false memories and that false memories do occur in short-term tasks. The results are consistent with activation-monitoring accounts of false memory (Roediger, Balota, & Watson, 2001).

Short-Term Memory and the DRM

Although false alarm rates to CLs in the current study are lower than those observed in long-term recognition studies (60–70%; e.g., Roediger & McDermott, 1995), they were still significantly greater than false alarm rates to weakly associated or unrelated items. After study of as few as five to seven related items and a delay of approximately 1 s, subjects still incorrectly endorsed the CL as *old* 22% and 20% of the time for Experiments 1 and 2, respectively. These values are comparable to those obtained by Robinson and Roediger (1997) for nine-item lists in a free recall task given after a filled delay. Additionally, our false alarm rates are comparable to those observed in many long-term memory tasks for comparable list lengths. Thus, it appears that false memory effects are not exclusive to long-term memory tasks, but can be reliably obtained in short-term memory tasks where conditions should maximize accuracy.

Prior studies that examined the effects of relatedness on short-term memory reported that correct rejection latencies and errors increased when the probe was related to a single item in the memory set (e.g., Bartha et al., 1998; Jones & Anderson, 1982). In the current experiments, correct rejections of CLs were significantly slower than correct rejections of weakly related or unrelated probes. The degree of relation of the probe to items in the set influenced latency and accuracy. Although more false alarms were made to weakly related items than to unrelated items in Experiment 1, false alarm rates were still lower than those to CLs. Additionally, the results of the current study confirm that semantic information plays a role in how items are stored and retrieved in short-term memory (see Nairne, 1996).

RT Data and the DRM

Whereas accuracy has been the primary measure used to compare list and CL memories, some studies have employed additional measures to compare accurate and false memories. Remember/know judgments (Roediger & McDermott, 1995) have revealed similarities between list and CL items (but see Mather, Henckel, & Johnson, 1997). However, other studies revealed that brain activity patterns associated with false memories were distinguishable from those associated with veridical memory (Curran, Schacter, Johnson, & Spinks, 2001; Fabiani, Stadler, & Wessels, 2000). Thus, true and false memory can be dissociated, although not all measures may be sensitive to these differences.

Some studies have reported latency data in the DRM procedure, arguing that latencies may be more sensitive to differences in list and CL item processing. Jou et al. (2004) found that false alarm RTs were consistently slower than hit RTs. Tun et al. (1998) and Thomas and Sommers (2005), however, failed to find consistent differences in response latencies between false alarms and hits. In the study by Tun et al., participants took a free recall test after the presentation of each list, followed by a final recognition test during which the latency data were recorded. Prior recall of lists increases both true and false memories on a later recognition test (Gallo & Roediger, 2002; Roediger & McDermott, 1995). If participants had recalled the CL in the immediate recall test, the CL may have seemed more like a list item than an unstudied item in the final recognition test. Thomas and Sommers used a study format that is atypical for the DRM paradigm (paired associates). They reported that correct rejections of CLs were slower than correct rejections of unrelated fillers, consistent with our findings. Thus, the picture emerging from latency data in DRM studies using long-term memory tasks remains unclear. Our findings are most similar to those of Jou et al., whose procedure most closely matched a standard DRM study, and indicate that RTs to CLs do differ relative to other items.

Our experiments contribute to the literature on RT data in the DRM paradigm by adding data from a short-term memory task. RTs might provide insights into cognitive processing that are not available through other measures (Luce, 1986). Even when subjects endorse a CL as *old*, thus rendering it indistinguishable from a studied list item in terms of accuracy, differences between item types can still emerge in RTs. Furthermore, as correct rejections of CLs were slower than correct rejections of weakly related items, it appears that the activation converging on the CL increases its familiarity making it more difficult to discriminate, and engaging slower checking processes. Results indicating differences in RTs between CLs and weakly related items across set sizes support the hypothesis that CLs yield high false-alarm rates because they are strongly associated with list items, not because subjects are relying on thematic relatedness to endorse these items as *old*.

Theoretical Accounts

Slower correct rejections and false alarms for CLs are consistent with activation-monitoring theories of false memory. With more activation, familiarity for CLs increased and required more monitoring than weakly related or unrelated items. The increased processing time was apparent even at set size three, when accuracy was at ceiling. Monitoring processes should take longer when checking items in a larger memory set, thus yielding the set size increases for correct rejections of CLs. Future studies might test whether further increases in set size contribute more activation, thereby allowing the familiarity of the CL to exceed the criterion for studied items, resulting in RTs to CLs that are equivalent to hit RTs (a result yet to be found). However, larger sets would exceed most participants' short-term memory capacity, reducing accuracy as the result of difficulties in monitoring the content of the memory store. Future studies should examine latency data to further qualify these findings.

One interesting aspect of our results is the larger set size effect observed for CLs than for other probes. This finding is consistent with a scanning process: When scanning the study set, latencies for studied items are faster on average as the search stops once the item is found, whereas the CL requires a more exhaustive and slower search process. Similarly, Gallo (2004) found that when subjects could exhaustively recall all the items in a category from long-term memory, they were able to correctly reject a related probe. When the categories could not be exhaustively recalled (because they were larger), subjects were more likely to falsely recognize the related probe. Gallo's data complement and parallel those reported here. It is possible that the ability to use recallable information to reject a highly activated probe underlies errors in short- and long-term recognition tasks.

The slower correct rejections for CLs support the hypothesis that activation and monitoring processes acted in opposition, requiring extra processing time. Future studies might investigate the effects of shorter response deadlines (e.g., 500 ms) that would require subjects to rely primarily on familiarity for responses. If activation processes are responsible for the current results, false alarm rates should increase even at the smaller set sizes under such conditions.

The current study employed speed-accuracy instructions. Perhaps asking subjects to respond quickly artificially enhanced the false alarm rate. However, three lines of evidence indicate our results were not entirely the result of task instructions. First, overall accuracy was high. Second, we examined correlations between speed and accuracy across probe types and set size in Experiment 2 and only one correlation ($r = -.54$, $p < .001$, for correct rejections of CLs at set size three in the related-lists condition) attained significance. Third, RTs in Experiment 1 were slower overall than those in Experiment 2; however, the same pattern of results was observed. Thus, we believe the results do reflect different cognitive operations for memory judgments of CLs and other probes. Future studies might directly address this issue using unspeeded instructions.

Summary

The findings in the current study add to the previous RT data for false memories in the DRM paradigm. The current findings indicate that (1) despite the reduced number of related items in the lists (three, five, or seven) and the minimal study-test delays used, false memories did occur in a short-term memory task with DRM lists; (2) slower RTs to CLs were consistent with predictions of the activation-monitoring account; and (3) set size showed predicted effects on accuracy and latency data for list and CL items. Future studies may use latency measures to further discriminate the current theories of false memory effects.

Acknowledgments

We thank Dave Gallo, Jeffrey Wagman, Dave McCabe, and an anonymous reviewer for helpful comments on earlier drafts of the manuscript.

References

Atkinson, R.C., & Juola, J.F. (1974). Search and decision processes in recognition memory. In D.H. Krantz, R.C. Atkinson, R.D. Lucek, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology: I. Learning, memory, and thinking* (pp. 243–293). San Francisco: Freeman.

Bartha, M.C., Martin, R.C., & Jensen, C.R. (1998). Multiple interference effects in short-term recognition memory. *American Journal of Psychology, 111,* 89–118.

Curran, T., Schacter, D.L., Johnson, M.K., & Spinks, R. (2001). Brain potential reflect behavioral differences in true and false recognition. *Journal of Cognitive Neuroscience, 13,* 201–216.

Fabiani, M., Stadler, M.A., & Wessels, P.M. (2000). True but not false memories produce a sensory signature in human lateralized brain potentials. *Journal of Cognitive Neuroscience, 12,* 941–949.

Gallo, D.A. (2004). Using recall to reduce false recognition: Diagnostic and disqualifying monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 120–128.

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.

Jones, W.P., & Anderson, J.R. (1982). Semantic categorization and high-speed scanning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 237–242.

Jou, J., Matus, Y.E., Aldridge, J.W., Rogers, D.M., & Zimmerman, R.L. (2004). How similar is false recognition to veridical recognition objectively and subjectively? *Memory and Cognition, 32,* 824–840.

Kucera, H., & Francis, W.N. (1967). *Computational analysis of present-day American English.* Providence: Brown University Press.

Luce, R.D. (1986). *Response times: Their role in inferring elementary mental organization. Oxford Psychology Series* (vol. 8). New York: Oxford University Press.

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.

Nelson, D.L., McEvoy, C.L., & Schreiber, T.A. (1998). *The University of South Florida word association, rhyme, and word fragment norms*. http://www.usf.edu/FreeAssociation/.

Nairne, J.S. (1996). Short-term/working memory. In E.L. Bjork & R.A. Bjork (Eds.), *Memory,* (pp. 101–126). San Diego: Academic Press.

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 arousal of false memories. In H.L. Roediger & J.S. Nairne. (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, 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 and Review, 8,* 385–407.

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

Sternberg, S. (1966). High-speed scanning in human memory. *Science, 153,* 652–654.

Thomas, A.K., & Sommers, M.S. (2005). Attention to item-specific processing eliminates age effects in false memories. *Journal of Memory and Language, 52,* 71–86.

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.

Received September 19, 2005
Revision received February 21, 2006
Accepted February 22, 2006

Jennifer H. Coane

Campus Box 1125
Washington University in Saint Louis
One Brookings Drive
Saint Louis, MO 63130
USA
Tel. +1 314 935-6524
Fax +1 314 935-7588
E-mail jhcoane@artsci.wustl.edu

