

Running head: COMPARING DECAY RATES FOR FALSE MEMORIES

Comparing Decay Rates for Accurate and False Memories  
in the DRM Paradigm

Jorie M. Colbert<sup>1</sup> and Dawn M. McBride<sup>2</sup>

<sup>1</sup>University of Utah

<sup>2</sup>Illinois State University  
Department of Psychology  
Campus Box 4620  
Normal, IL 61790-4620  
[dmcbride@ilstu.edu](mailto:dmcbride@ilstu.edu)  
(309) 438-7146

### Abstract

Although previous studies have consistently reported different forgetting rates for true and false memory when tested with recall, studies comparing the rates of decay for true and false recognition have reported inconsistent results. The current study attempted to clarify this inconsistency by comparing forgetting rates for true and false recognition in addition to addressing methodological differences among the previous studies. Recognition of list items and critical lures was assessed at 6 delays (2 min, 5 min, 10 min, 15 min, 20 min, and 2 days). A comparison of power function slopes across item type provided evidence of slope differences such that false recognition decayed more quickly than true recognition, which is inconsistent with predictions from activation-source and fuzzy trace theories.

Comparing Decay Rates for Accurate and False Memories  
in the DRM Paradigm

False memories occur when a person either believes that an event happened, when in fact it did not, or distorts the memory of an actual event (Roediger & McDermott, 1995). A popular method for creating false memories experimentally was first used by Deese in 1959. Deese presented participants with lists that consisted of 12 words each. The words in each list shared an association with a word that participants never saw. This word was referred to as the critical lure. For example, the list words *thread, pin, eye, sewing, sharp, point, pricked, thimble, haystack, pain, hurt, and injection*, are thematically associated with the critical lure *needle*. At test, participants were instructed to recall as many words from the list as they could. Deese found that participants falsely recalled the critical lure for some of the lists. Deese's study provided a simple way to investigate false memories in an experimental setting.

In 1995, Roediger and McDermott revived Deese's (1959) paradigm and published a pivotal study on false memory. The purpose of their study was to (a) replicate Deese's finding of false memory in a free recall paradigm and (b) extend the paradigm by testing false recognition. Roediger and McDermott's (1995) study showed that false memories are important to study because they can occur at the same rate as, or in some cases exceed the rate of, memories for list items in the middle of a list of associated words. Testing false memories by presenting lists

of associated words is now referred to as the Deese-Roediger-McDermott (DRM) paradigm.

In attempts to better understand the connection between true and false memories, many researchers have used the DRM paradigm to make comparisons of accurate and false memories using a wide range of manipulations (see Brainerd & Reyna, 2005, for a review of research in this area). One such comparison has focused on decay rates for each memory type. Studies that have tested false recall with the DRM paradigm over varying retention intervals suggest that there is a difference between the rate of decay for false memories and the rate of decay for true memories (McDermott, 1996; Seamon, Luo, Kopecky, Price, Rothchild, Fung, & Schwartz, 2002; Thapar & McDermott, 2001; Toggia, Neuschatz, & Goodwin, 1999). Results from these studies have suggested that accurate recall performance declines more quickly than false recall with longer test delays. However, results from recognition studies have been inconsistent. Some studies that tested false recognition suggest that false memories decay more slowly than true memories (Payne, Elie, Blackwell, & Neuschatz, 1996; Thapar & McDermott, 2001), while others suggest that accurate and false memories decay at the same rate over time delays (Lampinen & Schwartz, 2000; Neuschatz, Payne, Lampinen, & Toggia, 2001; Seamon et al., 2002).

### *Theoretical Framework*

Two current theories can explain differences in decay rates between accurate and false memories: (a) fuzzy trace theory and (b) activation/source monitoring theory. According to fuzzy trace theory, items are encoded with verbatim representations and gist representations (Brainard, Reyna, & Mojardin, 1999; Reyna & Brainard, 1995). These representations are encoded simultaneously (Payne et al., 1996). With respect to false memories for lists of words from the DRM paradigm, verbatim representations are stored for words from the list. At the same time, a gist representation is stored for the association shared among the list words. At test, participants rely on verbatim traces when deciding if a list word was presented. However, they rely only on the gist representation when deciding if the critical lure was presented because it shares the association with studied list items, but there is no verbatim representation available for the unstudied critical lure.

An assumption of fuzzy trace theory is that gist representations last longer than verbatim traces due to the rapid decay of item specific details (Brainard, Reyna, & Brandse, 1995). It follows then that false memories will last longer than true memories, because false recognition relies less on verbatim traces (that are decaying rapidly) than accurate recognition does.

Activation-monitoring accounts of false memories (e.g., Roediger, Balota, & Watson, 2001) can also account for differences in forgetting rates for accurate and false recognition. When participants are presented with an item for study,

other related items are activated in semantic memory (Collins & Loftus, 1975). Because critical lures are related in meaning to the DRM list words, they become activated during the presentation of the related lists. At test, participants monitor the source of the words they recollect (Israel & Schacter, 1997). The source can either be internally generated (e.g. critical lures generated by the participant due to activation at study) or externally generated (e.g. presented words from the list). Internally generated sources (critical lures) have fewer item specific features (surface and perceptual details) than external sources (Johnson, Hashtroudi, & Lindsey, 1993). An assumption of the theory is that over time participants rely on more on semantic information because item specific details decay quickly, (Johnson et al., 1993). Thus, according to activation-monitoring accounts, source information for internally generated items will fade less quickly than external generated items. Therefore, results indicating that false memories will last longer than true memories over time are also consistent with activation-source monitoring theory.

Both theories of false memories, fuzzy-trace and activation/source-monitoring lead to a prediction that accurate memories decay more quickly than false memories across test delays. Thus, results from recall studies are consistent with predictions of these theories. The results of recognition studies have been inconsistent with respect to how false memories operate over time as compared to true memories. This inconsistency leaves open the question of whether false

recognition, like false recall, decays at a different rate than accurate recognition. Therefore, it is theoretically important to compare rates of decay for true recognition and false recognition.

#### *Methodological Issues for the Current Study*

One difference among the previous recognition studies that might account for the inconsistent decay results is the recognition sensitivity measure used to correct the recognition scores for bias. False recognition was found to decay more slowly than true recognition in Payne et al.'s (1996) study when no correction procedure was used to measure sensitivity and in Thapar and McDermott's (2001) study when  $d'$  was used to measure sensitivity. Contrary to these findings, true and false recognition were found to decay at similar rates in Seamon et al.'s (2002) study when  $d'$ ,  $A'$ , and  $Pr$  correction measures were used. Similar results were found in Neuschatz et al.'s (2001) study when an  $A'$  measure was used and in Lampinen and Schwartz's (2000) study when they used their own correction procedure.

Another issue with the previous studies that examined effects of delay on false recognition is that they reported different rates of false memories for similar retention intervals. False recognition rates at the initial interval ranged from 45% to 73% across the five false recognition studies. This wide range in false recognition might be due to differences in associative strength of each DRM list. To illustrate this issue, Stadler, Roediger, and McDermott (1999) conducted a

norming study of 36 lists and found false recognition rates between 27% and 84%. Additionally, a multiple regression study by Roediger, Watson, McDermott, and Gallo (2001) revealed that 55 lists varied from 10% to 65% with respect to the percentage of false memories that they elicited. The variability in false recognition rates across DRM lists was not consistently addressed in the five forgetting studies reviewed above.

Another methodological issue that might have contributed to the large range of false recognition rates in the past studies is the varying retention intervals that occur in the immediate test condition for each study. The immediate delay condition does not refer to the same delay interval across the studies, because each study used a different number of lists and a different presentation rate for those lists. Additionally, each study had different instructions to present to the participants, which would also lengthen the immediate interval.

Therefore, due to the mixed results of the previous recognition studies and the methodological differences across these studies, the purpose of the current study was to develop a more consistent methodology to assess false memories over test delays. In the current study, the methodological issues described above for the studies that compared accurate and false recognition rates were addressed. Specifically, we attempted to (a) compare forgetting rate results for two sensitivity measures of true and false recognition, (b) balance lists across delays

in terms of false recognition rates, and (c) accurately estimate retention intervals by calculating the time to complete each portion of the experiment,

In order to accurately estimate retention intervals, we summed up item study time, average item test time, and the amount of time between study and test to determine the study-test delay times. To address the second issue, we matched lists across delays according to initial rates of false recognition. Lists were grouped by *low*, *moderate*, and *high* rates of false recognition elicited in a pilot study. The lists were then assigned to delay conditions such that on average lists tested at each delay had similar initial rates of false memory.

Finally, in the current study two sensitivity measures used in previous studies were calculated to compare true and false recognition. The past studies used four different measures and taken together reported mixed results. Seamon et al. (2002) recommend the high threshold measure  $Pr$  or the signal detection measure  $d'$  when using the DRM procedure. Based on this recommendation, we calculated both  $Pr$  and  $d'$  scores for the current study to allow comparison with measures used in previous studies and to determine if there is any effect of sensitivity measure used on forgetting differences for true and false recognition.

#### *Outline of Present Study*

Specifically, we assessed several short delays and one long delay in order to capture the forgetting curve for each item type (i.e., list and lure items) in the DRM paradigm. A review of the previous studies in this area revealed a large

variation in false memory across the immediate delays. Therefore, in the current study false recognition and true recognition were assessed at five delays near that range (2 min, 5 min, 10 min, 15 min, and 20 min). Six delay intervals (those listed above and a 2 day delay) were chosen. This range of delays allows for estimate of the decay with the slope of the power functions fit to the data. The minute delays also cover a range of retention intervals not consistently tested in previous false recognition studies. Forgetting rates were estimated from the slope of the function fits to the recognition data.

Differences in forgetting rates for true and false recall are consistent with current theories of false memory (fuzzy trace and activation-monitoring theories). Therefore, to extend our understanding of the persistence of false memories over time and to further contribute to research comparing rates of decay for accurate and false recognition, we measured forgetting rates for each item type (list items/critical lures) in the DRM paradigm. Past studies that assessed false recognition examined interactions of delay time and item type in order to compare decay rates for list and lure items. This procedure can be influenced by initial rates of recognition at immediate delays (Loftus, 1985), and the current review showed that these rates varied from study to study. Thus, estimating decay rates by fitting forgetting curves will provide a more accurate estimate of the rate of decay for each word type and thus allow a more valid comparison. This method tests differences in decay rates for true and false memories by directly measuring

those rates and statistically comparing them. Therefore, the contribution of this study is that we tested false recognition with a more refined methodology to address methodological inconsistencies in past studies comparing true and false recognition.

## Method

### *Participants*

Participants were 74 students from the University of Utah. Participation fulfilled part of a course requirement for an Educational Psychology course. Of these, four participants did not follow instructions correctly and 10 participants did not attend the second session. Thus, 60 participants were included in the analyses. Participants were native English speakers, 18 years of age or older. Informed consent for both testing sessions was obtained in the first testing session.

### *Design & Materials*

*Study lists.* In a pilot study, 75 DRM lists were normed to determine their respective false recognition rates. The lists included in the pilot study were taken from Stadler et al. (1999) and McBride and Coane (2002) and shortened to the nine strongest associates. Each list consisted of nine words that were associated with a theme word (critical lure) that was not presented at study. Lists containing nine associates have been shown to elicit moderate to high rates of false memories (McBride & Coane, 2002; McBride, Coane, & Raulerson, 2006; Robinson &

Roediger, 1997). Additionally, the shorter list length allowed for shorter intervals to be assessed.

In the norming session lists were tested in blocks of five lists (study of five lists and test of those five lists). This method tested lists at a delay of approximately 2 min for all lists, thus, allowing measurement of the false recognition rate for each list at a level similar to initial performance levels in the current study.

Based on the rate of false recognition they elicited in the norming session, the lists were placed into three list strength groups (*low*, *moderate*, *high*). Lists that elicited the 14 highest false recognition rates (51% - 73%) were placed in the *high* group. Lists that elicited the 14 lowest false recognition rates (11% - 29%) were placed in the *low* group. The seven lists below the median and the seven lists above the median were placed in the *moderate* group (40% - 45%). Three of the remaining lists from the norming session that were not placed in one of these three groups were randomly chosen to be filler lists during the study phase. These lists were used as fillers for both study sets (Sets A and B). The Appendix includes all of the lures from lists that were used in the current study and the rates of false recognition these lists generated in the pilot study.

Each of the 42 target lists in the study was assigned to one of two list sets (Set A or B). Each set contained 7 lists from each list strength group (i.e., *low*, *moderate*, *high*). Within each set, one list from each rate group was tested at each

of the minute delays (2, 5, 10, 15, or 20 min) and two lists from each rate groups were tested at the 2 day delay. Thirty of the participants were randomly assigned to view list Set A during the study phase, while the other 30 were randomly assigned to view list Set B. Items from the list set not studied served as unstudied list and lure items in the recognition test.

*Recognition tests.* Each recognition test assessed memory for two lists at a time. Within each test, items from one of the lists had been viewed in the study phase (studied list), while items from the other list were new to participants (unstudied list). Each recognition test contained a total of eight items: three words from the studied list (items from study positions one, four, and nine), three words from the unstudied list (also from positions one, four, and nine within the list), the critical lure for the studied list, and the critical lure for the unstudied list. The two lists included in each recognition test were matched in terms of the rate of false memory (i.e., list strength) they elicited from the pilot study (*low*, *moderate*, or *high*). For the 2 min to 20 min delays, participants were tested on three studied lists at each delay (one list from each of the *low*, *moderate*, and *high* rate groups). For the 2-day delay, each participant was tested on two studied lists from each false memory rate group.

Before each recognition test was presented, a set of math problems was given to fill the delay. In addition, before 10 min delay tests, a recognition test was given that was comprised of unrelated and unstudied (i.e., filler) words. The

filler recognition test was incorporated to shorten the math tests that filled the time delay before the longer delay (10 min, 15 min, and 20 min) tests.

### *Procedure*

Table 1 contains a detailed description of the procedure that was followed during study and testing of minute delay lists in each experimental session. Each participant completed two experimental sessions. Session 2 took place 2 days after Session 1. Participants were tested in small groups (1 to 4 people) for each session. At each session, each participant sat in front of a computer and viewed instructions that indicated they would be viewing lists of words and that their task was to remember the words as best they could for later memory tests. Study words were presented one at a time for 2 s each. In the instructions it was noted to participants that they would need to stay focused on the computer screen because words would be presented at a fast rate.

The study lists for the minute delays were presented in three blocks of six lists each (5 target lists for the delays and one non-target filler list that was not tested). Lists were presented successively with no indication when a new list began. Each block began with three unrelated words and the non-target filler list that was not tested to combat primacy effects. The block then included a list for each min delay (i.e., five lists). Three unrelated words were included at the end of the block to combat recency effects (see the Study Phase in Table 1).

After the presentation of the study phase, a block of recognition tests for the minute delays was given (each recognition test contained a studied list and an unstudied list, as described in the Materials section). Each recognition test was preceded by filler math problems. Immediately following the last word in a study block, instructions for the math problems appeared on the computer screen for 30 s asking participants to complete simple math problems. Math problems were included to fill the time between study and test to allow recognition tests to occur at the delays described above. Participants completed math problems for a range of times before each recognition test. See Table 1 for details of the testing phase.

At the end of each math problems set, instructions for the next recognition test appeared on the computer screen for 30 s. Participants were instructed to make an *old* (the word was presented at study) or *new* (the word was not presented at study) judgment for each recognition item that appeared on the screen. On the computer keyboard, the participants pressed “Y” for *old* and “N” for *new*. For each recognition test participants made judgments for eight words. Each word was presented one at a time on the computer screen, and participants had 5 s to make their judgment. The same procedure (math instructions, math problems, recognition instructions, recognition test) was repeated five times within the testing phase with varying math set times to allow recognition tests for the specific minute delays (see Table 1 for detailed timing of the testing phase).

After the last recognition test in the first testing block, participants viewed instructions on the computer screen informing them that a new study block would be presented. The procedure outlined above was then repeated with a second block of lists (i.e., the whole procedure presented in Table 1 was repeated with new lists in Session 1). No words from the first block appeared in the second block of study and test items. Participants were instructed that recognition tests in the second block would not contain items from the first block.

At the end of Session 1, participants viewed six lists that they were instructed to remember for Session 2. At the start of Session 2, participants were tested on these six lists presented at the end of Session 1. The same procedure from Session 1 was then repeated, but with only one block of lists (i.e., the Study and Test Phases in Table 1 were administered at Session 2 with a new set of lists not studied or tested in Session 1).

A total of six delays were assessed: 2 min, 5 min, 10 min, 15 min, 20 min, and 2 days. The test delays were determined by (a) calculating the time it took for participants to complete the recognition test and dividing that time in half to determine the average item time, (b) calculating the time it took for participants to study the list and dividing that time in half to determine the average item time, (c) determining the instruction time for the recognition test (30 s), and (d) determining the math problems block time for that list (see Table 1 for times used in these calculations). These four values were summed to determine the retention

delay for each list. For example, the first recognition test after each block assessed the 2 min delay. The time to present each list was 18 s. Therefore, on average, the study delay for each list item in the study portion is half of this time (9 s). To calculate the test delay for the last list (2 min list) of each study block, 9 s (average delay for an item in that list from the study portion) was added to the time it took to present the three unrelated words at the end of the block (6 s). This time was also added to the time that instructions for the math problems (30 s) and recognition test instructions (30 s) appeared on the screen in addition to the time participants completed the first set of math problems (25 s). Finally, half of the time it took to complete the recognition test for that list was added (20 s). Thus, a total of 120 s or 2 min elapsed from the midpoint of the 2 min list in the study block to the midpoint of its respective recognition test.

## Results

### *Preliminary Analyses*

The hit and false alarms rates were calculated for each subject for list items and critical lures by delay condition and list strength condition. For true recognition, a hit was correctly identifying a studied list items as *old* and a false alarm was incorrectly identify an unstudied list item as *old*. For false recognition, a hit was incorrectly identifying a critical lure for a studied list as *old* and a false alarm was incorrectly identifying a critical lure for an unstudied list as *old* (these definitions of hits and false alarms have been applied by previous researchers in

this area, e.g., see Miller & Wolford, 1999). Thus, false recognition was determined by responses to critical lure items when the accompanying list had been studied and when it had not been studied. These mean hit and false alarm rates are reported in Table 2.

Both  $Pr$  and  $d'$  scores were calculated from hit and false alarm rates for each individual subject. Mean  $Pr$  and  $d'$  scores are presented in Tables 3 and 4 respectively. For each subject, a  $Pr$  score was calculated for each delay by list strength (*low, moderate, high*) condition. For true recognition,  $Pr$  scores were calculated by subtracting the false alarm rate for the unstudied list items from the hit rate for the studied list items for each recognition test. Thus, a participants' true recognition  $Pr$  score for each list ranged from -1 (responding *old* to all three unstudied list items and *new* to all three studied list items) to 1 (responding *new* to all three unstudied list items and *old* to all three studied list items). For false recognition,  $Pr$  scores were calculated by subtracting the false alarm rate for critical lures from the unstudied lists from the false alarm rate for critical lures from the studied lists. Thus, a participants' false recognition  $Pr$  score for each list ranged from -1 (responding *old* to the unstudied list lure and *new* to the critical lure from the studied list) to 1 (responding *new* to the unstudied list lure and *old* to the critical lure from the studied list). Mean  $Pr$  scores for list items and critical lures and the standard errors of the  $Pr$  means for each delay (averaged across list strength condition) are shown in Table 3. Table 3 illustrates that mean accurate

recognition was higher than mean false recognition at each delay. In addition, recognition rates for each item type decreased with test delay.

$d'$  scores were calculated by subject for each retention interval with the following formula:

$$d' = z(M_{\text{hit}}) - z(M_{\text{false alarm}})$$

Mean hit and false alarm rates were calculated by averaging hits and false alarms across strength conditions for each subject. A correction for 0 and 1 hit and false alarm values was applied to the subject data before calculations were made.

$1/(2N)$  was substituted for values of 0, and  $1-1/(2N)$  was substituted for values of 1.0 (Macmillan & Creelman, 2005), where  $N$  equals the number of trials the data point was based on. The mean  $d'$  values for false and true recognition at each delay are presented in Table 4. Table 4 illustrates that true recognition was higher than false recognition at each delay. In addition, recognition rates for both item types decreased with retention interval.

To confirm the description of the data presented above and to allow comparison with previous studies that used this analysis, ANOVAs were conducted on average  $Pr$  and  $d'$  scores separately with Delay (2 min, 5 min, 10 min, 15 min, 20 min, 2 day) and Item Type (critical lure, list item) as within-subjects factors. For  $Pr$  scores, the main effect of item type was significant,  $F(1,59) = 164.20, p < 0.001$ , indicating higher recognition scores for list items than for critical lures (see Figure 1). The main effect of delay was also significant,

$F(5,295) = 45.89$ ,  $p < 0.01$ . Finally, the interaction of item type and delay was significant,  $F(5,295) = 3.15$ ,  $p < 0.05$ , indicating that accurate and false memories may decline at different rates. For  $d'$  scores, all effects were significant as well. The main effect of item type indicated higher recognition scores for list items than for critical lures,  $F(1,59) = 234.75$ ,  $p < 0.001$ . The main effect of delay was also significant,  $F(5,295) = 47.22$ ,  $p < 0.001$ . Finally, the interaction of item type and delay was also significant,  $F(5,295) = 3.69$ ,  $p < 0.005$ , indicating that accurate and false memories may decline at different rates. However, estimating decay rates directly from function fits provides a more accurate comparison of the decay rates for list and lure items. Therefore, the primary analyses were the function fits to the data (described below).

### *Function Fits*

To compare decay rates for true and false recognition, 2-parameter power and 2-parameter exponential functions were fit to  $Pr$  means for each item type and  $d'$  means for each item type (see Tables 3 and 4, respectively).<sup>1</sup> Power functions provided the best fits to the data overall so we focus on these analyses below. In addition, several studies have indicated that the power function best describes a range of forgetting data across different tasks (e.g., Rubin & Wenzel, 1996; Wixted, 2004; Wixted & Ebbesen, 1991) also found the power function to provide a better fit of recall and recognition data as compared to several other curvilinear functions. The power function fit to the data was of the form:

$$P(t) = \lambda t^{-\beta}$$

In this function,  $P(t)$  is the probability of obtaining a particular sensitivity score at time  $t$ ,  $\lambda$  is the initial rate of encoding,  $\beta$  is the slope, and  $t$  is the time delay. For each sensitivity measure ( $Pr$  and  $d'$ ), separate power functions were fit to data for each item type. The slopes were then compared by nested model  $F$  test (i.e., comparing the change in fit if a single slope is assumed for both item types). Parameter estimates and  $R^2$  values for power functions fit to data that included all delays are provided in Table 5.<sup>2</sup>

Overall, power functions fit the data for each item type well. As seen in Table 5,  $R^2$  values ranged from 0.67 to 0.93 and were lower for lure item fits than for list item fits. Fits to data for all delays indicated significantly different slopes for list and lure items for both sensitivity measures,  $F(1,8) = 8.65, p = 0.019$  for  $Pr$  data and  $F(1,8) = 5.47, p = 0.048$  for  $d'$  data. To confirm that slope differences were not due to the inclusion of the 2-day delay (where scores dropped considerably from the 20 min delay to the 2-day delay, (see Tables 3 and 4), fits were also conducted with the 2-day delay condition removed (i.e., power functions were fit to data that included only the minute delays). Results were similar for these fits without the 2-day delay: Significant slope differences were found for both  $Pr$  and  $d'$  measures,  $F(1,6) = 15.82, p = 0.007$  for  $Pr$  data and  $F(1,6) = 18.21, p = 0.005$  for  $d'$  data.

The power function fits are illustrated in Figure 1 for mean *Pr* scores. Functions overlaid are best-fit power functions. Parameter details of these fits are provided in Table 5. Test delays in Log 10 form are plotted on the x-axis. Mean *Pr* scores are plotted on the y-axis. Figure 1 illustrates that false recognition declined more quickly ( $\beta = -0.59$  with all delays) than accurate recognition ( $\beta = -0.15$  with all delays). The same results were found for  $d'$  function fits. The implications of these results for theories of false memory are discussed below.

Finally, to examine the effect of list strength on the slope difference results, separate power function fits were conducted for *Pr* data for list strength condition (we chose to focus on the *Pr* scores for rate group fits due to the small cell sizes; lure hit and false alarms by list strength group were based on a single observation). Power functions were fit as described above with slopes compared by nested model comparison where a single slope was assumed for both item types. In these analyses, significant slope differences were found only for the high strength lists,  $F(1,8) = 7.78, p = 0.02$ . The slope comparison for the moderate strength lists failed to reach significance,  $F(1,8) < 1.0, p = 0.59$ , as did the comparison for the low strength lists,  $F(1,8) = 2.81, p = 0.13$ . These results indicate that slope differences reported above were driven primarily by rate differences in the high strength lists. Implications of these results for previous studies in this area are discussed below.  $R^2$  values ranged from 0.46 to 0.95 (with highest  $R^2$  values for the high strength group). It should also be noted that

although non-significant results were found for low strength and moderate strength lists, the difference between the accurate recognition and false recognition slopes were in the same direction as the difference for high strength lists. Specifically, the slope parameter values estimated for false recognition were numerically greater (and thus, indicated a faster decline in recognition rate) than the slope values for accurate recognition. However, conclusions from these analyses should be made with caution, as our goal in this study was to balance list strength across retention interval and the separate analyses by list strength group described here reduce the number of observations available for function fits by one-third, thus, reducing power to detect differences.

#### Discussion

In the current study, decay rates for accurate and false recognition were compared for a range of delays with power function fits used to estimate and compare the slopes of the functions for list and critical lure items in the DRM paradigm. A comparison of power function slopes across item type (list and lure) provided evidence of slope differences such that false recognition decayed more quickly than true recognition. These results were consistent across analyses conducted for both  $Pr$  and  $d'$  sensitivity measures and for both analyses that included and excluded the 2-day delay where negative sensitivity scores were found for false recognition (see Tables 3 and 4). However, in analyses conducted separately by list strength group, slope differences were only found in the *high*

strength group, indicating that these lists may have driven the significant slope differences found in the overall analyses.

An issue regarding the current results that requires further discussion is the finding of negative sensitivity scores for false recognition in the 2-day delay. While the results did not differ when fits were conducted without from this data point included, it is nonetheless an unusual finding for false recognition in the DRM paradigm. The negative scores are a result of higher mean false alarms to critical lures from unstudied lists as compared to mean false alarms to list items from unstudied lists. Hits rates showed the opposite pattern: Hits to critical lures from studied lists were much lower than hits to list items from studied lists. The mean false alarm score to critical lures from unstudied lists at the 2-day delay is a clear outlier, especially for *low* and *moderate* strength list (see Table 2). Hits rates to critical lures from studied lists were also elevated for this delay condition as compared to the 20 min delay.

One possible explanation for the negative scores may be that with fewer observations per cell, the false recognition scores are more variable than those for true recognition and the negative scores resulted from random error. However, a more likely cause is the difference in study and testing of lists at the 2-day delay as compared to lists assigned to the minute delays. Subjects studied items for the 2-day delay at the end of Session 1 and were tested on these items at the beginning of Session 2. This procedure likely set these lists apart from the others

that subjects encountered and may have influenced them to encode or retrieve these items with a different strategy than the items assigned to the minute delays. Subjects were quite familiar with the procedure when the study lists for the 2-day delay were presented. They may have realized that the study lists had a thematic focus and mistakenly believed they had studied some of these theme items (i.e., critical lures). This familiarity may have increased the likelihood they would respond *old* to critical lure items and resulted in more *old* responses overall to critical lures. Alternatively, because the 2-day delays lists were tested in the last testing block in the experiment, subjects may have been less likely to respond *old* to lures that were consistent with studied list themes than to lures that corresponded to unstudied lists, resulting in negative *Pr* and *d'* scores for lures assigned to this delay condition.<sup>3</sup> The lists assigned to the 2-day delay were also subject to possible interference or rehearsal effects as a result of the time between the two sessions. The unusual sensitivity scores for the 2-day delay did not affect results in the current study. Nonetheless, future studies in this area should consider these factors when testing false memory with longer delays.

#### *Comparison with Past Studies*

As described in the Introduction, the recognition studies that have compared forgetting for list and lure items in the DRM paradigm have varied on several dimensions: (a) different sensitivity measures were used, (b) immediate delay conditions varied in length, and (c) lists varying in false recognition rates

were assigned to different delay conditions. The current study attempted to establish a more optimal method (with respect to these issues) with which to compare forgetting rates for true and false recognition. Therefore, methodological differences between the current study and each of the previous studies exist and may have contributed to the difference in results across (the current and previous) studies. The current study found a faster decline in false recognition than in true recognition across retention delays of 2 min to 2 days, while Lampinen and Schwartz (2000), Neuschatz et al. (2001), and Seamon et al. (2002) each reported findings indicating similar declines in true and false recognition rates across time delays. In their studies, Payne et al. (1996) and Thapar and McDermott (2001) found forgetting differences between true and false recognition. However, in contrast to the results of the current study, their findings indicated a slower decline for false recognition than true recognition. Methodological differences may be the cause for the different results between these studies and the current study.

Studies testing recall have consistently reported decay rate differences in true and false recall (McDermott, 1996; Seamon et al., 2002; Thapar & McDermott, 2001; Toggia et al., 1999). However, the current findings indicate forgetting rate differences for true and false recognition that are in the opposite direction and those reported for true and false recall. Differences in processing during retrieval in each of the tasks (recall and recognition) may also be a cause

of the inconsistent results across recognition and recall. There is evidence that items generated during recall in the DRM paradigm may influence false recall of lures, while recognition items seem to have small effects (at least much smaller than study effects) on false recognition of lures. Roediger and McDermott (1995) found that lures tended to appear later in the response output in free recall. In recognition, however, testing effects (i.e., likelihood of false recognition of lures following recognition of related studied list items) appear to be small (Coane & McBride, 2006; Dodd, Sheard, & MacLeod, 2006; Marsh, McDermott, & Roediger, 2004). It is possible that these differences in testing effects across recall and recognition result in very different forgetting functions for false recall and false recognition. Further discussion of this point in the context of theories of false memory is presented below.

### *Theories of False Memory*

The results of the current study seem inconsistent with current accounts of the creation of false memories (i.e., fuzzy trace and activation-monitoring theories). In fuzzy trace theory (Reyna & Brainerd, 1995) it is assumed that memory for list items relies primarily on verbatim traces, while memory for critical lures relies primarily on gist traces. In activation-source monitoring theory (Roediger et al., 2001), it is assumed that critical lures are activated along with list items at study, increasing familiarity for both types of items. At test, source

judgments allow subjects to correctly remember list items and incorrectly remember critical lures as having been studied.

Results reported in previous recall studies indicating that false memories decay less quickly than true memories (McDermott, 1996; Seamon et al., 2002; Thapar & McDermott, 2001; Toggia et al., 1999) seem consistent with these accounts. This is primarily due to the assumption that item-specific details decay more quickly over time than relational representations. According to these theories, participants rely primarily on relational representations to determine if the critical lure is old or new. Given this explanation of false memory effects, participants' false memories should decay more slowly over time than their true memories. This prediction has been consistently supported in DRM studies testing recall, but results from recognition studies have been mixed. The findings of similar decay rates reported by Lampinen and Schwartz (2000), Neuschatz et al. (2001), and Seamon et al. (2002) do not support the fuzzy trace and activation-source monitoring theory predictions for recognition tests. Further, the results in the current study of faster decay of false than true recognition are at odds with the theories' predictions.

Processing differences across recall and recognition (as described above) may be one possible explanation for the lack of support for the theories' predictions regarding decay of false recognition. Some DRM studies have found differences between false recall and false recognition. Rates of false recall and

false recognition tend to differ, with false recognition occurring more frequently than false recall (Roediger & McDermott, 1995). In addition, as described above, Seamon et al. (2002) reported different results for recall and recognition in comparisons of forgetting for true and false memory. Neither of the false memory theories (activation-source monitoring or fuzzy trace) account for possible differences across memory tasks and thus, do not account for differences in false recognition and false recall that have been reported.

It may also be the case that item-specific representations are more important in recognition judgments of critical lures than they are in the generation of critical lures as responses in recall tests. If item-specific information is relied on more heavily for lure judgments in recognition tests than in recall tests, it could result in faster decay rates for false than true recognition, while false recall decays more slowly than true recall because the faster decaying item-specific information is less important in recall of critical lures. There is some evidence to support this idea. Several studies have reported that subjects will endorse lures as *remembered* in remember/know recognition tasks (Roediger & McDermott, 1995) and will claim to remember a presentation source (e.g., voice gender in auditory presentations) for critical lures (Gallo, McDermott, Percer, & Roediger, 2001). Further, Gallo et al. found that reductions in false memory with visual study presentation (as compared to auditory presentation) were larger for recall tests than recognition tests, indicating that false memories in recognition tests may be

caused by subjects' reliance on source details of the list presentation. Hicks and Starns (2006) also provided evidence for the importance of source information in recognition judgments in a recent study that examined causes of the source-strength effect. This effect is produced in DRM studies when subjects report source judgments (e.g., presentation voice gender) for the critical lure that are consistent with the more strongly associated items in the list. Their study indicated that subjects were falsely remembering sources for the critical lures as opposed to inferring a source for critical lures from their memory of the source of remembered list items. These results illustrate the importance of source information in false recognition and may suggest an explanation for the different findings reported for forgetting in false recall and false recognition. Therefore, separate predictions may need to be made by both theories of false memories to account for effects of certain factors (e.g., retention interval) on false recall and false recognition since item-specific information is important for verbatim memory in fuzzy trace theory and source judgments in activation-source theory.

### *Summary and Conclusions*

To summarize, in contrast with previous studies that compared true and false recognition, results of the current study indicate that false recognition declines at a faster rate than true recognition. The difference in results between the current study and past studies may be due to methodological differences. In the current study, an attempt was made to provide an accurate test of forgetting

differences between list and lure items in the DRM paradigm by (a) testing several retention intervals and fitting curvilinear functions to directly compare slopes for list and lure items, (b) counterbalancing list strength across retention interval to remove (and test) effects of list strength on forgetting rate comparisons, (c) including more than one sensitivity measure ( $Pr$  and  $d'$ ) to compare results across measures, and (d) calculating accurate average item retention intervals tested. The results of the current study are also inconsistent with those in previous studies that found accurate recall to decline more quickly than false recall. Current theories of false memory (activation-source monitoring and fuzzy trace) predict a faster decline in true memory than false memory. Results from recall studies are consistent with this prediction, while the results from the current study stand in contrast to the theories' prediction. Future researchers may need to consider the possibility that false recall and false recognition are influenced by different processes and thus, different predictions may need to be made for false memory across these tasks.

## References

- Brainerd, C. J., & Reyna, V. F. (2005). *The science of false memory*. New York: Oxford University Press.
- Brainerd, C. J., Reyna, V. F., & Brandse, E. (1995). Are children's false memories more persistent than their true memories? *Psychological Science, 6*, 359-364.
- Brainerd, C. J., Reyna, V. F., & Mojardin, A. H. (1999). Conjoint recognition. *Psychological Review, 106*, 160-179.
- Coane, J. H., & McBride, D. M. (2006). The role of test structure in creating false memories. *Memory & Cognition, 34*, 1026-1036.
- Collins, A. M., & Loftus, E. F. (1975). Spreading-activation theory of semantic memory. *Psychological Review, 82*, 407-428.
- Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions on immediate recall. *Journal of Experimental Psychology, 58*, 17-22.
- Dodd, M. D., Sheard, E. D., & MacLeod, C. M. (2006). Re-exposure to studied items at test does not influence false recognition. *Memory, 14*, 115-126.
- Gallo, D. A., McDermott, K. B., Percer, J. M., & Roediger, H. L. III. (2001). Modality effects in false recall and false recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 339-353.

- Hicks, J. L., & Starns, J. J. (2006). The role of associative strength and source memorability in the contextualization of false memory. *Journal of Memory and Language, 54*, 39-53.
- Israel, L., & Schacter, D. L. (1997). Pictorial encoding reduces false recognition of semantic association. *Psychonomic Bulletin & Review, 4*, 577-581.
- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*, 3-28.
- Lampinen, J. M., & Schwartz, R. M. (2000). The impersistence of false memory persistence. *Memory, 8*, 393-400.
- Loftus, G. R. (1985). Evaluating forgetting curves. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 397-406.
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide*. New Jersey: Lawrence Erlbaum Associates.
- 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.
- McBride, D. M., & Coane, J. (2002). *False memory norms*. Unpublished manuscript, Illinois State University.
- McBride, D. M., Coane, J. H., & Raulerson, B. A. III. (2006). An investigation of false memory in perceptual implicit tasks. *Acta Psychologica, 123*, 240-260.

- McDermott, K. B. (1996). The persistence of false memories in list recall. *Journal of Memory & Language, 35*, 212-230.
- Miller, M. B., & Wolford, G. L. (1999). Theoretical commentary: The role of criterion shift in false memory. *Psychological Review, 106*, 398-405.
- Neuschatz, J. S., Payne, D. G., Lampinen, J. M., & Toggia, M. P. (2001). Assessing the effectiveness of warnings and the phenomenological characteristics of false memories. *Memory, 9*, 553-571.
- Payne, D. G., Elie, C. J., Blackwell, J. M., & Neuschatz, J. S. (1996). Memory illusions: Recalling, recognizing, and recollecting events that never occurred. *Journal of Memory & Language, 35*, 261-285.
- Reyna, V. F., & Brainard, C. J. (1995). Fuzzy trace theory: An interim synthesis. *Learning and Individual Differences, 7*, 1-75.
- 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, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 21*, 803-814.
- Roediger, H. L., III, Balota, D. A., & Watson, J. M. (2001). Spreading activation and arousal of false memories, In H. L. Roediger III, J. S. Nairne, I. Neath, & A.M. Suprenant (Eds.), *The nature of remembering: Essays in*

- honor of Robert G. Crowder* (pp. 95 – 115). Washington, DC: American Psychology Association.
- 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.
- Rubin, D. C., & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review*, *103*, 734-760.
- Seamon, J. G., Luo, C. R., Kopecky, J. J., Price, C.A., Rothchild, 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.
- Stradler, M. A., Roediger, H. L., III, & McDermott, K. B. (1999). Norms for word lists that create false memories. *Memory & Cognition*, *27*, 494-500.
- Thapar, A., & McDermott, K. B. (2001). False recall and false recognition induced by presentation of associated words: Effects of retention interval and level of processing. *Memory & Cognition*, *29*, 424-432.
- Toglia, M. P., Neuschatz, J. S., & Goodwin, K. A. (1999). Recall accuracy and illusory memories: When more is less. *Memory*, *7*, 233-256.
- Wixted, J. T. (2004). The psychology and neuroscience of forgetting. *Annual Review of Psychology*, *55*, 235-269.

Wixted, J. T., & Ebbesen, E. B. (1991). On the form of forgetting. *Psychological Science*, 2, 409-415.

## Footnotes

<sup>1</sup>Functions were not fit to individual subject data due to small cell sizes. Individual subject fits resulted in unstable fits with highly variable  $R^2$  values.

<sup>2</sup>Negative values were replaced with 0 for the purposes of power function fits. However, negative values only occurred in the 2-day delay condition for lures (see Tables 3 and 4) and slope comparisons were significant in fits both with and without this delay condition included.

<sup>3</sup>We thank John Wixted for this suggestion.

### Author Notes

The authors wish to thank Anne Cook for her assistance in setting up the computer program for the experiment, Jennifer Coane for helpful discussion, and Jeff Wagman, John Wixted, and two anonymous reviewers for helpful comments on earlier versions of the manuscript. The authors also wish to thank Tim Rarrick for his help in data collection.

## Appendix

## False Recognition Rates of DRM Lists Organized by Group with the Critical

## Lures Representing Each List

Low Rates of <u>False Recognition</u>	Moderate Rates of <u>False Recognition</u>	High Rates of <u>False Recognition</u>
Deep (11 %)	Music (40%)	Stove (51%)
Lord (12%)	Dog (40%)	Heavy (51%)
Summit (15%)	School (41%)	Soldier (52%)
Hammer (15%)	Weather (41%)	City (53%)
Black (21%)	Bread (42%)	Trash (53%)
Girl (21%)	Lion (42%)	Man (54%)
Chair (23%)	River (42%)	Smoke (55%)
Kill (24%)	Baby (42%)	Short (56%)
Right (25%)	Carpet (43%)	Smell (56%)
Spider (26%)	Hungry (44%)	Holiday (61%)
Butterfly (27%)	Work (44%)	Flower (62%)
Whistle (28%)	Tests (44%)	Movie (65%)
Street (29%)	Clean (44%)	Cold (70%)
Green (29%)	Eagle (45%)	Sleep (73%)

Filler Lists: Earth (31%), Window (42%), Sweet (46%)

Table 1

*List Presentation and Recognition Test Procedure for Minute Delays (Target lists are indicated in italics)*

Procedure Order	Time to Present
<i>Study Phase (5 target lists, one for each delay, and 1 filler list presented)</i>	
3 unrelated words	6 s
untested/filler list	18 s
<i>20 min list</i>	18 s
<i>15 min list</i>	18 s
<i>10 min list</i>	18 s
<i>5 min list</i>	18 s
<i>2 min list</i>	18 s
unrelated words	6 s
<i>Testing Phase</i>	
Math Instructions + Math Set 1	55 s
Recognition Instructions	30 s
<i>Recognition Test for 2 min list</i>	40 s
Math Instructions + Math Set 2	62 s
Recognition Instructions	30 s
<i>Recognition Test for 5 min list</i>	40 s

Math Instructions + Math Set 3	41 s
Recognition Instructions	30 s
Recognition Test for 8 non-studied words	40 s
Math Instructions + Math Set 4	41 s
Recognition Instructions	30 s
<i>Recognition Test for 10 min list</i>	40 s
Math Instructions + Math Set 5	182 s
Recognition Instructions	30 s
<i>Recognition Test for 15 min list</i>	40 s
Math Instructions + Math Set 6	182 s
Recognition Instructions	30 s
<i>Recognition Test for 20 min list</i>	40 s

---

*Note.* This procedure was completed twice with different lists in the first experimental session and completed once with a third set of lists in the second experimental session.

Table 2

*Hits and False Alarms for Each Test Delay and List Condition Organized by Item Type with Standard Errors of the Means in Parenthesis.*

Delay	Hits		False Alarms	
	Critical Lures	List Items	Critical Lures	List Items
<i>Low Condition</i>				
2 min	0.32 (0.06)	0.78 (0.03)	0.22 (0.05)	0.10 (0.02)
5 min	0.33 (0.06)	0.83 (0.03)	0.25 (0.06)	0.19 (0.03)
10 min	0.17 (0.05)	0.77 (0.04)	0.32 (0.06)	0.20 (0.03)
15 min	0.27 (0.06)	0.74 (0.04)	0.17 (0.05)	0.13 (0.03)
20 min	0.23 (0.06)	0.82 (0.03)	0.22 (0.05)	0.14 (0.03)
2 day	0.33 (0.05)	0.56 (0.03)	0.62 (0.05)	0.42 (0.03)
<i>Moderate Condition</i>				
2 min	0.42 (0.06)	0.73 (0.04)	0.18 (0.05)	0.14(0.03)
5 min	0.23 (0.06)	0.64 (0.04)	0.05 (0.03)	0.07 (0.02)
10 min	0.32 (0.06)	0.75 (0.04)	0.13 (0.04)	0.10 (0.03)
15 min	0.17 (0.05)	0.83 (0.03)	0.33 (0.06)	0.24 (0.03)
20 min	0.40 (0.06)	0.66 (0.04)	0.42 (0.06)	0.29 (0.04)
2 day	0.38 (0.05)	0.53 (0.03)	0.58 (0.05)	0.35 (0.03)
<i>High Condition</i>				

2 min	0.60 (0.06)	0.79 (0.03)	0.10 (0.04)	0.10 (0.02)
5 min	0.53 (0.06)	0.74 (0.04)	0.12 (0.04)	0.07 (0.02)
10 min	0.52 (0.07)	0.80 (0.03)	0.10 (0.04)	0.17 (0.03)
15 min	0.35 (0.06)	0.69 (0.04)	0.17 (0.05)	0.09 (0.03)
20 min	0.28 (0.06)	0.67 (0.04)	0.13 (0.04)	0.11 (0.03)
2 day	0.38 (0.05)	0.56 (0.03)	0.38 (0.05)	0.35 (0.04)

*Averages*

2 min	0.44 (0.04)	0.77 (0.02)	0.17 (0.03)	0.11 (0.02)
5 min	0.37 (0.04)	0.73 (0.03)	0.14 (0.02)	0.11 (0.02)
10 min	0.33 (0.03)	0.76 (0.02)	0.18 (0.03)	0.16 (0.03)
15 min	0.26 (0.03)	0.76 (0.02)	0.22 (0.03)	0.15 (0.02)
20 min	0.31 (0.04)	0.72 (0.03)	0.25 (0.04)	0.19 (0.02)
2 day	0.36 (0.03)	0.55 (0.02)	0.53 (0.03)	0.37 (0.02)

---

Table 3

*Pr Means and Standard Errors of the Means for Each Delay and Item Type**Condition*


---

<u>Delay</u>	<u>List Items</u>		<u>Critical Lures</u>	
	<i>Pr</i>	<i>SE</i>	<i>Pr</i>	<i>SE</i>
2 min	0.66	0.03	0.28	0.05
5 min	0.62	0.03	0.23	0.05
10 min	0.61	0.03	0.15	0.05
15 min	0.60	0.03	0.04	0.05
20 min	0.53	0.03	0.05	0.05
2 day	0.18	0.03	-0.16	0.04

---

Table 4

*d'* Means and Standard Errors of the Means for Each Delay and Item Type*Condition*


---

<u>Delay</u>	<u>List Items</u>		<u>Critical Lures</u>	
	<i>d'</i>	<i>SE</i>	<i>d'</i>	<i>SE</i>
2 min	1.65	0.07	0.57	0.09
5 min	1.59	0.08	0.44	0.10
10 min	1.47	0.08	0.31	0.09
15 min	1.54	0.08	0.07	0.10
20 min	1.40	0.07	0.07	0.12
2 day	0.48	0.07	-0.40	0.11

---

Table 5

*Parameter Estimates from Power Functions Fit to Pr and d' Data (All Delays Included)*

<u>Parameter</u>	<u>List Items</u>	<u>Critical Lures</u>
<i>Pr</i> Data		
$\lambda$ (Intercept)	0.80 (0.06)	0.45 (0.10)
$\beta$ (Slope)	-0.15 (0.03)	-0.59 (0.15)
$R^2$	0.91	0.87
<i>d'</i> Data		
$\lambda$ (Intercept)	1.73 (0.06)	0.93 (0.25)
$\beta$ (Slope)	-0.06 (.02)	-0.61 (0.19)
$R^2$	0.81	0.82

Note. Standard errors are in parentheses.

Figure Caption

*Figure 1.* Log-log plot of the mean  $Pr$  scores by delay with power functions fits to list items and critical lures (the negative 2-day delay data point for lure items was substituted with 0.001 to be represented in the figure)

Figure 1

