Available online at www.sciencedirect.com
ScienceDirect
Journal of Memory and Language 56 (2007) 305–320
Journal of Memory and Language
www.elsevier.com/locate/jml

The roles of spreading activation and retrieval mode in producing false recognition in the DRM paradigm ☆

Michelle L. Meade [a,*], Jason M. Watson [b], David A. Balota [c], Henry L. Roediger III [c]

[a] Department of Psychology, Montana State University, P.O. Box 173440, Bozeman, MT 59717-3440, USA
[b] Department of Psychology, The Brain Institute, University of Utah, 380 S., 1530 E., Salt Lake City, UT 84112, USA
[c] Department of Psychology, Washington University in St. Louis, Campus Box 1125, St. Louis, MO 63130-4899, USA

Received 23 February 2006; revision received 25 July 2006
Available online 26 September 2006

Abstract

The nature of persisting spreading activation from list presentation in eliciting false recognition in the Deese–Roediger–McDermott (DRM) paradigm was examined in two experiments. We compared the time course of semantic priming in the lexical decision task (LDT) and false alarms in speeded recognition under identical study and test conditions. The results revealed priming on the LDT only when a test item occurred immediately (1 s) after the last list item. In contrast, robust false recognition occurred across all delays in both experiments. We interpret the data as indicating that the automatic activation processes evidenced in lexical decision do not persist sufficiently long to produce the false recognition obtained in the DRM paradigm. False recognition occurs because episodic retrieval instructions and a related probe item create reactivation of a list's associative structure, but such reactivation does not occur in LDT under conditions in which subjects are discouraged from retrospective checking of the list.

Keywords: False memory; Lexical decision; Spreading activation; DRM

The Deese–Roediger–McDermott (DRM) paradigm involves presenting subjects with a list of semantically related words (e.g., *bed, rest, wake, tired, dream,* etc.) that converge on a single, non-presented critical item such as *sleep* (Deese, 1959; Roediger & McDermott, 1995). Subjects have been shown to misremember the critical item at remarkably high rates as having been presented in the study list across a variety of experimental situations (see Gallo, 2006; Roediger & Gallo, 2005 for reviews). The purpose of the current paper is to examine the ways in which persistent spreading activation may operate to create the DRM memory illusion. Spreading activation serves as a fundamental retrieval mechanism across a wide variety of cognitive tasks (e.g., Anderson, 1983). The notion is that related concepts are linked in

☆ This research was supported by NIA Grants PO1 AG-03991, AGO2049, and NIA Training Grant AG00030. We thank Jerry Markman for help with the computer program used for testing subjects, personnel at Lackland Air Force Base for help in testing subjects in Experiment 1, Jen Coane and David McCabe for comments on an earlier version of the manuscript, and Darius Tse and James Neely for helpful discussions regarding these studies.
* Corresponding author. Fax: +1 406 994 3804.
E-mail address: mlmeade@montana.edu (M.L. Meade).

0749-596X/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2006.07.007

memory, and that when one item or concept in memory is activated (via encoding or retrieval), the activation spreads to other related concepts (Collins & Loftus, 1975). The concept of spreading activation has been supported in a wide variety of semantic priming experiments (see Neely, 1991 for a review). The present experiments explore the role of spreading activation in producing the high levels of false memory effects with DRM materials by directly comparing the time course of semantic priming effects in a lexical decision task and false memories in a recognition test under identical study-test procedures. The aim is to see whether the same (or similar) sorts of activation operate in the two paradigms or whether their forms are different.

Of course, several theories have been proposed to account for the DRM illusion, and we consider alternative frameworks in the General Discussion. However, the implicit associate response (IAR) theory originally proposed by Underwood (1965) and the more recent Activation Monitoring theory (AMT) advanced by Roediger, Balota, and Watson (2001) (see too Balota et al., 1999; McDermott & Watson, 2001; Roediger, Watson, McDermott, & Gallo, 2001) are most relevant to the current project. The IAR theory relies on the concept of spreading activation: when a word is studied, the meaning of that studied item is activated and the words associated to the studied item are also implicitly activated. To account for associative memory illusions, IAR theory suggests that subjects may falsely recognize that a non-presented associate occurred in the list because activation of the list items has spread and has heightened activation levels of the associate. IAR theory has received support from past research showing that false recognition increases in relation to associative proximity (e.g., Vogt & Kimble, 1973). More recently, Robinson and Roediger (1997) presented subjects with study lists ranging in length from 3 to 15 semantic associates per list and found that false recall and recognition increased with greater numbers of associates studied. An activation account of this finding suggests that increasing the number of associates also increases the convergence of activation on the critical non-presented item. Further evidence for an activation account of false memory comes from a regression analysis conducted on predictors of false memory in the DRM paradigm. Across 55 associative word lists, Roediger, Watson et al. (2001) found the highest levels of false recall and false recognition for the lists with the greatest average associative strength from list items to the non-presented item. The greater the associative strength of the list, the more likely list items activate the non-presented critical item and the more probable is its false recall or false recognition on a later test.

The activation monitoring theory (AMT) is based on the idea that spreading activation works in conjunction with a more controlled, monitoring process that allows subjects to make attributions about the source of the activation (Johnson, Hastroudi, & Lindsay, 1993). That is, when making a memory judgment, subjects may use information from heightened activation, but must also rely on a monitoring process to discriminate those activated items that were studied from those that were not studied. Strongly activated items may be misattributed to having occurred in the list if there is no information to distinguish list items from critical items. Several lines of evidence support the idea that monitoring processes are critical. First, numerous experiments have now shown that when list items are made distinctive in some way (e.g., by presenting them with pictures, or presenting them visually rather than auditorily), false recall and false recognition is lessened (e.g., Israel & Schacter, 1997; Smith & Hunt, 1998). Such evidence supports the distinctiveness heuristic as a means of a person rejecting a candidate memory and reducing levels of false recall or false recognition (see Schacter, Cendan, Dodson, & Clifford, 2001). Second, warning subjects before the lists are presented about the presence of critical non-presented items decreases false recall in the DRM paradigm, presumably due to monitoring processes invoked during encoding which then are carried forward during retrieval (e.g., Gallo, Roberts, & Seamon, 1997; Gallo, Roediger, & McDermott, 2001; McDermott & Roediger, 1998). Third, Roediger, Watson et al. (2001) found a negative correlation between veridical and false memory so that the better remembered the list items were (presumably due in part to greater source monitoring), the less likely subjects were to falsely remember the critical item. Fourth, evidence in support of the AMT comes from investigations of age-related changes in false memories in the DRM paradigm. Because evidence indicates that activation patterns in younger and older adults are equivalent, but that monitoring processes suffer age related decline (see Balota, Dolan, & Duchek, 2000), one might expect declines in veridical recall for older adults relative to young adults, but similar or even heightened false recall in older adults. Indeed, this is the pattern observed by Balota et al. (1999) and Norman and Schacter (1997), among others. Finally, repetition of the study list has been shown to increase false recognition for older adults, but to decrease false recognition for younger adults, presumably because older adults are impaired in their ability to monitor the source of the increased activation from repetition and so have higher false alarms (Benjamin, 2001; Kensinger & Schacter, 1999; Watson, McDermott, & Balota, 2004). These studies and others (e.g., McDermott & Watson, 2001) have been viewed as providing converging evidence that, in addition to activation processes, monitoring processes are an important part of memory decisions on the DRM task.

Although there has been considerable literature viewed as supporting the IAR and AMT, the evidence of a persistent activation process from study list to the test context has been equivocal. The purpose of the current experiments is to explore whether the spreading activation mechanism proposed in false memory theories such as in the IAR and the AMT will be reflected in the standard measure of spreading activation processes in the semantic priming literature, i.e., the lexical decision task (LDT). The LDT involves deciding whether or not a string of letters is a word or a nonword, and, under appropriate conditions, can be a relatively pure measure of activation processes. The LDT has been the standard task used in semantic priming paradigms, and performance in this task has repeatedly been shown to be facilitated when a target word is preceded by a related word compared to an unrelated word (see Neely, 1991, for a review). Hence, this task would appear to be ideally suited to be sensitive to persistent activation from a DRM study list to the test context.

Available evidence suggests reasons that one might find such persistent activation effects in lexical decision performance from DRM lists. For example, Balota and Paul (1996) reported a series of semantic priming studies demonstrating that multiple associatively related primes can produce heightened effects in standard semantic priming paradigms. They found additive activation from related primes on an immediate lexical decision task; that is, the influence of two related primes was nicely predicted by the sum of the activation from each of the individually presented primes. Hence, activation from primes does appear to summate, at least at a short prime target stimulus onset asynchrony (SOA). With the large number of converging primes from DRM lists, one might expect considerable heightened activation that may persist into the test context.

There is also evidence suggesting that multiple primes can influence priming on implicit memory tests for DRM lists. McDermott (1997) found significant levels of semantic priming on word stem completion and word fragment completion after subjects had studied the 15 related items on the DRM word lists. McKone and Murphy (2000) also found significant levels of priming on stem completion, and Lovden and Johansson (2003) demonstrated significant levels of priming on an anagram task (but see McBride, Coane, & Raulerson (in press) for evidence that priming is not always obtained on word stem completion and graphemic cued response tasks). However, as Tse and Neely (2005) have recently argued, it is possible that the past studies that have used these implicit tasks may have suffered from the problem of explicit contamination; that is, subjects may have noticed the relation between the indirect measure and the initial study list and used explicit retrieval strategies. Hence, these studies may not provide pure measures of activation processes, although it should be noted that the authors of these studies argued against such an interpretation for various reasons. Tse and Neely further argue that the LDT may be a better reflection of pure activation processes.

However, there is also available research that questions the potential utility of the spreading activation mechanism reflected by the LDT as an underlying mechanism accounting for the DRM memory illusion and suggests that the activation might be of a different sort. For example, DRM false memory effects persist after sizeable delays (e.g., Seamon et al., 2002), while semantic priming in LDT can be greatly attenuated or eliminated with intervening items (e.g. Dannenbring & Briand, 1982; Masson, 1995). In addition, spreading activation effects can be eliminated by simply switching attention to a different semantic category (e.g., Balota, Black, & Cheney, 1992; Neely, 1977). Further, under some conditions, the LDT discourages explicit retrieval strategies because judgments are made relatively quickly and subjects have little time to retrieve the study episode (Zeelenberg & Pecher, 2002). In contrast, performance on a recognition test is relatively slower and explicitly demands retrieval of earlier list information.

One way to conceive of such differences in task demands is through Tulving's concept of episodic retrieval mode; as Tulving (1983) originally put it, "The same stimulus reminds a person of a particular episode only when the individual's mind is in a particular state; the episodic system must be in the 'retrieval mode' before a stimulus change in the environment can serve as an effective retrieval cue to stored episodic information" (p. 46). Perhaps subjects must be in a retrieval mode (or, equivalently, be given test instructions usually used in explicit memory tests) to obtain the heightened false memory to critical non-presented items. That is, when one deliberately attempts to retrieve from episodic memory, the retrieval cues used (whether list items like *slumber* or critical lures like *sleep*) will cause reactivation of the encoded associative network. Because the formation of this network will be influenced by, and interconnected with, pre-existing semantic representations, as suggested by Anderson (1983), it is likely that the activation from the related words that are part of this network will converge on the critical non-presented item and therefore produce false recollections. If reactivation only occurs when subjects are in retrieval mode, then one might expect to see false recall and false recognition at much longer delays than on the LDT (the instructions for which, as with other implicit memory tasks, discourages subjects from an episodic retrieval mode). On the other hand, it is possible that the remarkably high activation produced by the convergence of DRM lists will persist from the study episode to the test, as suggested by the AMT, and influence lexical decision performance even when retrieval is not directly encouraged.

The studies that have directly compared the activation produced by DRM materials on the LDT have produced conflicting results (see Tse & Neely, 2005 for a review). Zeelenberg and Pecher (2002) presented subjects with 36 DRM lists and then measured subjects' reaction time to the critical words, list words, and nonwords on an LDT. Across four experiments, they obtained no evidence of long term semantic priming. McKone (2004) also found no effects of long term semantic priming on LDT after presenting subjects with 8 DRM lists. Note, however, that in both of these studies, the LDT was not given until after all lists had been presented. Tse and Neely argued that results from these studies are inconclusive because any activation from lists presented early in the study phase may have decayed by the time subjects were asked to respond on the LDT. However, because DRM studies often use a similar delayed test and find robust false recognition (e.g, Roediger & McDermott, 1995, and many more), the absence of priming on the LDT under this condition is still notable. Also, McKone (2004) presented subjects with a final yes-no recognition test following the LDT, and obtained significant levels of false recognition on this final test (in absence of prior priming on the LDT). Although in McKone's study the items in the LDT were later repeated on the recognition test, and so the cause of the false recognition effect is unclear due to this contaminating effect, we tend to believe that it is real because so many DRM experiments have obtained high levels of false recognition on delayed tests (see Gallo, 2006, for a thorough review).

Hancock, Hicks, Marsh, and Ritschel (2003) also examined activation from DRM lists on the LDT by presenting subjects with DRM lists, a filler task, and a lexical decision test for studied items, critical items, filler words and nonwords. Unlike Zeelenberg and Pecher (2002) and McKone (2004), Hancock et al. gave subjects the LDT immediately following each DRM list. The results of their experiments showed significant levels of priming (also see Whittlesea, 2002). More specifically, when subjects were presented with 15-item DRM lists, reaction time for the non-presented critical item was actually faster than for list items that had been studied. When subjects studied only 3-item DRM lists, reaction time for the non- presented critical item was equivalent to reaction times for the studied items. Hancock et al. concluded that this superadditive priming effect was due to the activation of many studied items converging to activate the critical item, an idea consistent with findings of Balota and Paul (1996) and Robinson and Roediger (1997), discussed previously. However, as noted by Tse and Neely (2005), one potential limitation with the Hancock et al. results is that the baseline used for priming on the critical items was an unrelated word. Although the control word was matched to the critical item for frequency, number of letters, and syllabic length, Tse and Neely noted problems with control performance on these baseline words. A more powerful test would be to compare the critical word when it was preceded by a related list to that same critical word when it was preceded by an unrelated list.

Tse and Neely (2005) employed a procedure similar to that used by Hancock et al., but used a baseline that compared reaction times to the same critical non-presented word when it was preceded by a related versus an unrelated list. Replicating Hancock et al., Tse and Neely found significant long term semantic priming effects on the LDT. However, procedures in the Tse and Neely study may have encouraged episodic retrieval, and hence subjects being in a retrieval mode during the LDT. Specifically, the LDT contained (as always) both words and nonwords, but the study lists Tse and Neely used did not include nonwords. Because of this difference, subjects were potentially able to use the relation between the critical item and the list items to aid performance on the LDT (in a postlexical process). That is, if the stimulus word was related to the previous list of 14 highly related DRM words, the subject could be assured it was a word and so speed performance in judging that the target was a word. Such retrospective checking processes have been clearly shown to produce priming not because of spreading activation from the prime to the target, but because of checking for a relationship between the target and the priming context. Such postlexical checking processes have been critical in understanding the nature of underlying semantic priming effects (see Balota & Lorch, 1986; Neely, 1991) and episodic priming effects (see Durgunoglu & Neely, 1987) in the LDT. Of course, simply because subjects could have used this information is not a guarantee that they did, but in our experiments we ruled out this possibility by including nonwords in the study list.

Present experiments

The present experiments explored the role of activation in a study-test paradigm in which we were able to examine the decay of activation across two tasks (LDT and speeded recognition) under identical study test conditions. Subjects in the LDT and recognition conditions studied both nonwords and words which came from DRM lists that converged on a critical non-presented item. During tests given immediately after each study list, subjects either made lexical decisions or episodic recognition decisions on sequences of items that were the same for both tests; only the operations during the test (lexical decision or episodic recognition) differed. Both words and nonwords were equally likely to be old or new during both the lexical decision task and the episodic recognition task. Hence, in the lexical decision task, subjects could not use the presence of an item in the previous study list as a cue that the stimulus was a word. In this way, we eliminated subjects' strategy of using the presence of an item in the study list as a cue to drive the lexical decision task (see Neely, Keefe, & Ross, 1989). If response latencies are faster for the critical non-presented words in the lexical decision task when they follow related lists compared to unrelated lists, then this would reflect the influence of the activation aroused during the study phase as persisting during the LDT. The inclusion of a comparable episodic recognition test is quite important here, because this test allows us to insure that one finds the expected robust false recognition of the critical non-presented words when episodic retrieval is engaged under identical conditions. Prior work leads us to expect that to be so (e.g., McDermott & Roediger, 1998), but prior DRM work did not, of course, include nonwords in the study lists. The current recognition tests will permit us to determine if this methodological change eliminates the effect.

In addition to including nonwords on the study list to minimize the utility of retrieval during the later LDT, we also tested the position of the critical lure at four different positions (1, 3, 6, or 11) within the test phase. This manipulation allowed us to examine differences or similarities in any decay in activation across time (and intervening positions) in the LDT and the episodic recognition tests. This comparison is critical in determining whether similar or different forms of activation underlie priming on the LDT and false recognition in episodic recognition.

Experiment 1

Method

Subjects

Subjects were 288 recruits tested at Lackland Air Force Base in San Antonio, Texas as part of their training requirements.

Design

The experiment consisted of a $2 \times 4 \times 2$ mixed factorial design. Relation of the critical lure on the test list to items on the study list (related or unrelated) and the position of the critical lure on the test list (1, 3, 6, or 11) were both manipulated within subjects. Type of test (lexical decision or speeded recognition) was manipulated between subjects. The dependent variables were the reaction time and accuracy of subjects' responses on the test lists.

Materials

DRM lists, which contained 15 semantic associates to a related critical lure, were used to create forty-eight study lists. The study lists were identical for the LDT and speeded recognition conditions. All study lists contained the fifteen items from the DRM lists (based on lists from Stadler, Roediger, & McDermott, 1999, or Watson, Balota, & Roediger, 2003), six nonwords, and six unrelated filler words for a total of 27 items on each list. Nonwords were pronounceable and were matched to the list items in length. Filler words ranged in length from 4 to 8 letters and had an average frequency of 80 as measured by Kucera and Francis (1967) word frequency norms. They were fully counterbalanced in relation to related and unrelated lists so that the same words served as controls in both related and unrelated conditions. Each list was designed so that a combination of four nonwords and filler items were presented (in random order) at the beginning of the list, then the first five DRM words were presented, then four more nonwords or fillers, then five more DRM words, then four more fillers and nonwords, and finally the last five items from the DRM list. As noted earlier, we intermixed the DRM associates with nonwords and filler items during the study list so that we could eliminate backward checking during the later LDT that might bias word decisions for non-presented critical lures. It is also important to note here that random presentation of DRM associates results in lower, yet still highly reliable, levels of false memory than blocked presentation of DRM associates (e.g., McDermott, 1996; Toglia, Neuschatz, & Goodwin, 1999). However, no prior experiments have included nonwords in DRM study lists.

Test lists for each of the 48 study lists were identical for the lexical decision and the recognition test. Each test list contained 12 items: three previously studied nonwords, three new nonwords, three previously studied filler words, and three new words (one of which was the critical lure). For half of the trials, the critical lure was related to the items in the previously studied DRM list and for the other half the critical lure was taken from a different DRM list so that it was unrelated to the study items. The critical lure was the only item in the test list potentially related to the DRM list (so there could be no potential for test-induced priming; Marsh, McDermott, & Roediger, 2004; Coane & McBride, in press). Multiple test lists were created for each study list so that the critical lure appeared as the first item, the third item, the sixth item or the eleventh item in the 12-item sequence. In addition, multiple versions of lists were created so that even when the position of the lure was held constant, the order of the old and new nonwords and fillers was different.

All study and test lists were completely counterbalanced. Each subject saw every list and each list was presented in each of the experimental conditions an equal number of times across subjects. Test List Position was presented in a blocked order. Specifically, test lists with the critical lure in Position 1 were presented first, followed by test lists with the critical lure presented in the third position, and so on. This feature represents a confounding in Experiment 1, but, as discussed later, in Experiment 2 test list presentation position was completely randomized and the results of this experiment yielded an identical pattern. See Appendix A for an example of a study list and various forms of test lists.

Procedure

The experiment consisted of a series of 48 study and test trials. During each study phase, the items appeared centered on the computer screen and each word was presented at a rate of 1.5 s with an interstimulus interval of 500 ms. Subjects were instructed that no keyboard response was necessary, but to pay attention to the study items because they would later be tested on the items.

After seeing all 27 items from the study list, subjects heard a tone for 1 s that signaled the end of the study phase and that the test phase was about to begin. Thus, the total time between the last item of the study list and the first item of the test list was 1 s. During the test phase, 12 items were presented serially in the middle of the screen. Each of the items remained on the screen until subjects made a lexical decision response or a speeded recognition judgment. Specifically, subjects in the lexical decision condition were told to press the 'M' key if they thought the item was a word and the 'X' key if the item was not a word; subjects in the speeded recognition condition were instructed to press the "M' key if they thought the item had been presented in the previous study list and the 'X' key if the item had not been presented in the previous study list. Hence, all features were equated across the two tasks, with the exception of the type of information used to drive the lexical decision task or the episodic recognition task. Subjects in both conditions were instructed to respond as quickly and as accurately as possible.

After responding to all 12 items on the test list, subjects were presented with a reminder of which keys to press and were then prompted to press the enter key to continue on to the next study and test trial. The words "Next Trial" then appeared on the screen and the study and test sequence was repeated.

Results

In the following analyses, we compare the time course of activation for critical lures on LDT (via the magnitude of semantic priming effects) to the time course of activation for critical lures on a speeded recognition test (via the magnitude of false recognition). To anticipate the results, this study provided clear evidence indicating that activation on the LDT (evident in faster reaction times for related than unrelated critical items) was short-lived, while the effect of relatedness on the speeded recognition test (evident in higher false alarm rates for related than unrelated critical items) persisted across all test positions. Speeded recognition and lexical decision were analyzed separately because they are fundamentally different measures: in speeded recognition, subjects' responses are based on direct retrieval of past experience, whereas in the LDT, subjects' responses are based on the persistent activation and lexical status. In this light, as discussed previously, the two tests qualify as explicit and implicit measures of memory, respectively. Statistical significance was set at $p \leqslant .05$ unless otherwise noted.

Lexical decision

Response latencies. Correct response latencies that were made between 250 and 1500 ms and were within 3 *SD*s of the mean were included in this analysis. After screening and trimming procedures, to avoid missing cells in the position by relatedness ANOVA for critical items in the LDT, it was necessary to remove 37 of the 144 subjects tested in Experiment 1 (though, to preview, this evidently did not influence the outcomes we report, because the findings of Experiment 1 were replicated in Experiment 2 where only 1 subject out of an *N* of 48 had to be removed due to missing cells for the same reason).[1]

The mean reaction times for related and unrelated critical lures across the four possible test positions on the LDT are displayed in Table 1 (along with error rates, to be discussed next). The difference scores (semantic priming) are shown at the bottom of the table. To examine the magnitude of semantic priming on LDT across test positions, a 2 (related or unrelated) × 4 (Test Position 1, 3, 6, or 11) ANOVA was conducted on the mean reaction times. The ANOVA revealed a significant main effect of test position, $F(3, 318) = 132.87$, $MSE = 23{,}262$, which simply indicated that response latencies decreased quite dramatically after the first trial. Priming was greatest for Test Position 1 ($M = 49$ ms) and then vanished across the remaining test positions ($M = -18$ ms for Position 11). More importantly, there was no main effect of relatedness ($F < 1$), but there was a reliable interaction between relatedness and test position,

[1] Additional analyses revealed that the overall pattern of results in the Experiment 1 lexical decision task did not depend on the deletion of these 37 subjects. Specifically, when available data from these subjects were included in the analyses, the overall pattern of the means was qualitatively similar to the lexical decision results presented in Tables 1 and 2. Inferential statistics were not performed due to missing cells in the analyses. We suspect that the reason for the great variability in LDT performance is that the education levels of our Air Force recruits were quite mixed and some had difficulty performing the LDT. In addition, motivation of the recruits in performing the LDT in the midst of basic training might not be great. These speculations are borne out in part by the fact that when we tested Washington University students in highly similar procedures in Experiment 2, the error rates dropped, reaction times were faster, and only one subject had to be eliminated from the analysis. Despite these differences, the same patterns of data occurred across the independent variables in both experiments.

Table 1
Mean reaction times (in ms) and mean proportion of errors on the lexical decision test for the related and unrelated critical words as a function of test position in Experiment 1 ($N = 107$)

	Test Position			
	1	3	6	11
Related critical word				
RT	917 (22)	715 (13)	698 (11)	720 (11)
Error	.01 (.01)	.02 (.01)	.03 (.01)	.03 (.01)
Unrelated critical word				
RT	966 (25)	690 (10)	683 (10)	702 (11)
Error	.01 (.00)	.02 (.01)	.02 (.01)	.02 (.01)
Semantic priming effect				
RT	+49	−25	−15	−18

Parenthetical values represent the standard error of the mean.

$F(3, 318) = 4.89$, $MSE = 12{,}968$. Further analyses on the interaction indicated a significant difference between related and unrelated conditions on test Position 1 only, $F(1, 106) = 4.62$, $MSE = 27{,}678$. Semantic priming effects did not obtain on Test Positions 3, 6, or 11 (F's < 3.6), and in fact all effects are slightly inhibitory as opposed to facilitatory. This pattern clearly suggests that the facilitatory effect of spreading activation as measured on LDT in the DRM paradigm is short-lived.

Errors. The mean overall error rate was .02 when collapsed across all conditions (see Table 1). Errors did not vary as a function of relatedness, $F < 1.00$, or test position, $F(3, 318) = 2.16$, $MSE = .004$. More importantly, the interaction between relatedness and test position was not significant ($F < 1$), suggesting no speed/accuracy tradeoff.

Repetition priming. To insure that our lexical decision task was sensitive to the presentation of the previous study list, we conducted a 2 (old or new) by 2 (word or nonword) within subjects ANOVA to examine the repetition effects. As is evident in Table 2, subjects were faster to respond to previously presented items ($M = 809$) compared to new items ($M = 846$), $F(1, 106) = 212.50$, $MSE = 691$. Subjects were also faster to respond to words ($M = 772$) than to nonwords ($M = 883$), $F(1,106) = 219.46$, $MSE = 5{,}963$. Finally, the repetition by lexicality interaction was significant, $F(1,106) = 105.78$, $MSE = 768.3$. This interaction reflects the common finding that repetition of words produce greater facilitation than repetition of nonwords. Nonwords generally produce smaller repetition effects than words, and this outcome suggests that repetition of nonwords produces an increase in their familiarity which in turn increases the difficulty of making a nonword decision (e.g., see Balota & Spieler, 1999).

Table 2
Mean reaction times (in ms) of old and new words and nonwords in Experiment 1 ($N = 107$)

	Old	New
Words	740 (8)	804 (9)
Nonwords	878 (11)	887 (11)

Parenthetical values represent the standard error of the mean.

Speeded recognition

False recognition. The speeded recognition test allowed us to determine if significant levels of false recognition were obtained on lists identical to those used to measure activation in LDT. We were also interested in whether the levels of false recognition across test positions would track the levels of activation found on the LDT. The mean false alarm rates for the critical items after related and unrelated lists across test positions are displayed in Table 3, with the difference score at the bottom representing the DRM effect. We included in our analyses of the recognition data only responses that were made between 250 and 5000 ms and were within 3 standard deviations of the mean (1 subject's data out of 144 subjects were excluded because the majority of the responses did not meet these criteria).

To examine the level of false recognition across test positions (as evidenced by false alarms to the critical

Table 3
Mean proportion of false alarms to the critical lure on the speeded recognition test for the related and unrelated critical lures as a function of test position in Experiment 1 ($N = 143$)

	Test Position			
	1	3	6	11
Related Critical Lure	.49 (.03)	.55 (.03)	.54 (.02)	.56 (.02)
Unrelated Critical Lure	.17 (.02)	.23 (.02)	.24 (.02)	.30 (.02)
DRM Effect	.32	.32	.30	.26

Parenthetical values represent the standard error of the mean.

item when it was in the related condition relative to when it was in the unrelated condition), we conducted a 2 (related or unrelated) × 4 (Test Position 1, 3, 6, or 11) ANOVA on the mean false recognition rates for the critical non-presented items. The ANOVA revealed a highly reliable main effect of relatedness, $F(1, 142) = 273.59$, $MSE = .10$, which indicated that subjects were more likely to falsely recognize a critical item when it was directly preceded by a related list ($M = .54$) than when it was preceded by an unrelated list ($M = .23$). Clearly, this large 30% effect of relatedness indicates that this paradigm is sensitive to high levels of false recognition. This finding is of course consistent with past research demonstrating robust false recognition on the DRM lists even on immediate recognition tests (McDermott & Roediger, 1998). This is the first report of such robust false recognition when no items related to the critical lure were included during the recognition test and when nonwords were included in the study list (see too Coane & McBride, in press). Of further interest is the magnitude of false recognition across test positions. Based on past research showing persistent (or even increasing) levels of false recognition and false recall over delays (e.g., McDermott, 1996; Seamon et al., 2002), we expected the magnitude to remain relatively stable across test positions. There was also a reliable main effect of test position, $F(3, 426) = 12.31$, $MSE = .04$, indicating that overall false alarms appeared to increase at later test positions. However, because critical items were blocked by test position this may simply reflect participant strategies. This is remedied in the second experiment. More importantly, there was no interaction between relatedness and test position ($F < 2$). Therefore, in contrast to the LDT, the higher false recognition following semantically related lists, compared to unrelated lists, occurred equally across all test positions.

Hit and false alarm rates for non-critical items. The mean hit and false alarm rates for non-critical items are presented in Table 4. A 2 (word or nonword) × 2 (old or new) ANOVA computed on the proportion of "old" responses revealed a significant main effect of study status, $F(1, 142) = 854.85$, $MSE = .03$, and a significant main effect of lexicality, $F(1, 142) = 39.39$, $MSE = .01$. Subjects had higher hit rates ($M = .60$) than false alarm rates ($M = .20$), and subjects had more "old" responses for nonwords ($M = .43$) than words ($M = .37$). The interaction between study status and lexicality was not significant ($F < 1.00$). Importantly, when the hit and false alarm rates of non-critical items are examined in relation to the false alarm rate for the critical items reported previously, it is clear that the DRM lists with interleaving blocks of filler words and nonwords used in the current experiment elicited robust levels of false recognition. Specifically, the mean false alarm rate for critical items ($M = .54$, see Table 3) was higher than the false alarm rate to the non-critical items ($M = .17$ for words) and comparable to the hit rate for non-critical studied items ($M = .57$ for words).

Table 4
Mean proportion of "old" responses given to old items (hit rate) and new items (false alarm rate) for non-critical words and nonwords in Experiment 1 ($N = 143$)

	Old	New
Words	.57 (.01)	.17 (.01)
Nonwords	.63 (.01)	.23 (.01)

Parenthetical values represent the standard error of the mean.

Response latencies. The reaction time data from the speeded recognition test was analyzed to examine the potential for a similar time course in response latencies across the LDT and the recognition test for the critical lures. No difference was obtained between the mean reaction time for false alarms to critical items in the related condition ($M = 1201$) relative to the unrelated condition ($M = 1200$), $F(1, 121) < 1$. Note that these reaction time data are somewhat unstable, particularly in the unrelated condition where subjects had fewer false alarms to use as a comparison baseline. Also, position was not used as a factor in these reaction time analyses due to missing cells in the position by relatedness design that varied between most of the subjects. However, it seemed reasonable to collapse across test position to address the influence of relatedness on reaction times because we did not observe an interaction of position by relatedness in the preceding analysis on the magnitude of false alarms.

Reaction time data were also analyzed for correct rejections of critical items in the related and unrelated conditions. As expected, subjects were slower to correctly reject critical items in the related condition ($M = 1368$) relative to the unrelated condition ($M = 1189$), $F(1, 140) = 36.39$, $MSE = 62{,}263$. Thus, even when subjects were able to correctly reject critical DRM lures, they were slower to do so.

Experiment 2

Experiment 1 revealed that semantic priming effects for critical lures in lexical decision performance and false recognition for critical lures in speeded recognition performance are similar (i.e., greater for related than unrelated lists) only at the first test position; on subsequent test positions (3, 6, and 11) the data clearly diverge. Specifically, there is no evidence beyond Test Position 1 that activation persists in the LDT, whereas a large influence of relatedness occurs in speeded recognition across all test positions. Hence, these results would appear to indicate that pure activation processes persisting from the list presentation at study are short-lived in the DRM paradigm and that the activation during retrieval (or reactivation) drives false recognition (and presumably false recall). However, before discussing these issues in more detail, we report the results of a second experiment that had two goals. First, because of the inconsistent results reviewed in the introduction in the LDT with DRM materials, we wanted to replicate the results of Experiment 1 showing different time courses of priming in the two tasks. Second, as noted earlier, Experiment 1 involved a blocked design wherein tests of particular list positions were blocked and hence confounded with practice across the experiment. In Experiment 2 we introduced a random presentation of the critical items across the four test positions. It is possible (albeit unlikely) that the blocked presentation used in Experiment 1 may have encouraged strategic processes that highlighted a given test position and might have influenced the observed pattern of results. Thus we wanted to replicate the pattern of data obtained on the LDT and recognition tests with random presentation of the critical lure across test positions. We also used a different source of subjects in Experiment 2.

Method

Subjects

Subjects were 96 Washington University undergraduates who participated in the experiment for partial fulfillment of a course requirement. Because students were used rather than Air Force recruits, we tested fewer subjects (expecting their data to be faster with fewer errors, assumptions that turned out to be correct). Importantly, as described below, our data show the same pattern of results can be obtained with two different subject populations.

Design

The $2 \times 4 \times 2$ mixed factorial design of Experiment 2 is identical to the design of Experiment 1. Relation of the critical lure on the test list to items on the study list (related or unrelated) and the position of the critical lure on the test list (1, 3, 6, or 11) were both manipulated within subjects. The position of the critical lure randomly varied across test position within subjects. Type of test (lexical decision or speeded recognition) was again manipulated between subjects. The dependent variables were the same as in Experiment 1.

Materials

The same study and test lists used in Experiment 1 were also used in Experiment 2.

Procedure

The procedure was similar to the procedure of Experiment 1. Minor changes included using asterisks presented at the center of the computer screen instead of a beep to indicate the start of the test phase. This change allowed subjects to be tested in small groups of 3–5 people without disturbing one another. The asterisk sequence lasted 1 s so that the total time between the last item of the study list and the first item on the test list was 1 s (the same as in Experiment 1). Additional changes to the procedure included the introduction of practice trials and a break half way through the experiment. Before beginning the experimental trials, subjects were presented with 3 practice study-test trials. Subjects were also given a one minute break half way through the experimental trials. Both the practice trials and the break were introduced to improve the subjects' ability to accurately perform the task (because 37 subjects in the LDT condition were excluded in Experiment 1).

Results

Lexical decision

Response latencies. The mean reaction times for related and unrelated critical lures across the four possible test positions in LDT are displayed in Table 5 (along with error rates) for critical lures following related and unrelated lists, with the difference (priming score) shown at the bottom of the table. As in Experiment 1, analyses included only correct responses that were made between 250 and 1500 ms and were within 3 standard deviations of the mean. Data from 1 of the 48 subjects were excluded because the majority of the responses did not meet these criteria.

To examine the relative amount of priming across test positions, a 2 (related or unrelated) $\times$ 4 (Test Position 1, 3, 6, or 11) ANOVA was conducted on the mean response latencies. As in Experiment 1, the ANOVA revealed a main effect of test position, $F(3, 138) = 14.78$, $MSE = 20{,}407$, which indicated that response latencies were slowest at the first position. Further replicating Experiment 1, there was no main effect of relatedness, $F(1, 46) = 2.82$, $MSE = 4{,}216$. More importantly, the ANOVA did reveal a significant interaction between relatedness and test position, $F(3, 138) = 3.23$, $MSE = 3{,}406$. Follow up analyses revealed significant priming only when the critical item was in Test Position 1, $F(1, 46) = 8.23$, $MSE = 5{,}026$. Comparable statistical tests on Test Positions 3, 6, and 11 were not significant (all Fs < 1.00). This pattern nicely replicates the results from Experiment 1, and indicates that the facilitatory influence of activation from the DRM lists does not persist throughout the list and is only available at the first test position.

Errors. As shown in Table 5, the overall error rate for critical items on LDT was quite low and the results from the ANOVA did not yield any main effects or interactions, all Fs < 1.1.

Table 5
Mean reaction times (in ms) and mean proportion errors on the lexical decision test for the related and unrelated critical words as a function of test position in Experiment 2 ($N = 47$)

	Test Position			
	1	3	6	11
Related critical word				
RT	746 (30)	654 (21)	651 (20)	656 (18)
Error	.01 (.00)	.01 (.01)	.01 (.01)	.01 (.01)
Unrelated critical word				
RT	788 (34)	664 (20)	650 (19)	650 (16)
Error	.01 (.00)	.01 (.01)	.01 (.01)	.01 (.01)
Semantic priming effect				
RT	+42	+10	−1	−6

Parenthetical values represent the standard error of the mean.

Repetition priming. Repetition priming was significant in all conditions of the experiment. As is evident in Table 6, subjects were faster to respond to old items ($M = 719$) than new items ($M = 747$), $F(1, 46) = 76.5$, $MSE = 482$, and faster to respond to words ($M = 705$) relative to nonwords ($M = 761$), $F(1, 46) = 45.86$, $MSE = 3{,}145$. The lexicality by repetition interaction was also significant, $F(1,46) = 49.41$, $MSE = 384$, again indicating that words showed repetition priming to a greater degree than nonwords.

Speeded recognition

False recognition. The mean false recognition rates for related and unrelated critical lures across test positions are presented in Table 7, along with the difference score (corrected false recognition) at the bottom of the table. As in Experiment 1, analyses only included responses that were made between 250 and 5000 ms, and were within 3 standard deviations of the mean. No participants' data were excluded using these criteria.

A 2 (related or unrelated) × 4 (Test Position 1, 3, 6, or 11) ANOVA conducted on the mean proportion of false recognition revealed a highly reliable main effect of relatedness, $F(1, 47) = 95.40$, $MSE = .20$. As in Experiment 1, subjects were much more likely to false alarm to the critical items when they had been preceded by a related list ($M = .57$) compared to an unrelated list ($M = .12$). In addition, there was a main effect of test position, $F(3, 141) = 2.76$, $MSE = .03$, reflecting a slightly higher false alarm rate at the third position within the test list (see Table 7). Most importantly, however, there was again no hint of an interaction between test position and relatedness, $F < 1.0$. Hence, in contrast to the LDT, the speeded recognition results indicated that robust levels of false recognition following related lists, compared to unrelated lists, remained stable across test positions, providing a clear replication of the pattern obtained in Experiment 1.

Table 6
Mean reaction times (in ms) of old and new words and nonwords in Experiment 2 ($N = 47$)

	Old	New
Words	681 (16)	729 (18)
Nonwords	757 (17)	765 (18)

Parenthetical values represent the standard error of the mean.

Table 7
Mean proportion of false alarms to the critical lure on the speeded recognition test for the related and unrelated critical lures as a function of test position in Experiment 2 ($N = 48$)

	Test Position			
	1	3	6	11
Related critical lure	.54 (.04)	.62 (.04)	.56 (.04)	.54 (.04)
Unrelated critical lure	.11 (.03)	.16 (.03)	.11 (.03)	.11 (.03)
DRM effect	.43	.46	.45	.43

Parenthetical values represent the standard error of the mean.

Hit and false alarm rates for non-critical items. Table 8 displays the hit and false alarm rates for non-critical words and nonwords. We conducted a 2 (word or nonword) × 2 (old or new) ANOVA on non-critical old and new items presented in the study lists. The ANOVA revealed a significant main effect of lexicality, $F(1, 47) = 19.06$, $MSE = .02$, showing that subjects responded "old" more to nonwords ($M = .42$) than words ($M = .32$), and a significant effect of study status indicating that subjects were more likely to respond "old" to studied items ($M = .58$) than distractors ($M = .16$), $F(1, 47) = 205.81$, $MSE = .04$. A significant interaction between lexicality and repetition, $F(1, 47) = 59.55$, $MSE = .003$, revealed that subjects had higher hit rates for nonwords ($M = .66$) than words ($M = .50$), and this difference was in the same direction but smaller in the false alarm rate ($M = .18$ for nonwords; $M = .14$ for words). It is noteworthy that this pattern replicates the

Table 8
Mean proportion of "old" responses given to old items (hit rate) and new items (false alarm rate) for non-critical words and nonwords in Experiment 2 ($N = 48$)

	Old	New
Words	.50 (.02)	.14 (.03)
Nonwords	.66 (.02)	.18 (.02)

Parenthetical values represent the standard error of the mean.

tendency for higher positive responses on the recognition test for pseudowords than words that was observed in Experiment 1, which is inconsistent with the standard mirror observation (see Greene, 2004, for a discussion of the pseudoword effect in recognition memory). More importantly, when these are findings compared to the false alarm rates for critical items (discussed above), we gain leverage on the strength of the false memory effect obtained in the current experiment. Specifically, the mean false alarm rate for critical items ($M = .57$, averaging across positions; see Table 5) was much greater than the false alarm rate for non-critical items ($M = .14$ for words) and higher than the hit rate for non-critical studied words ($M = .50$). This finding provides further support that the lists used in the current experiment with interleaved blocks of DRM items and filler words produced robust levels of false recognition.

Response latencies

As in Experiment 1, reaction time data for the critical lures on the speeded recognition test were analyzed using an ANOVA. Although the latencies were 91 ms in the predicted direction, the difference between the related condition ($M = 899$) and the unrelated condition ($M = 990$) did not reach significance, $F(1, 34) = 2.82$, $MSE = 50{,}441$, $p = .10$. As in Experiment 1, there were many missing cells in this analysis especially in the unrelated condition, as subjects did not always false alarm to the critical lure in that condition.

Reaction time data were also computed for correct rejections of critical items in the related and unrelated conditions. As in Experiment 1, subjects were slower to correctly reject critical lures in the related condition ($M = 942$) than in the unrelated condition ($M = 878$), $F(1, 47) = 13.59$, $MSE = 7{,}367$.

General discussion

The experiments reported here are the first to directly compare the time course of spreading activation as measured by LDT to the time course of spreading activation underlying false memories elicited in the DRM paradigm under nearly identical conditions. Several important findings emerged from this comparison. First, activation levels on the LDT and speeded recognition tasks did not track each other across test positions on an immediate test. In LDT, priming from the study list was evident only when the critical item appeared as the first item on the test list and was eliminated at later positions. In speeded episodic recognition, false recognition of the critical item persisted across multiple intervening items and increasing temporal delays between study and test. This same pattern of results was found when test position and relatedness were presented in a blocked order (Experiment 1) and a random order (Experiment 2). These results are summarized in Fig. 1, which displays the mean critical item difference score for priming on the LDT and false recognition on the speeded recognition test, combining the data from Experiments 1 and 2. For the LDT measure in the top panel of Fig. 1, the difference score is based on reaction time differences to the same critical items following

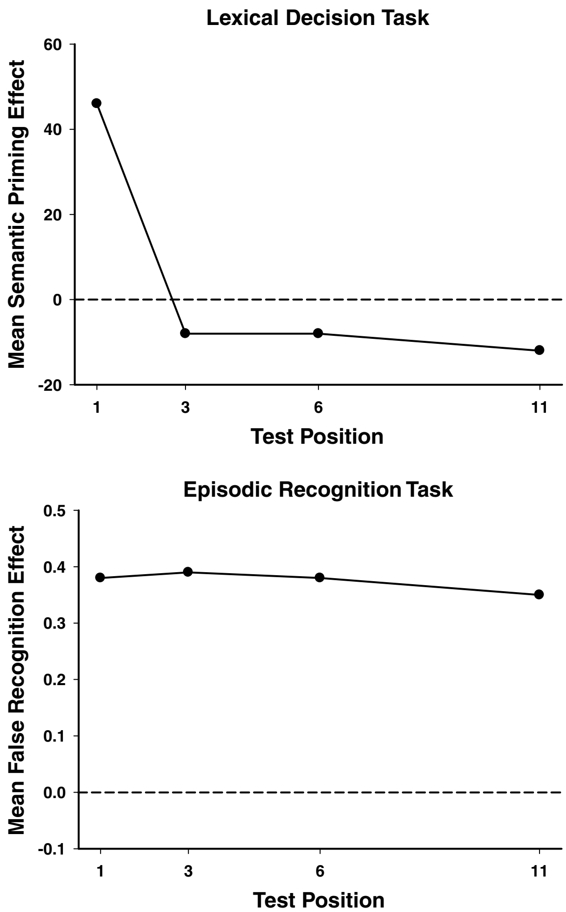

Fig. 1. Mean semantic priming effects and mean DRM false recognition effects for critical items presented in test positions 1, 3, 6, and 11 for a lexical decision task ($N = 154$, top panel) and a speeded recognition task ($N = 191$, bottom panel) data used in figure are collapsed across Experiments 1 and 2.

related and unrelated lists (i.e., semantic priming effects). For the episodic recognition measure in the bottom panel of Fig. 1, the difference score is based on false alarm proportion differences to the same critical items following related and unrelated lists (i.e., DRM false recognition effects). The time course of the relatedness effect on the LDT was short-lived and disappeared with just two intervening items, whereas the effect of relatedness on the speeded recognition test was quite strong and persisted across multiple intervening items and extended temporal delays.

As noted in the introduction, Tse and Neely's (2005) experiments were similar to ours, and also produced important results. Across four experiments, they found evidence that appears to support the notion that activation persists from presentation of DRM lists beyond the first test position that we observed. Our experiments were not conducted in response to Tse and Neely (our first experiment was conducted in 1997–1998), but nonetheless the critical question we must answer is how to reconcile our results (showing activation for only the first item tested in LDT) and Tse and Neely's results (showing activation persisting across much longer intervals). There are at least three probable differences in our sets of experiments that may account (singly or jointly) for the differences in outcome. First, in the Tse and Neely study, subjects were presented all 14 words from a DRM list on each study trial and were given intentional study instructions. Therefore, the critical non-presented item may be consciously activated during list presentation in rehearsing the list in anticipation of a memory test (Goodwin, Meissner, & Ericsson, 2001). If subjects did indeed think of the critical non-presented item for some lists, the priming on LDT in Tse and Neely's experiments may represent repetition priming rather than priming from automatic spreading activation. As shown in the present experiments, the LDT is very sensitive to lexical repetition. Second, because the results from their first experiment produced only a small (17) ms effect in the LDT, Tse and Neely decided to include pseudohomophone nonwords (e.g., brane) and to reduce the luminance contrast in their subsequent experiments. Both the reduction of contrast (Becker & Killion, 1977) and the presence of pseudohomophones (Becker, Moscovitch, Behrmann, & Joordens, 1997) have been shown to increase the reliance on semantic context in making lexical decisions. As predicted, Tse and Neely found larger effects in their subsequent experiments. However, both perceptual degradation and the presence of pseudohomophones, which increased the difficulty of the task, may have caused subjects to increase their reliance on additional sources of information to aid them in making the "word/nonword" discrimination; in particular, as detailed in the next point, they may have retrieved the strong relation between the critical item and the highly integrated list of 14 DRM items that were just presented in a retrospective check (see Neely et al., 1989). Moreover, if such methodological constraints as used by Tse and Neely are necessary to observe the priming effects in LDT, one might question the relevance to normal DRM experiments, in which the luminance of the recognition study words is not degraded. Third, and perhaps most importantly, Tse and Neely did not include any nonwords on the study list. Thus, subjects could use the relation between the critical item during the LDT and the earlier study episode to help drive the "word" response. As just noted, evidence for such retrospective checking processes in the LDT has been shown in many experiments (Neely et al., 1989; see Neely, 1991, for a review). Although the critical nonpresented item was not studied, a strong relationship existed between it and the earlier highly integrated studied list in those conditions when the relevant list was presented. Therefore, retrieving this relationship between the critical item and the list would facilitate the "word" response when the critical item followed the related lists compared to when it followed unrelated lists, and this factor (rather than automatic spreading activation) may have created the priming observed in Tse and Neely's experiments. The inclusion of nonwords on the study list in our experiments reduced any utility of retrieving the earlier list context and thus our LDT procedure probably constituted a purer measure of activation without the influence of retrospective processes.

Taking a broader perspective, our results may not disagree too drastically with those of Tse and Neely, because both sets of experiments agree that priming from DRM lists is relatively short-lived (although they vary with the definition of "short-lived"). We refer to Tse and Neely's Experiment 4, which contained conditions most similar to the conditions in the current study (with no opportunity for test-induced priming and an attempt to minimize, but not eliminate, the potential influence of backward checking). In this experiment, they obtained significant long term semantic priming effects only in the first half of the LDT, when the critical lure was presented in Positions 5 through 9; they did not find priming when the critical lure was presented in Test List Positions 25 through 29. Of course, in our experiments, LDT priming was diminished by Test Position 3. Semantic priming indeed appears to last longer with the blocked lists used in Tse and Neely (2005) than with mixed lists (containing both nonwords and unrelated words) used in the current experiments. However, we note that both sets of experiments agree that priming effects from DRM lists on the LDT are relatively short-lived when the opportunity to use retrospective retrieval processes in the LDT are minimized. This supports our contention in the introduction that subjects must be in an episodic retrieval mode (either through retrieval instructions or through task demands) to see long-lasting effects of associative materials on a related, but not previously presented, item.

A possible criticism of our experiments is that the inclusion of nonwords on our study list could have disrupted the activation produced by the study list, and so our results do not generalize to the standard DRM paradigm (which was used by Tse & Neely, 2005). We have three responses to such a concern: First, because we included a recognition test under identical conditions, we were able to provide a direct measure of the size of the false recognition effect. The results yielded very high false recognition performance under identical study and test conditions. In fact, the false alarm rates for the critical items following related lists were comparable or potentially higher than the hit rates for matched words that were actually presented on the list ($M = .54$ false alarm rate, $M = .57$ hit rate in Experiment 1, and $M = .57$ false alarm rate, $M = .50$ hit rate in Experiment 2). Such high levels of false recognition are a signature finding in DRM studies. Second, high levels of false recall and false recognition in the DRM paradigm have been observed in longer lists of randomly intermixed related items, although random presentation does reduce false recall and false recognition (McDermott, 1996; Tussing & Greene, 1997). Hence, the mixed list procedure used in the present study clearly has applicability to such DRM findings. In addition, Robinson and Roediger (1997) manipulated list length of DRM lists at levels of 3, 6, 9, 12 or all 15 words. In one experiment they used unrelated filler words to make each list 15 items long whereas in another experiment they did not. The false recall and false recognition rates were roughly the same, indicating that adding unrelated words did not diminish the effect. Third, and importantly, the LDT procedure used in our experiments was indeed sensitive to activation from the list presentation: reliable facilitation of 49 ms occurred at the first test position in Experiment 1 and the corresponding figure was 42 ms in Experiment 2. These priming effects are slightly larger than those reported by Tse and Neely, which ranged between 17 and 36 ms at later positions within the list, which makes the critical point that our present LDT was sensitive to activation from the study list. However, consistent with classic notions of the automatic spreading activation process, these effects were short-lived. In particular by the third test position in the LDT, i.e., after two unrelated items intervened between the list's presentation and test, the automatic activation had disappeared.

If simple spreading activation is shown to dissipate shortly after list presentation following DRM lists (in agreement with LDT experiments using more standard single semantic primes), why do the lists cause such high levels of false recognition under nearly identical conditions? As anticipated in the introduction, we believe that our results indicate that DRM false recognition is driven by reactivation processes during retrieval, and these retrieval processes can be minimized and dissociated from the short-lived activation processes that occur in LDT. Indeed, our experiments provide strong evidence for this assumption. Below we describe a tentative account that features activation and monitoring processes and can be considered a further specification of activation/monitoring theory (Roediger, Balota et al., 2001; Roediger, Watson et al., 2001 see Hutchison & Balota, 2005; Watson et al., 2003 for further discussion).

Along with many other researchers (e.g., Anderson, 1983; Raaijmakers & Shiffrin, 1981), we assume that the event of studying a list of words brings on line related information via a spread of activation to related items stored in memory networks. Hence, the encoding of a related list like a DRM list can therefore be represented as an integrated network of associatively related information. The representation may include both studied items and nonstudied (but related) items that were associatively activated during study. The greater the tendency of list items to arouse the critical item during study (the greater the backward associative strength from the list to the critical item), the more probable is false recall and false recognition (Deese, 1959; McEvoy, Nelson, & Komatsu, 1999; Roediger, Watson et al., 2001). Unless attention remains focused on the network, however, this activation dissipates quickly, as reflected by the present LDT results. However, we propose that there is also an important role for activation processes during episodic retrieval, because the well-integrated network encoded during the study phase is still present at the time of test. When subjects are in an episodic retrieval mode (Tulving, 1983) or, more prosaically, when they are given explicit retrieval instructions and a strongly related recognition probe during the test, the probe serves to reactivate the episodic associative network. The degree of overlap between the probe and the reactivated network provides information that is used in making episodic recognition decisions. In this light, we would argue that for DRM lists high in backward associative strength, the critical item recreates a high level of activation that is useful for the recognition decision. In the LDT, there is no utility in consulting the episodically instantiated network (at least under the conditions of our experiment) and consequently the LDT is sensitive only to forward activation from the items for a brief period of time after list presentation. However, in recognition, subjects must use this network to make an episodic decision, and hence, the network is reactivated during retrieval. Because the critical non-presented item receives considerable activation upon its presentation during retrieval from the related words in this network, unless subjects can use source monitoring processes to reject its occurrence in the list, it will be falsely recognized (Jacoby, 1991; Israel & Schacter, 1997). In this light, it is surprising that our experiments and others (e.g., McDermott & Roediger, 1998) show very high levels of false recognition even at very short retention intervals, when source information should be high. One solution to the puzzle is that our experiments used speeded recognition, and therefore subjects may have minimized efforts towards carefully monitoring their responses. However, subjects did show good discrimination between presented and nonpresented words and nonwords, so the high false alarm rates to critical lures does not seem attributable to some general criterion shift in calling all items "old." As noted above, the hit rate to studied words was actually similar to the false alarm rate to critical items in both experiments.

We have interpreted our results within the activation monitoring framework, and have argued that the activation causing the DRM memory illusion is in large part due to a reactivation of a highly interconnected network during episodic retrieval. Of course, alternative theories have been proposed to explain the high levels of false recognition with DRM items without the appeal to activation mechanisms (e.g., Arndt & Hirshman, 1998; Brainerd & Reyna, 2002; Toglia et al., 1999; Whittlesea, 2002). For example, Arndt and Hirshman (1998) have argued that such effects naturally arise out of Hintzman's (1996) MINERVA model of recognition memory in which the vector of features in the critical probe produces a high level of familiarity due to the overlap with multiple episodes of the individual list items. Brainerd and Reyna (1990) have interpreted the high level of false recognition results within a fuzzy trace framework in which a high degree of semantic gist-based information is provided by DRM lists and this causes false recognition. Gallo (2006) reviews various theories postulated to account for associative memory illusions, and it is beyond the scope of the present paper to evaluate the relative merits of all the various theories. We have emphasized the activation/monitoring framework in this paper because our experiments were designed around it and because it is most compatible with standard interpretations of the lexical decision task. To our knowledge, other theorists explaining false recognition have not extended their theories to semantic priming effects in LDT and so their theories would have relatively little to say about the current pattern of results.

In sum, the present results have important implications for further understanding the mechanisms underlying the DRM memory illusion. We propose that subjects establish a well-integrated associative network during encoding of DRM lists. Although automatic spreading activation processes probably play a role in the instantiation of such a network, these processes do not persist long beyond list presentation, as reflected by our lexical decision results. Rather, false recognition and false recall are caused by reactivation of the integrated network that occurs during retrieval when explicit retrieval instructions create an episodic retrieval mode which in turn leads to false recognition. Activation during retrieval, if unopposed by careful monitoring processes via recollection gives rise to the DRM false memory illusion.

Appendix A. The study list for the critical lure, sleep, and examples of possible test lists

Study list: SLEEP
style, measure, slable, trock, bed, rest, awake, tired, dream, older, session, bandle, drep, wake, snooze, blanket, doze, slumber, view, twin, loke, hirning, snore, nap, peace, yawn, drowsy.

Test list: (related critical lure in Position 1)
sleep, metro, style, slable, session, smitar, mamp, view, drep, kapen, hirning, terrace.

Test list: (unrelated critical lure in Position 1)
flag, metro, style, slable, session, smitar, mamp, view, drep, kapen, hirning, terrace.

Test list: (related critical lure Position 3)
slable, terrace, sleep, hirning, style, session, mamp, view, drep, kapen, smitar, metro.

Test list: (unrelated critical lure Position 3)
slable, terrace, flag, hirning, style, session, mamp, view, drep, kapen, smitar, metro.

Test list: (related critical lure Position 6)
mamp, metro, drep, slable, view, sleep, kapen, style, hirning, smitar, session, terrace.

Test list: (unrelated critical lure Position 6)
mamp, metro, drep, slable, view, flag, kapen, style, hirning, smitar, session, terrace.

Test list: (related critical lure Position 11)
style, terrace, smitar, session, drep, hirning, metro, kapen, view, mamp, sleep, slable.

Test list: (unrelated critical lure Position 11)
style, terrace, smitar, session, drep, hirning, metro, kapen, view, mamp, flag, slable.

References

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

Arndt, J., & Hirshman, E. (1998). True and false recognition in MINERVA2: explanations from a global matching perspective. *Journal of Memory and Language, 39*, 371–391.

Balota, D. A., Cortese, M. J., Duchek, J. M., Adams, D., Roediger, H. L., McDermott, K. B., et al. (1999). Veridical and false memories in healthy older adults and in dementia of the Alzheimer's Type. *Cognitive Neuropsychology, 16*, 361–384.

Balota, D. A., & Lorch, R. F. Jr., (1986). Depth of automatic spreading activation: mediated priming effects in pronunciation but not in lexical decision. *Journal of Experimental Psychology: Learning, Memory and Cognition, 12*, 336–345.

Balota, D. A., Black, S., & Cheney, M. (1992). Automatic and attentional processes in young and older adults: a reevaluation of the two-process model of semantic priming. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 485–502.

Balota, D. A., & Paul, S. T. (1996). Summation of activation: Evidence from multiple primes that converge and diverge within semantic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 827–845.

Balota, D. A., & Spieler, D. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology: General, 128*, 32–55.

Balota, D. A., Dolan, P. O., & Duchek, J. M. (2000). Memory changes in healthy older adults. In E. Tulving & F. I. M. Craik (Eds.), *The oxford handbook of memory*. New York: Oxford University Press.

Becker, C. A., & Killion, T. H. (1977). Interaction of visual and cognitive effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 389–401.

Becker, S., Moscovitch, M., Behrmann, M., & Joordens, S. (1997). Long-term semantic priming: a computational account and empirical evidence. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 1059–1082.

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

Brainerd, C. J., & Reyna, V. F. (1990). Gist is the grist: Fuzzy-trace theory and the new intuitionism. *Developmental Review, 10*, 3–47.

Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science, 11*, 164–169.

Coane, J.H., and McBride, D.M. (in press). The role of test structure in creating false memories. *Memory & Cognition*.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*, 407–428.

Dannenbring, G. L., & Briand, K. (1982). Semantic priming and the word repetition effect in a lexical decision task. *Canadian Journal of Psychology, 36*, 435–444.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate free recall. *Journal of Experimental Psychology, 58*, 17–22.

Durgunoglu, A. Y., & Neely, J. H. (1987). On obtaining episodic priming in a lexical decision task following paired-associate learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 206–222.

Gallo, D. A. (2006). *Associative memory illusions*. New York: Psychology Press.

Gallo, D. A., Roberts, M. J., & Seamon, J. G. (1997). Remembering words not presented in lists: Can we avoid illusory memories? *Psychonomic Bulletin & Review, 4*, 271–276.

Gallo, D. A., Roediger, H. L., III, & McDermott, K. B. (2001). Associative false recognition occurs without strategic criterion shifts. *Psychonomic Bulletin & Review, 8*, 579–586.

Goodwin, K. A., Meissner, C. A., & Ericsson, K. A. (2001). Toward a model of false recall: experimental manipulation of encoding context and the collection of verbal reports. *Memory & Cognition, 29*, 806–819.

Greene, R. L. (2004). Recognition memory for pseudowords. *Journal of Memory and Language, 50*, 259–267.

Hancock, T. W., Hicks, J. L., Marsh, R. L., & Ritschel, L. (2003). Measuring the activation level of critical lures in the Deese–Roediger–McDermott paradigm. *American Journal of Psychology, 116*, 1–14.

Hutchison, K. A., & Balota, D. A. (2005). Decoupling semantic and associative information in false memories: Explorations with semantically ambiguous and unambiguous critical lures. *Journal of Memory and Language, 52*, 1–28.

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

Jacoby, L. L. (1991). A process dissociation framework: separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*, 513–541.

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

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

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

Lovden, M., & Johansson, M. (2003). Are covert verbal responses mediating false implicit memory? *Psychonomic Bulletin & Review, 10*, 724–729.

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.

Masson, M. E. J. (1995). A distributed memory model of semantic priming. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 21*, 3–23.

McBride, D. M., Coane, J. H., & Raulerson, B. A., III (in press). An investigation of false memory in perceptual implicit tasks. *Acta Psychologica*.

McDermott, K. B. (1996). The persistence of false memories in list recall. *Journal of Memory and Language, 35*, 212–230.

McDermott, K. B. (1997). Priming on perceptual implicit memory tests can be achieved through presentation of associates. *Psychonomic Bulletin & Review, 4*, 582–586.

McDermott, K. B., & Roediger, H. L. III, (1998). Attempting to avoid illusory memories: robust false recognition of associates persists under conditions of explicit warnings and immediate testing. *Journal of Memory and Language, 39*, 508–520.

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

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

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

McKone, E. (2004). Distinguishing true from false memories via lexical decision as a perceptual implicit test. *Australian Journal of Psychology, 56*, 42–49.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: roles of inhibition-less spreading activation and limited capacity attention. *Journal of Experimental Pscyhology: General, 106*, 226–254.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: a selective review of current findings and theories. In D. Besner & G. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264–336). Hillsdale, NJ: Erlbaum.

Neely, J. H., Keefe, D. E., & Ross, K. L. (1989). Semantic priming in the lexical decision task: roles of prospective prime-generated expectancies and retrospective semantic matching. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 1003–1019.

Norman, K. A., & Schacter, D. L. (1997). False recognition in younger and older adults: Exploring the characteristics of illusory memories. *Memory & Cognition, 25*, 838–848.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88*, 93–134.

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, III, J. S. Nairne, I. Neath, & A. M. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 95–115). Washington, DC: American Psychological Association Press.

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 create false recall: a multiple regression analysis. *Psychonomic Bulletin & Review, 8*, 385–407.

Roediger, H. L., III, & Gallo, D. A. (2005). Associative memory illusions. In R. Pohl (Ed.), *Cognitive illusions: Fallacies and biases in judgment, thinking, and memory*. Oxford: Oxford University Press.

Schacter, D. L., Cendan, D. L., Dodson, C. S., & Clifford, E. R. (2001). Retrieval conditions and false recognition: testing the distinctiveness heuristic. *Psychonomic Bulletin & Review, 8*, 827–833.

Seamon, J. G., Luo, C. R., Kopecky, J. J., Price, C. A., Rothschild, L., Fung, N. S., et al. (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.

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

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

Toglia, M. P., Neuschatz, J. S., & Goodwin, K. A. (1999). Recall accuracy and illusory memories: When more is less. *Memory, 7*, 233–256.

Tse, C-S., & Neely, J. H. (2005). Assessing activation without source monitoring in the DRM false memory paradigm. *Journal of Memory and Language, 53*, 532–550.

Tulving, E. (1983). *Elements of episodic memory*. Oxford: Clarendon Press.

Tussing, A. A., & Greene, R. L. (1997). False recognition of associates: How robust is the effect? *Psychonomic Bulletin & Review, 4*, 572–576.

Underwood, B. J. (1965). False recognition produced by implicit verbal responses. *Journal of Experimental Psychology, 70*, 122–129.

Vogt, J., & Kimble, G. A. (1973). False recognition as a function of associative proximity. *Journal of Experimental Psychology, 99*, 143–145.

Watson, J. M., Balota, D. A., & Roediger, H. L. III, (2003). Creating false memories with hybrid lists of semantic and phonological associates: over-additive false memories produced by converging associative networks. *Journal of Memory and Language, 49*, 95–118.

Watson, J. M., McDermott, K. B., & Balota, D. A. (2004). Attempting to avoid false memories in the Deese/Roediger-McDermott paradigm: Assessing the combined influence of practice and warnings in young and old adults. *Memory & Cognition, 32*, 135–141.

Whittlesea, B. W. A. (2002). False memory and the discrepancy-attribution hypothesis: the prototype-familiarity illusion. *Journal of Experimental Psychology: General, 131*, 96–115.

Zeelenberg, R., & Pecher, D. (2002). False memories and lexical decision: even twelve primes do not cause long-term semantic priming. *Acta Psychologica, 109*, 269–284.