DRM in Context: Story Context and the False Memory Effect in the Deese-Roediger-McDermott Paradigm

By

Michael D. Lee

University of Manitoba

A Thesis Submitted to the Faculty of Graduate Studies
In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

Department of Psychology
University of Manitoba
Winnipeg, Manitoba

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION

DRM in Context: Story Context and the False Memory Effect in the Deese-Roediger-McDermott Paradigm

BY

Michael D. Lee

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of Manitoba in partial fulfillment of the requirement of the degree

Of

Doctor of Philosophy

Acknowledgements

The road that has finally brought me to the point of writing this section of my completed dissertation has been a long and arduous one. Fortunately, I had the invaluable support of many people who came along for the ride. I am deeply indebted to my committee members, Drs. James Clark, Murray Singer, Jason Leboe, Beverly Zakaluk and my external examiner Steve Dewhurst for their insightful and valuable comments, which were crucial for designing and editing the dissertation. Murray, Jason, and Bev particularly were a great source of inspiration by the knowledge and expertise, from their own respective perspectives, they shared with me in the development of my own ideas on the nature and outcome of this work. A very special thanks to my research advisor, Jim Clark, whose tremendous guidance, friendship, patience and devotion to this dissertation will not soon be forgotten. For an incredible amount of patience and social and emotional support, a loving thank you to my wife, Heather, whose role in the completion of this work was simply indispensable.

Table of Contents

List of Tables

List of Figures

Abstract

The Deese-Roediger-McDermott (DRM) paradigm (Roediger & McDermott, 1995) has become a widely known and highly reliable procedure for demonstrating a powerful memory illusion whereby people falsely remember a word (critical lure) after studying a list of semantic associates of that word (e.g., *bed*, *rest*, *wake*, *dream*, etc., for the lure *sleep*). The DRM paradigm is a notable exception in the false memory literature in that the false memory effect it produces is based on reproductive memory processes as opposed to the reconstructive nature of memory traditionally believed to underlie the production of false memories. The goal of the current set of studies was to add to the nearly fourteen years of empirical and theoretical attention the DRM effect has generated by examining for the first time the nature of the DRM-elicited false memory effect when traditional DRM stimuli are embedded within a text (story) structure. Experiment 1 demonstrates a way to assess the prediction that a story context that relates to the overall theme of a single DRM list enhances false recognition of the critical, nonpresented word for that list of associated items. The results of Experiment 1 showed a clear story effect whereby false alarms to critical lures were stronger following the presentation of a DRM story context than after the presentation of a scrambled, non-semantic version of the same story context. Three additional experiments are reported here that further investigated the hypothesis that narrative text processing enhances the false memory effect in the DRM paradigm. Experiment 2 replicated the first set of results by manipulating context within-participants. Experiment 3 introduced an instructional manipulation so that some participants were given thematic versus list-based instructions. In addition to replicating the results of the first two experiments, the results of Experiment 3 found a clear benefit across context of thematically-based instructions. In the fourth experiment, a levels-of-text-processing manipulation was provided to participants before the presentation of the DRM stories in their narrative format. Depth of processing had a clear effect on false recognition in that study. Overall, the results of this investigation support a story effect with a story context as a novel experimental variable that appears to moderate the standard DRM effect. The results are discussed in terms of two of the principal theories put forth to explain the DRM memory illusion: the activation/monitoring account and the fuzzy trace theory account, in addition to the potential for using DRM narratives, rather than DRM lists, as a method to explore false memories as produced by more naturalistic materials.

DRM in Context: Story Context and the False Memory

Effect in the Deese-Roediger-McDermott Paradigm

Ever since Bartlett (1932) first began empirically investigating what he termed reconstructive memory, as opposed to reproductive memory, the scientific study of memory has continued to explore the malleability of memory. Bartlett's work was in many ways a reaction to the traditional approach used by Ebbinghaus, whose early experimental investigations of memory involved materials that invoked simple rote reproductive strategies, such as lists of nonsense syllables. It was Bartlett's belief that memory is a constructive process and that the comprehension, assimilation, and remembering of information conveyed through more natural materials, such as stories or prose passages, was guided by one's organized mental structures or schemata. Since Bartlett, investigations of false memory have focused primarily on thematic material, guided by the assumption that reconstructive memory processes are more likely to be invoked and that false memories will be more likely to occur. Likewise, it was assumed that learning lists of words would lead to reproductive memory processes, and hence, fewer, if any, false memories.

One notable exception to the emphasis on reconstructive memory has been the Deese-Roediger-McDermott (hereafter, DRM) paradigm which places the study of false memory within a more traditional list-learning framework. By creating false recognition or recall of words not presented in lists, DRM memory researchers have effectively demonstrated the conditions under which reproductive memory becomes reconstructive, perhaps to the point of making the distinction between the two memory processes unnecessary (Roediger & McDermott, 1995).

In a report that has perhaps become better known during the last decade then at the time after its publication, Deese (1959) investigated the probability of extralist intrusions in an immediate free recall task. Using lists developed from Russell and Jenkins (1954) word-association norms, Deese presented participants with lists composed of 12 words associated to a critical, nonpresented word (e.g., *thread, pin, eye, sewing, sharp, point, prick, thimble, haystack, thorn, hurt, injection*, which are all associates to the word *needle*). After presentation of some of these lists, participants' recall of the critical nonpresented words was remarkably high. For other lists, intrusions were low or did not occur at all. Deese (1959) determined that the main factor differentiating these lists (and the probability of false recall) was the mean backward associative strength of the list items. The mean probability with which list items (e.g., *thread*) produced the critical word (*needle*) on a free association test correlated highly (+.87) with the probability of obtaining an intrusion of the critical word for that list.

Roediger and McDermott (1995, Experiment 1) adopted Deese's paradigm using six of his lists which had produced the most intrusions in order to study false recall and recognition. Roediger and McDermott confirmed Deese's (1959) results with these lists; participants recalled critical nonpresented words from which lists were constructed on 40% of trials. The probability of recall of critical words in fact equaled the probability of correctly recalling words that had actually been presented in the middle of the list. On a recognition test that followed the study and recall of all six lists, participants were just as likely to report critical words as "old" as studied items. Roediger and McDermott (1995) conducted a second experiment for which they created 18 new lists of 15 words, including expanded versions of the original six lists. Lists were composed of the first 15 words appearing in the Russell and Jenkins (1954) norms with the strongest associates appearing first in the lists. Following the presentation of each auditorally-presented list, participants were instructed to recall the list or do math problems. The math task was added here because of the possible influence of recall on the subsequent recognition rates in

Experiment 1. Participants recalled 62% of the words they had actually studied and 55% of the time also produced the critical nonpresented item, an even higher rate of false recall than that obtained in their first experiment. On a final recognition test, Roediger and McDermott found remarkably high false alarm rates for the critical lures (.81 for the lists that had been followed by the recall tests and .72 for the lists presented but not followed by initial tests) and these were virtually identical to hit rates for presented items (.79 and .65 respectively). The recognition test also employed the remember/know procedure developed by Tulving (1985) in which participants are asked to classify each item judged old according to whether they can recollect some specific aspect of its moment of occurrence in the list (remember response) or know it was on the list but cannot remember the moment it was presented (a know response). Participants were just as likely to claim to remember some specific aspect of the presentation of the critical items (.48) as they were for studied items (.49).

The levels of false recall and false recognition reported by Roediger and McDermott (1995) were especially surprising given that their experiments were designed to discourage false memories. First, word lists were used, which are generally believed to promote reproductive processes, rather than reconstructive processes. Second, their recall tests immediately followed each list and included instructions against guessing. Finally, unlike other false memory paradigms (e.g., Loftus & Palmer, 1974), there was no attempt to mislead the participants by explicitly presenting false information. Still, the levels of false memories, as indexed by both objective and subjective measures, were among the strongest ever reported in the literature (Roediger, McDermott, & Robinson, 1998).

The basic results obtained by Roediger and McDermott (1995) were so salient that many other researchers subsequently attempted relatively direct replications of these experiments (e.g., Payne, Elie, Blackwell, & Neuschatz, 1996; Schacter, Verfaellie, & Pradere, 1996). The results reported by Roediger and McDermott (1995) have proven to be easily replicable under the conditions they used, and in some ways, the false memory effects obtained from the DRM paradigm were seen as somewhat routine only a few years after they were first reported (Stadler, Roediger, & McDermott, 1999). A spate of research has extended the original report by Roediger and McDermott (1995), as many subsequent researchers pursued a variety of different questions regarding this memory illusion.

DRM Empirical Findings

As stated, the results originally reported by Roediger and McDermott (1995) are so robust, that, similar to a perceptual illusion, the DRM-induced memory illusion is difficult to avoid. Since much has been learned about ordinary perception by the study of perceptual illusions, memory researchers and theorists can similarly benefit from investigations of memory illusions (Roediger, 1996). As such the extant database on the DRM is currently enormous and still growing so rapidly that a complete account of this work would be impossible within these pages. In addition, a large amount of work and attention has been and continues to be devoted to the theoretical processes underlying the false memory effect. In the remainder of this brief introduction, I will highlight some of the empirical findings stemming from DRM investigations and review the major theoretical approaches to understanding and explaining false memory effects.

Backward Associative Strength

It would seem appropriate to begin with an examination of the materials that produce the false memory effect itself, the lists. Recall that Deese (1959) found intrusions for some of his lists, but not for others. To explain this, he proposed that false recall was a function of the average associative strength of the words in a list, or mean backward associative strength. If this is true then shorter lists of strong associates to the critical word should be more likely to elicit false recognition of the critical word than longer lists containing strong and weaker associates to the critical item, because shorter lists will have greater average associative strengths. Using only the strongest associates, Robinson and Roediger (1997) presented participants with lists 3, 6, 9, 12, or 15 items long. Contrary to what one might predict from Deese's hypothesis, false recall and recognition of the critical item showed a pronounced list-length effect with the probability of intrusion at test increasing linearly as a function of list length. In a second experiment, Robinson and Roediger (1997) added unrelated filler words to equate list length and thereby varied the associative strength of each list while holding list length constant. The increase in false recall as a function of number of associated words was not affected by filler items, although veridical recall of studied items did show a decline. The authors concluded that it is not the mean backward associative strength (MBAS) of a list that is crucial to false memory, but rather the total backward associative strength (TBAS) of the list items that predicts false recall and false recognition. What is even more interesting is the inverse relationship between the probability of true and false memory: as list length increased, the probability of accurate recall decreased (list-length effect), whereas the probability of false recall increased. Thus, the critical nonpresented item does not seem to behave as though it were another item in the list, as the results of Roediger and McDermott (1995) appeared to suggest.

It is clear from Deese (1959) and Robinson and Roediger (1997) that the probability of false recall is largely determined by the degree of associative relationship from list items to the critical item. Thus, although the lists used in DRM studies are all constructed by the same means, and include the first 15 associates of a single word that is itself not presented, there is striking variability among different lists in eliciting false recall and false recognition of the associatively related critical item. That backward associative strength is critical to producing false memory is supported by a multiple regression study by Roediger, Watson, McDermott, and Gallo (2001) that determined that the largest contributing factor accounting for most of the variance in false memory across lists was backward associative strength (BAS). Also, given the high positive correlation between BAS and false memory (Deese, 1959; McEvoy, Nelson, & Komatsu, 1999; Stadler, Roediger, & McDermott, 1999), it appears that only lists with strong BAS to the critical item will tend to lead to false memory. In addition, items from lists that have a high probability of eliciting false recall or recognition are also accompanied by more "remember" and high confidence judgments than are lists weak in BAS (Gallo & Roediger, 2002; Gallo & Roediger, 2003). In sum, greater associative strength of list items not only elicits greater levels of false recall and false recognition, but is also positively related to the degree of illusory recollection that accompanies false memory.

Associative Relationships

Yet another characteristic of the word lists used in the DRM paradigm that may add to the reported variability in eliciting false memory is the nature of the associative relation between list items. A closer examination reveals that some lists contain words that are linked by coordinate (horizontal) associations, and others by subordinate (vertical) associations (Park, Shobe, & Kihlstrom, 2005). For example, the words *door*, *glass*, *pane, shade*, etc., are all horizontally-related to the critical lure *window*, whereas the words *apple*, *orange*, *kiwi*, *citrus*, etc., are vertically-related to the critical lure *fruit*. The first set of words are linked at the same level of categorization whereas the second set are linked at different levels of categorization. In two experiments, Park, Shobe, and Kihlstrom (2005) compared the probability of two sets of word lists based on type of association in eliciting false recall and false recognition. They found that the false memory effect was induced by lists composed of coordinate associations but not by lists composed of subordinate or vertical associations. It would seem then that the type or level of association is yet another important variable that mediates the false memory illusion in the DRM paradigm.

Phonological Relationships

The lists used in the DRM paradigm have traditionally been constructed from purely semantic associates. Several researchers have extended the scope of the DRM paradigm by constructing lists of phonological associates that converge upon nonpresented critical items (Chan, McDermott, Watson, & Gallo, 2005; McDermott & Watson, 2001). Phonological associates are typically generated by adding, deleting, or substituting phonemes from critical items. In addition to semantic false memory findings, robust levels of false recall and false recognition have been obtained with lists of phonological associates. Moreover, Watson, Balota, and Roediger (2003) have found that hybrid lists of semantic and phonological associates to critical items produce levels of false memory twice that for lists of pure semantic or pure phonological associates.

Persistence of False Memory

In addition to list characteristics, researchers using the DRM paradigm have also examined the effects of various instructional manipulations at study. For example, there has been considerable interest in the possibility that having knowledge about the false memory phenomenon elicited in the DRM paradigm might reduce or eliminate false recall or recognition. If the false memory illusion is similar to a perceptual illusion then foreknowledge of the effect may have little or no impact. But, if this memory illusion is different from perceptual illusions, in that there is a greater opportunity for memory performance to be influenced by encoding or retrieval strategies, foreknowledge of the illusion should attenuate the effect.

Gallo, Roberts, and Seamon (1997) tested this possibility by comparing groups of participants who differed according to the type of instructions they received. One group received the standard instructions (uninformed condition) similar to the general procedure used by Roediger and McDermott (1995) and a second group (cautious condition) was asked to be cautious on the recognition test in order to minimize the possibility of false recognition. A third group (forewarned condition) was provided with explicit information and examples of the false memory effect prior to the presentation of the study lists. Whereas merely asking participants to be cautious had no effect on false memory, forewarning participants about the nature of the effect reduced the proportion of falsely recognized critical lures. Gallo et al. (1997) also reported that forewarned participants were likely to adopt a strategy of attempting to identify the critical lure during study. Interestingly, however, is the finding that although the forewarned participants actively attempted to identify critical lures they still falsely recognized nearly half of them. Thus, although the effect was diminished under conditions of forewarning, it was certainly not eliminated.

The persistence of the false memory illusion in the study by Gallo et al. (1997) may have been due to the fact that participants engaged in a single trial. If warnings are combined with repeated trials, then it is possible to virtually eliminate the false memory effect (Watson, McDermott, & Balota, 2004). This is similar to the result McDermott (1996, Experiment 2) reported when using a multiple study-test trial procedure in an attempt to reduce or eliminate the false recall effect. Participants were presented with a 45-item list composed of three of the standard lists in blocked order followed immediately by a recall test which was then repeated for four additional trials. Although veridical memory increased over repeated study-test trials, false recall diminished across trials, with the proportion of intrusions dropping between Trials 1 and 5 from .57 to .32. But again, even after five presentations of the list, participants still produced substantial levels of false recall. Thus, it appears that both an explicit warning about the false memory phenomenon for a single trial and an opportunity to improve true memory performance over several trials can attenuate, but not eliminate, the false memory effect. If these procedures are combined, however, there is a greater likelihood of reducing the effect altogether.

Item-Specific Processing

The finding that the false memory illusion is reduced over repeated study-test trials is consistent with the idea that allowing the participant to encode more item-specific information during study can aid in the later discrimination of true from false memories during testing. In addition, it has been reported that allowing more time to study the lists also reduces the false memory effect (Gallo & Roediger, 2002). It may be that both of these situations allow for list items to become more distinctive over time which later prohibits the false recall or recognition of nonpresented items. Factors that make the list items more distinguishable from the critical items may encourage retrieval strategies that aid in the rejection of critical lures. This was clearly evident when participants studied pictures along with list items during study (Israel & Schacter, 1997). They presented participants with lists comprising the top 12 semantic associates to critical nonpresented items, with the provision that all items could be pictorially represented. In one condition, study words were simultaneously presented auditorily and visually and in another condition study words were simultaneously presented auditorily and as black and white line drawings. False recognition of the critical items was considerably lower in the latter condition.

Retention Intervals

An opposite pattern emerges when lists are presented in the typical manner but there is an increase in the time between study and test. McDermott (1996, Experiment 1) introduced a 2-day retention interval and found that the proportion of critical nonpresented items recalled exceeded the proportion of studied items recalled. It may be that as studied items are forgotten over time and item-specific information is lost, there will be a greater reliance on the processes that tend to produce false memories. Payne, Elie, Blackwell, and Neuschatz (1996, Experiment 1) also found that neither false recognition rates nor remember judgments declined over a 24-h retention interval although normal forgetting did occur for the studied items. Thapar and McDermott (2001), on the other hand, demonstrated a decline in false recall and false recognition across retention intervals of 0, 2, and 7 days, although this reduction was less for false memory than it was for veridical memory.

Modality effects

The concept of distinctiveness has also been applied to the so-called modality effect in false memory (Smith & Hunt, 1998). Smith and Hunt argued that simply presenting words visually during study is sufficient to reduce false remembering, relative to the standard condition in which study words are presented auditorily (e.g., Roediger & McDermott, 1995). Smith and Hunt found dramatic reductions in false recall and recognition, about 50%, by merely switching presentation modality from auditory to visual. They proposed that visual presentation enabled more distinctive item-specific processing than auditory presentation. As a result, participants who studied lists visually could more readily discriminate between items actually studied and those that were only thought of during study. These results were replicated by Gallo, McDermott, Percer, and Roediger (2001) in a series of experiments that found both false recall and false recognition on visual tests to be greater following auditory than visual study. These outcomes suggest, therefore, that false remembering is sensitive to study modality.

Levels of processing

Instructional manipulation of levels of processing has also been investigated using the DRM paradigm. Since critical lures tend to be remembered at about the same frequency as studied items, one might expect nonpresented associates to show the same level-of-processing effects as studied words. On the other hand, it is reasonable to predict that deeper processing of the studied items might lead to improved veridical recall, and thereby a diminished false memory effect. Generally, semantic processing has led to increased false recall and false recognition (and veridical memory) over superficial levels of processing (Rhodes & Anastasi, 2000; Thapar & McDermott, 2001; Toglia, Neuschtaz, & Goodwin, 1999), although there are also reports of null levels of processing effects on false memory (Read, 1996; Tussing & Greene, 1997).

An interesting exception to this levels effect was recently reported by Chan, McDermott, Watson, and Gallo (2005). They presented participants either with lists constructed from phonological associates to a critical word (for example, the words *steep*, *weep*, *sweep*, etc. phonologically converge on the critical word *sleep*) or lists comprised of semantic associates (*bed*, *rest*, *yawn*, etc.). Participants had to attend to either the sound of the presented items or their meaning. Chan et al. (2005) found that for lists of converging phonological associates false recall and false recognition was greater when the orienting task required processing the phonological characteristics of the words. Likewise, when lists of converging semantic associates were encoded, an orienting task that focused on the meaning of words enhanced false memory relative to the phonological encoding task.

Studies that have manipulated levels of processing indicate that encoding factors affect later probability of false recall or false recognition. This is consistent with the finding that blocked versus random presentation of the list items enhances recall of both correct items and critical nonpresented items (McDermott, 1996). When associates are presented together at study, the context may increase the probability that each individual item arouses the common associate more than if the associates are presented randomly. Taken together, these studies suggest that participants adopt organizational strategies during the study of list items and that perhaps elaborative, relational processing of associates plays a role in the production of false memory. If true, one would expect that dividing attention during the study phase would diminish false recall and recognition.

Alternatively, several lines of evidence suggest that false memory in the DRM paradigm can be based on largely automatic processing (Seamon, Goodkind, Dumey, Dick, Aufseeser, Strickland, Wouflin & Fung, 2003). For example, when list items are presented at a rapid presentation rate of 20 msec, veridical recall is severely reduced whereas false memory is still observed (Seamon, Luo, & Gallo, 1998). Similarly, dividing attention at study impairs accurate memory but not false memory, specifically recall (Perez-Mata, Read, & Diges, 2002; Seamon et al., 2003). In contrast, false recognition has been reduced when attention is divided at study by secondary tasks that inhibit semantic elaboration of list items (Dewhurst, Barry, & Holmes, 2005; Dewhurst, Barry, Swannell, Holmes, & Bathurst, 2007).

Finally, Dodd and MacLeod (2004) presented DRM lists as coloured words in a Stroop test. After participants had read coloured words, they showed strong memory for list words and strong false memory for nonstudied critical words. However, after participants had named colours (instead of reading words), accurate memory was reduced for list words, but false memory remained high for critical words. This finding is intriguing in that false memory was observed under incidental learning conditions, whereas prior studies have all emphasized or at least encouraged intentional learning of the DRM lists.

Individual Differences

The DRM paradigm has also been used to study populations that may be more susceptible to memory illusions. For instance, Norman and Schacter (1997) found that older adults show more false recognition than do younger adults, but show less accurate memory than do younger adults. Similarly, Balota, Cortese, Duchek, Adams, Roediger, McDermott, & Yerys (1999) investigated false memories in healthy older adults and in adults with dementia of the Alzheimer's type. They found that susceptibility to the false memory effect increased with both age of the participant and the severity of the dementia. Gallo and Roediger (2003) presented younger and older adults lists of associated words in either a visual or auditory format and found that in older adults, source judgments for list items and related lures were less likely to match the modality of the corresponding study list. These results suggest that older adults are less able to encode or to retrieve item-specific details for studied items (Butler, McDaniel, Dornburg, Price, & Roediger, 2004) and that false memory may be due to deficits in source monitoring. Such deficits may reflect poorer frontal lobe functioning in some older adults. Butler et al. (2004) found that older adults who score higher on tests that measure working memory capacity and executive control processes do not show a substantial difference from younger adults in either veridical or false recall. This explanation is consistent with the finding from Melo, Winocur, and Moscovitch (1999) that, relative to controls, nonamnesic patients with frontal lobe damage showed greater susceptibility to the false memory effect.

Individual differences in working memory capacity in young adults has also been investigated (Watson, Bunting, Poole, & Conway, 2005). Watson et al. reported that people with high memory spans were less likely than people with low spans to recall nonpresented critical words and that high spans were more likely to benefit from experimenter-provided warnings about the false memory effect prior to encoding to reduce their susceptibility to false memories. Presumably young adults with low working memory spans demonstrated a breakdown in their ability to actively maintain task goals, such as identifying but not recalling the critical nonpresented word. Other groups have also been reported to show enhanced susceptibility to the false memory effect produced by the DRM paradigm. Women who have claimed to have recovered memories of sexual abuse (Clancy, Schacter, McNally, & Pitman, 2000) and people who report high levels of dissociative experiences and vivid mental imagery (Winograd, Peluso, & Glover, 1998) tend to show elevated levels of false memory.

One group that does not appear to show susceptibility to the false memory effect, however, is children. Dewhurst and Robinson (2004) tested children of three age groups (5, 8, and 11) with a variant of the DRM task. They used five of the lists from Roediger and McDermott (1995) and chose words that would be familiar to children and would allow both semantic and phonological associates. The words were read aloud by the experimenter at a rate of one every 2 s. The child was asked to recall the words at the end of each list. Intrusions were then classified as semantic (semantically related to the list theme), phonological (rhymes of words in the study list), or unrelated. Not surprisingly, there was an age-related increase in the number of correctly recalled words. A similar increase was found for semantic intrusions with 11-year-olds showing significantly more semantic intrusions than phonological or unrelated intrusions. 8-year-olds produced a similar number of each type of intrusion, whereas 5-year-olds were most likely to produce phonological intrusions. Although all children produced false memories, the nature of the false memory varied qualitatively as a function of age, which perhaps represents a developmental shift from phonological processing to semantic processing (Dewhurst & Robinson, 2004). Brainerd, Reyna, and Forrest (2002) have shown that unlike other types of false memory, the false memory effect elicited by the DRM paradigm using semantic associates is virtually absent in very young children but shows a developmental increase. Research on false memory in children is, of course, quite extensive, but specific use of the DRM paradigm with children, among other populations, is just beginning (Brainerd et al., 2002; Howe, 2005).

Theoretical Underpinnings

As can be seen from this brief review, the false memory effect in the DRM paradigm has generated a great deal of empirical attention. Much of this work has gone beyond simple replication and has examined a large number of variables believed to affect the false memory illusion originally reported by Deese (1959) and subsequently Roediger and McDermott (1995). In addition, a great deal of work has been devoted to explaining the false memory effect produced by the DRM procedure. The underlying mechanisms of the DRM-induced false memory illusion remain the subject of much debate. Given the continually expanding corpus of false memory data, theoretical work not surprisingly lags behind. Of course, any theory of the false memory produced by the DRM paradigm will have to be able to accommodate the variety of empirical results in this burgeoning research literature.

Implicit Associative Response

One of the earliest explanations of how false recall and false recognition arise made use of Underwood's (1965) concept of the implicit associative response. Using a continuous recognition paradigm, Underwood had participants read a list of words, which they were to judge as occurring or not occurring earlier in the list. Some words were "critical stimulus words" such as UP. If a participant later is presented with the word DOWN in the list and decides that it was presented earlier in the list, when in fact it had not, then the word UP was assumed to have influenced the false recognition of the word DOWN. The word *down* in this case represents what Underwood referred to as an implicit associative response (IAR) to the word *up,* based on the fact that an IAR is the most frequent associate produced in response to a critical stimulus word in word association norms. A critical stimulus word may therefore produce an IAR, or the unintentional conscious activation of a word that is strongly related to that studied word. In the DRM procedure, critical nonpresented items that are strongly associated to study items in a list may be encoded during study along with the actual study words. Since all of the items in a list converge on a critical lure, that lure is repeatedly activated by the list items through implicit associative responses. False memories can be understood as reality-monitoring errors in which a participant cannot recollect whether the critical lure was simply imagined during study or was actually perceived.

From this perspective, a nonstudied critical lure should behave as though it was actually studied. This is in fact what Roediger and McDermott (1995) observed as the nonpresented items were in fact recalled at the same frequency as presented words in the middle serial position of a study list and were judged to be old at rates comparable to the hit rates for studied items on a test of recognition. The critical lure, however, does not always behave as a studied item. For example, veridical recall tends to decrease when more items are studied (list-length effect) but the same rate of forgetting does not occur for nonpresented items (Robinson & Roediger, 1997). Also, conscious activation of a critical lure during study is not necessary to produce false recognition or false recall (Seamon, Luo, & Gallo, 1998). Thus, whereas the IAR hypothesis alone appears too simple to explain some reported effects, the idea of implicit arousal of a critical lure is a central one in association-based theories of the false memory effect.

Spreading Activation

Another way to interpret the IAR is to assume that the activation of a critical word occurs automatically and unconsciously. That is, false recall or recognition can be explained by assuming that activation of a critical lure spreads throughout a large semantic network, as in the spreading activation theory of Collins and Loftus (1975). False recall and false recognition would then result from spreading activation whereby concepts related in semantic memory are thought to be linked in such a way that accessing one "node" (e.g., *bed*) sends activation across these linked pathways to associated nodes (e.g., *sleep*). Thus, the node of an item such as *sleep* is highly primed by having a list of related words recently presented, and this converging activation might later trigger false recall or recognition. Such a theory can explain the finding that blocked presentation of lists leads to greater false remembering than does random presentation (McDermott, 1996). The spreading activation account is also consistent with the finding that increasing the total number of associates in a list elevates false remembering (Robinson & Roediger, 1997). Also, the degree to which list items evoke associations to the critical item nicely predicts false recall. The regression analysis conducted by Roediger et al. (2001) found the correlation between backward associative strength and false recall was .73. Additional evidence that spreading activation of a critical lure occurs during encoding comes from the finding that false memories occur on implicit memory tests such as word stem and word fragment completion (McDermott, 1997). McDermott (1997) argued that such priming was due to lexical activation of the critical item at study.

One problem with this account is that automatic spreading activation does not explain the subjectively compelling nature of false remembering as measured by the remember/know procedure. Participants in DRM studies typically report the experience of remembering the illusory items. It may therefore be necessary to assume that activation can be based on the conscious process of covertly verbalizing semantically related words in the presence of list associates as in Underwood's IAR (1965) hypothesis (Roediger, McDermott, & Robinson, 1998). Roediger and McDermott (1995) proposed that false remembering may occur because the critical lure is repeatedly activated by the studied items through implicit associative responses, which results in the critical lure's coming to mind (either consciously or unconsciously) during the study episode. During retrieval, participants erroneously falsely recognize or recall the critical item as a studied item due to an inability to distinguish an event that did occur from an event that did not happen.

Activation/Monitoring Framework

The view that false memory occurs as a result of either conscious or unconscious activation during encoding and a failure of source monitoring during retrieval has been combined into a two-component framework typically referred to as an activation/monitoring account (Roediger et al. 2001), although there is no necessary alignment between activation and encoding on the one hand, and monitoring and retrieval on the other. By this account, processing the DRM-list items, at study or test, activates the critical nonpresented associated item and false remembering reflects a failure to correctly monitor the source of this activation. Activation may be due to automatic spreading activation in a lexical/semantic network and/or more conscious elaborative processing as when list items produce implicit associative responses. In either case, false remembering occurs when a participant during study erroneously attributes this activation to the actual presentation of a critical, nonpresented item. Thus, the monitoring component of this theory points to the importance of more strategic, controlled processes which can influence whether activation is translated into a later false memory. Factors that increase activation of the critical item, such as the associative strength of the list items, also increase false remembering. Conversely, factors that increase item-specific processing, thereby facilitating the discriminability between studied and nonstudied items, reduce false remembering. Consistent with a source monitoring framework, older adults tend to show equal or greater levels of false memories, compared to younger adults, even though they tend to show a decline in veridical memory (Balota et al. 1999; Norman & Schacter, 1997). Older adults, relative to younger adults, are also less able to take advantage of multiple study-test trials and experimenter-provided warnings about the nature of the false memory effect (Watson et al. 2003; Watson, McDermott, & Balota, 2004). One explanation for this difference between younger and older adults is that although automatic associative processes are spared with age, monitoring processes are sometimes impaired by frontal cortex dysfunction (Balota et al. 1999). In sum, the activation/monitoring theory favoured by Roediger and his colleagues is currently the most widely accepted framework for explaining and understanding the DRM-induced false memory effect.

Fuzzy-Trace Theory

A second major theoretical explanation of the DRM illusion is based on fuzzy-trace theory (Brainerd & Reyna, 1996; 1998; Brainerd, Wright, Reyna, & Mojardin, 2001). An important theoretical principle of this framework is the fuzzy-to-verbatim continua principle, which asserts that multiple mental representations of experience are stored in memory and these differ with respect to the precision with which they specify those experiences (Payne et al. 1996).

Specifically, the presentation of items in a DRM list is hypothesized to result in the formation of two types of memory representations. Verbatim representations are memory traces thought to correspond to the individual items presented to participants during a study phase. Verbatim traces contain information about surface forms and other item-specific information. Gist representations are memory traces that specify the more general semantic content and thematic characteristics of the list items. The process of establishing a gist representation involves what is called gist extraction which allows people to store the semantic, relational, and elaborative properties of studied items (Reyna & Brainerd, 1998). Although these representations are believed to be encoded in parallel, they are dissociated during retrieval. At test, verbatim representations should support veridical memory for studied items, whereas gist representations support the false recognition or false recall of nonstudied, but highly associated items. From this perspective, false memory is primarily a retrieval phenomenon in that the presentation of a critical lure at test evokes gist traces which convey the information that the list items represent a common theme. Words that are consistent or similar with the gist of the list should be highly familiar to participants, and therefore, falsely remembered as having been presented. Thus, false memory is caused by the similarity between the critical item and the studied items, as opposed to the associative activation of the critical item (Roediger & Gallo, 2004).

Retrieval of verbatim representations supports feelings of item-specific information for studied items, which in turn facilitates acceptance of target items and recollection rejection of critical lures. Retrieval of gist representations supports feelings of nonspecific resemblance between studied items and critical lures and hence, the false alarms to critical, nonpresented items. According to Brainerd and Reyna (1998) there should be variables that increase or decrease the probability that retrieval will be based on either verbatim or gist representations.

One such variable is retention interval; specifically, the time course of accessibility to gist and verbatim representations differs in that gist representations are said to be more durable than are verbatim representations which decay more rapidly over time. In support of this assumption, it has been found that veridical memory decreases more over a delay than false memory (McDermott, 1996; Toglia, Neuschatz, & Goodwin, 1999). Fuzzy-trace theory can accommodate the effects of retention interval because it assumes that gist traces are more resistant to forgetting than verbatim traces. Activation-based accounts have trouble accounting for this finding, particularly when the initial levels of true and false memory are equivalent. Fuzzy-trace theory can also account for findings originally interpreted in terms of spreading activation. For example, the blocked/random effect reported by McDermott (1996), in which false memory is higher when lists are blocked by theme, could be explained by supposing that blocked presentation facilitates gist extraction from the lists, which in turn, facilitates false memory of the critical lures from those lists.

For the most part, both theories make similar predictions regarding the effects of several variables, since the opponent processes of activation and monitoring are operationally similar to the opponent processes of generation and recollection rejection proposed by fuzzy-trace theory (Seamon et al. 2003). The processes underlying the subjective reports of illusory recollections that often accompany false memory of critical nonpresented items, however, are not well understood (Gallo & Roediger, 2003). In some cases, such retrospective reports have involved rather detailed recollection. For example, when lists are presented in different modalities (auditory vs. visual) or different voices (male vs. female), participants will often assign a source to critical lures that are falsely remembered. Using the Memory Characteristics Questionnaire (MCQ), participants often claim to recollect specific details about lures that were not actually presented, such as perceptual features, their position in lists, and personal reactions to the words (Gallo & Roediger, 2003).

One explanation for these subjective phenomena is a fluency-based attribution process (Gallo & Roediger, 2003; Roediger & Gallo, 2004). Gallo and Roediger (2003) proposed that when making a remember/know judgment about a critical lure, participants might try to imagine features of the related lure's presentation in an effort to try to remember. If the prior associative activation of the critical lure makes the subsequent processing of that lure more fluent, these imagined features may be falsely attributed to true memories. At the same time, this attribution process can also occur automatically, producing a phenomenological experience of remembering the critical lure. Similarity-based mechanisms, such as those proposed by fuzzy-trace theory (Brainerd et al. 2001) may also contribute to fluent processing of the critical lure. According to this approach, false memory is based on the semantic overlap between the critical and studied items, which produces a strong sense of familiarity for the critical items. When this overlap is particularly salient, as with highly associated lists, the retrieval of strong gist representations may be associated with vivid perceptual details that are characteristic of verbatim traces, or what Brainerd et al. (2001) refer to as phantom recollection. Brainerd et al. (2001) extended fuzzy-trace theory by introducing a parameter for phantom recollection to mathematically model illusory conscious experiences. Gallo and Roediger (2002; 2003) have argued, however, that associative mechanisms play a larger role than do similarity-based mechanisms in producing both false memories and subjective illusory recollections. Increasing associative activation (via list length or manipulating mean backward associative strength) not only increases levels of false recall and false recognition, but it also enhances the attribution process they believe drives illusory recollection.

Despite an abundance of theoretical development pertaining to the DRM illusion, there is clearly a lack of evidence directly comparing these two major competing accounts against each other. This lack of evidence leaves open a critical and, as yet, unresolved issue. That is, what is the exact nature of the associative relationship of the items in DRM lists that produces false memory? More fundamentally, this question relates to whether the mechanism underlying the false memory effect is one of lexical associative activation or the formation of meaning-based representations. In other words, how important is meaning to the DRM memory illusion? Part of the problem is that words in the typical DRM lists are derived from the same word association norms, which may or may not be based on semantic relatedness. Given the strong correlation between backward associative strength and false memory, it is important to note that lists high in BAS do not necessarily consist of words that are related at the semantic level. As noted by Gallo and Roediger (2002), there are a variety of factors, such as statistical co-occurrence in natural language, which could cause one concept to activate another (e.g., *cradle-baby*). A further problem is that semantically similar words also tend to be associated (e.g., *hot-warm*).

Therefore, it would be interesting to compare lists constructed of words associated by non-semantic features with lists that are primarily based on semantic relatedness. This could be done by comparing existing DRM lists based on this distinction or creating new lists that are either high in non-semantic, or at least minimally semantic, associative structure or high in semantic overlap but which do not rely on associations. Lists of the first type should be expected to produce high levels of false recall or recognition compared to lists of the latter sort if false memory is indeed driven by the activation of preexisting associations rather than the extraction of thematic representations based on similarity of meaning. Another way, using current DRM lists, might be through the disruption of gist-based processing by presenting mixed-item lists. It would be possible to compare standard lists with varying levels of multiple-theme lists comprised of 2, 3, or 4 themes in a single list. Since both gist and verbatim representations support veridical memory, and assuming that implicit associations summate to support false memory, one might expect a dissociation between true and false memory with a greater reduction in veridical memory for studied items as a function of increasing the number of themes per list as opposed to false memory for critical lures. Whatever the methodology, certainly a critical goal for future research will be to present a clearer picture of the underlying mechanisms of the DRM-induced memory illusion as proposed by leading theories. One way to further this goal is to examine the effects on false memory of higher level semantic processing, as in text comprehension. That is the approach taken in the present series of studies.

DRM in Context

As described in the introduction, the DRM paradigm arose as a list-learning paradigm to examine the nature of false memories, which have traditionally been investigated with sentence- and prose-based materials. Using standard word lists in place of contextual materials to elicit false memory effectively served to bridge the traditional distinction between reconstructive and reproductive memory. With a few exceptions (e.g., Deese, 1959; Underwood, 1965), researchers interested in false memory had primarily followed in the path started by Bartlett (1932) by examining errors in memory from the perspective of reconstructive memory. By resurrecting methods used by Deese (1959), Roediger and McDermott (1995) introduced a new paradigm for investigating false memory that has since become the dominant method and a hugely popular one, due in large part to its simplicity and reliability, for a whole new generation of memory researchers interested in examining the nature of false memory. This paradigm relies on presenting semantically-related items in a list followed by a standard test of memory recall and/or recognition of those items, including words not presented, but associated with those items, in the list.

In the last 14 years, researchers have elucidated a variety of factors and variables within the DRM paradigm that have provided a wealth of empirical information about how false memory, as elicited by this paradigm, is strengthened, weakened, modulated, etc. Whereas there is no reason to believe that these basic investigations will not continue for many years to come, some researchers are beginning to take the study of false memory in new directions. For example, McDermott (2007) and colleagues (e.g. Chan & McDermott, 2006) have looked at whether the typical results of DRM studies are applicable to understanding false memories as they occur in more ordinary, natural situations. Specifically, these researchers have been looking into the phenomenological experience underlying pragmatic inferences when such inferences are illusory representations of sentences that were never explicitly stated. Using Tulving's (1985) remember / know procedure, as in Roediger and McDermott (1995), the general pattern is that 'remember' responses to falsely recognized pragmatic inferences are similar to those for actually studied sentences, as they are for critical lures and studied items in the DRM list-learning paradigm (Chan & McDermott, 1996; McDermott & Chan, 2006; Roediger & McDermott, 1995). These findings help support the assertion that the mechanism(s) underlying the false memory effect found in the DRM paradigm may be similar to the types of false memory elicited by a wide variety of alternative procedures, including memory distortions produced by script-based inferences (Dewhurst, Holmes, Swannel, & Barry, 2008) and discourse-based inferential processes (Singer & Remillard, 2008).

The purpose of the current series of studies was similar in that it represented a departure from traditional DRM research and an extension into prose, effectively returning the investigation of false memory to its more traditional beginnings. Unlike the research just described, however, the principal goal of this investigation was to examine the effects of story context in eliciting false memory with traditional DRM stimuli and materials. Based on extant theories for typical DRM findings (e.g., activation-monitoring, fuzzy trace theory), and findings that word lists show a levels-of-processing effect, embedding DRM items into short stories that are consistent with the overall theme of the list would appear to be another, as yet largely unexplored, valuable avenue of investigation. It is possible that presenting DRM words in this way would produce a unique levels-of-processing effect by encouraging semantic processing. A thematic story context consistent with and including associated items that converge on a non-presented critical lure could increase the likelihood of falsely recalling or recognizing that critical item relative to the standard list condition. That is, the effects of story context could raise false recall and/or recognition to levels above that which is already found under standard list presentation conditions. While a number of studies, as seen in the above review of the literature, have investigated many different variables shown to modulate the standard effect, no study to date has looked at the effect of story context on DRM list items with adults.

A study recently conducted by Dewhurst, Pursglove, and Lewis (2007) represents pioneering work in that it was the first report on the effects of story context within the DRM paradigm among a younger population, and as such has become a foundation for the current set of studies. As described above, younger children typically fail to show the false memory effect in the DRM paradigm, at least when semantically-associated lists are used. Their lack of susceptibility to the effect could be explained by the fact that younger children simply do not have sufficient semantic knowledge to either consciously or unconsciously relate the list associates to their overall theme, and therefore, fail to activate the critical lure. Alternatively, their relative invulnerability to false recognition or recall could be due to their inability to extract the gist representation of the list and without such gist traces of the list theme, there is greater protection against falsely remembering a critical lure (Brainerd, et. al., 1998). Dewhurst et al. (2007) proposed that children might become more susceptible to the illusion if their ability to identify the theme of a list is enhanced by presenting list words in a story context that highlights its overall theme.

To test this hypothesis, Dewhurst et al. (2007) introduced an interesting paradigm by creating eight short stories using the list items from eight, standard DRM lists. Both standard list presentations and story presentations were compared among groups of 5-, 8-, and 11-year-old children. Compared to the older age groups, 5-year-olds falsely recognized significantly more critical lures when the DRM stimuli were presented in stories and fewer when presented in lists, providing tentative support for the prediction that younger children's false memories in the DRM paradigm might increase as a function of story context. Curiously, and more importantly for present purposes, story context did not lead to the same increase in false recognition compared to list presentations in the older children, producing overall lower rates of false recognition. Contrary to the goal of the four experiments reported here, which were based on the prediction that story contexts might facilitate or enhance the false memory effect in adults over and above the traditional list format, the results of Dewhurst et al. (2007) provide negative evidence for such an effect. Is it because, as Dewhurst et al. (2007) offered, that older children (and presumably therefore, adults as well) are already able to establish a more efficient gist representation of list items during study, and use verbatim traces to aid in the monitoring of items at test? If so, it is possible that the facilitation created by story contexts in younger childrens' false memories may be offset by the greater ability in older children (and adults) to rely on more effective source monitoring to reject those related items that were not actually presented in lists or stories.

Although interesting and informative, these developmental trends in the DRM effect were not the focus here. Only through an investigation of story context with adult participants might some of the questions with respect to the negative finding in older children reported by Dewhurst et al. (2007) above be answered. Thus, the principal objective was to examine further the effects, if any, on false memories in adults when DRM stimuli are presented in story contexts. The timing of the Dewhurst et al. (2007) study was fortunate as it provided both a novel methodology and a set of materials (all eight of the original stories used in that study were kindly provided upon request) to examine, for the first time, the effects of story context with a more typical population of adult participants.

As mentioned above, the fact that Dewhurst et al. (2007) failed to find an effect of story context with older children suggests, by implication, that the same may hold true for adults, perhaps for the reasons just described. There may also be, however, some other explanations for those negative results. A closer examination of the methodology and materials used by Dewhurst et al. (2007) revealed some possible confounds and other limitations that may have contributed to the lack of a story effect in the older children relative to the younger age group. In some cases, in the context of their original study, these would not constitute limitations but were perhaps necessitated by virtue of working with children, which would still nonetheless need to be appropriately modified for use with adult participants.

First, Dewhurst et al. (2007) compared stories that were created using the DRM stimuli and those stimuli as presented in a traditional list format. This meant that participants in the story condition would have heard at study many more words than in the list condition, including a number of additional function words. Whereas the number of words in each of the lists in the list condition consisted of only 14 items, the number of words in the stories ranged from 65 to 104 words. Second, words were presented at a rate of 2 seconds per word (for a total of 28 seconds) in the list condition, whereas it took anywhere from 30 to 40 seconds to read each of the eight stories. This means that participants in the story condition not only heard a significantly higher number of words, but also at a significantly higher reading rate, necessarily reducing the duration or time for each of the individual words to be processed at study by participants, compared to those in the list condition.

Finally, single recognition tests for the stories and lists were given immediately after the presentation of each of the eight individual stories or lists, and each recognition test consisted of all of the fourteen studied items and only five unrelated lures, in addition to the critical lures. The standard procedure is to present on a final test a smaller number of studied items from each list and an equal number of items from nonstudied lists, in addition to the critical lures.

It is possible that any or all of the above factors may have confounded the comparisons between story and list presentations in the Dewhurst et al. (2007) study. Eliminating these confounds could reveal the positive effect of story context on false memory that is expected given the hypothesized role of semantic processes in the DRM effect. The goal of Experiment 1 was to attempt to control or at least reduce these factors and see whether the negative results reported by Dewhurst et al. (2007) replicate with adults, or if holding these factors constant across study conditions would reveal that story context indeed enhances the typical false memory effects elicited by DRM stimuli, as originally predicted. This prediction is based on assumptions about the underlying mechanisms that have been put forth to explain the relatively high false alarm rates to critical lures, which can in fact approach the levels of hits to studied items. The assumptions about spreading activation and semantic or gist processing that serve as the basis for the current prediction are consistent with either of the two major, current theoretical accounts of the false memory effect in the DRM paradigm. The assumptions also parallel the rationale given by Dewhurst et al. (2007) for the use of story context in bringing out the false memory effect in children, who typically do not demonstrate the standard false memory effect.

If false recognition is tied to gist extraction of a list's theme, as according to the FTT account, then embedding DRM stimuli in a story context that promotes gist processing even further predicts higher rates of false recognition compared to the presentation of the same DRM stimuli in a list of associated items. Alternatively, if false recognition is based on the unconscious activation or conscious elaboration of a critical associate as a function of semantically-related stimuli presented in lists, as according to the activation/monitoring account, then one predicts a greater degree of false recognition of that critical associate when those stimuli are presented in a story context that further enhances semantic processing, relative to a list of semantic associates absent text cues.

In other words, if the story context as a whole relates to the theme of a list, there may be an even stronger tendency under those conditions to gist extraction or associative activation leading to enhanced false recognition of those items or critical lures that represent the theme of the story. False recognition might be higher because in addition to presenting list items that already converge on a semantic associate, a thematic story context itself is an additional factor over and above the individual list items that may drive the underlying mechanisms toward semantic processing and the production of false memory. Regardless of the actual underlying mechanism that is responsible for the effect, it is predicted that the processing of a thematic story context in which DRM stimuli are embedded will lead to even higher rates of false recognition of items highly related to those stories, compared to when false recognition is obtained by the presentation of DRM stimuli in only lists.

EXPERIMENT 1

Although there now exists a finding in Dewhurst et al. (2007) that contradicts the prediction outlined above, the goal of Experiment 1 was to modify the approach used in that study to provide a more appropriate test of the predicted benefits of story context. In addition to some other changes described below, one major addition to this experiment was the inclusion of a third condition, along with the story and list presentations. Specifically, Experiment 1 included a version of the stories in which the additional words were randomly scrambled to weaken and ideally eliminate any text effect, yet keep the number of items the same as in the story condition. This scrambled text was meant to serve as a more appropriate comparison than a simple list condition to the normal story condition where the coherence of the theme was maintained, but where there were also an equal number of identical words. The filler words in each story were randomly scrambled to create these scrambled lists of words, and the associated items remained in their respective positions in both versions of the context lists. Also the stories, adapted from Dewhurst's et al. (2007) study, were presented as lists, one word at a time, but in a coherent, story format that retained the theme of the story. This was done to ensure that the story contexts were presented and processed in a manner identical to that of the normal lists and scrambled context lists. The lists presented in the standard list context condition consisted of only the associated items but with added spacing between those items wherever a word from the story or scrambled context lists was present. This further equated all three sets of lists in that critical items occurred at exactly the same point in time. That is, the lists of associated items were interspersed with blanks corresponding to each additional word in the story and scrambled conditions so that the associated items appeared in the same positions in each of three sets of lists.

The above procedures resulted in three sets of list conditions, identical in terms of presentation time and appearance of studied items: list context, scrambled context, and story context. These conditions were presented between-participants in Experiment 1. An example of these three conditions for one story is shown in Appendix A. The stories were also re-worked to make them shorter than the originals, by reducing the number and frequency of extraneous function and content words, where possible, without losing the thematic nature of the stories. In particular, added content words that arguably could directly prime the associated item were removed.

Method

Participants

One hundred and thirty-one Introductory Psychology students at the University of Winnipeg voluntarily participated in this experiment as part of a course requirement. The only requirement for participation was that participants be fluent in English.

Design and Materials

The eight stories used by Dewhurst et al. (2007) were adapted for this experiment. These eight stories were based on lists selected from Roediger and McDermott (1995) and consisted of semantic associates for each of the following critical lures: sleep, smell, doctor, lion, cold, music, thief, and fruit. Each list consisted of 15 converging associates to their respective lures, although, following the procedure of Dewhurst et al.(2007), only 14 out of the 15 associates were used in the stories, with the one omitted item serving as a second critical lure on the recognition test.

The eight stories (see Appendix B) were divided into two sets for each of the three conditions for counterbalancing purposes, and each provided a set of nonstudied items in the recognition test. The two sets of stories, A and B, consisted of associated items that were equal with respect to their frequency of recognition probabilities, ($M = .67$) based on norms established by Stadler et al. (1999). The average number of words for each set of stories was practically identical ($M = 51.0$ for set A, $M = 49.5$ for set B). For each set of four stories, three presentation conditions were constructed corresponding to the example in Appendix A: a Story context condition in which the list of items were presented in the order that they appeared in the newly created thematic story; a List context condition in which all of the words from the stories except for the associated items were omitted; and a Scrambled context condition in which the stories from the first condition were presented in a randomized order except for the associated items which were held in the same positions as they appeared in the story and list context conditions. Additional counterbalancing was achieved by creating two orders of list presentations within each of the three conditions for both sets of stories, and two orders of items for the recognition test.

The recognition test consisted of 48 items. Studied lists contributed 8 critical lures (2 per list) and 16 studied items (four items per list corresponding to positions 2, 4, 6, and 8). The lists from the second, non-studied set of lists also contributed eight non-critical lures and sixteen non-studied items from their corresponding positions in the lists. For additional counterbalancing, two orders of the recognition test were created. Participants were asked to provide a confidence rating from 1-4 to indicate whether or not they had heard each item presented earlier in the study phase. A rating of 1 corresponded to *sure new*, 2 to *probably new*, 3 to *probably old*, and 4 to *sure old*. Thus, participants were instructed that if they were positive an item on the recognition test was not presented earlier they were to give that item a rating of 1, and if they were positive that an item on the test was presented earlier they were to give that item a rating of 4, and so on.

The presentation of each list was prepared by creating powerpoint files for each of the six orders within the two sets of stories, resulting in twelve different stimulus files. Words in the Story, Scrambled, and List context conditions were put into a slideshow format so that the timing of presentation could be controlled. In the List context condition, where there were many absent words, blank slides were created to maintain the timing of critical word presentation constant across conditions. The rate of presentation of each slide was set at 1 second, with a short pause between each list and cues signaling the start and end of lists. While running the powerpoint files, the stimuli were then read and recorded by a digital voice recorder and the twelve media files were burned to a compact disc for later playback on a portable stereo system.

Procedure

Participants were run in groups of 10-12 each and each group was randomly assigned to one of the 12 possible conditions defined by three conditions, two sets of items, and two presentation orders. The actual experiment took place in a small classroom. The instructions for each of the three main conditions were identical. At the start of each experiment, all participants were told that they were going to hear four short lists of words and to listen carefully as their memory for some of those words in the lists would be later tested. The study part of the experiment took about 5 minutes to complete. Following this study phase, participants opened booklets and were instructed to complete a mental arithmetic task, which served as a filler task between the study and test phases of the experiment. This task involved completing as many multiplication problems as they could in three minutes.

Following the filler task, participants received instructions for the recognition task. Participants were allowed to complete the recognition task at their own pace but took no longer than five minutes, on average, once started. The entire experimental session took approximately 25-30 minutes to complete.

Results and Discussion

Table 1 shows mean recognition ratings for primed and unprimed lures and for list items in the Story, Scrambled, and List Context conditions. Primed lures refer to critical and/or weakly-related non-presented items when their related list associates were studied, and unprimed lures to those items when their related list associates were not studied. List items (studied or nonstudied) refer to the presented items in the lists of related associates. A comparison of recognition performance by participants in the Story vs. Scrambled list conditions provided the strongest test of a story effect on false recognition (bold values in Table 1). As such, initial analyses focused only on the rate of false alarms to the 16 critical lures in the 48-item recognition test across both sets of lists when their list associates were either studied or not studied. Thus, the 8 critical lures per set served as their own control on the recognition test.

Table 2 reports mean recognition ratings for the lures when their respective list associates were either studied or not studied in either the Story or Scrambled context conditions. These analyses also took into account the type of lure, since half the lures in each set consisted of the standard associates from the norms (e.g., *sleep* for list items *bed*, *rest*, *wake*, etc.) and half were items that Dewhurst et al. (2007) had selected from the DRM list (e.g., *nap*). In these and subsequent analyses, normative lures were referred to as *critical* lures, and the Dewhurst lures as *weakly-related*, or simply, *weak* lures.

Table 1. Mean recognition ratings for lures and list items as a function of Context in Experiment 1.

Item Type	Context: Scrambled	Story	List
	Critical Lures		
Primed	3.39	3.68	3.43
Unprimed	2.19	1.92	1.72
Corrected False Recognition	***1.20***	***1.76***	*1.71*
	Weak Lures		
Primed	2.89	3.16	2.84
Unprimed	2.06	1.76	1.58
Corrected False Recognition	***0.83***	***1.40***	*1.26*
	List Items		
Studied	3.35	3.50	3.64
Nonstudied	2.23	1.94	1.72
Corrected Recognition	*1.12*	*1.56*	*1.92*

Table 2. Mean recognition ratings combined for primed and unprimed critical and weakly-related lures as a function of Context in Experiment 1.

	Context	
Lures	Scrambled	Story
Primed	3.14	3.42
Unprimed	2.13	1.84

A 2 x 2 x 2 analysis of variance was performed with one between-subject variable: Context (Scrambled versus Story context), and two within-subject variables: lure type (critical versus weakly-related lures) and lure status (unprimed versus primed). A main effect for lure status was found: F (1, 84) = 276.48, MSE = .523, p < .001 as well as an interaction between lure status and Context: F (1, 84) = 13.43, MSE = .523. p < .001. There was also a main effect for lure type F (1, 84) = 43.58, MSE = .217, p < .001. Thus, overall, the set of standard, critical lures were more often falsely recognized than the set of weakly-related lures. There was no significant interaction, however, between lure type and Context: F (1, 84) = .072, MSE = .217, p = .788.

In summary, for both types of lures, the difference in false alarm rates between primed lures and unprimed lures was greater in the Story context condition than in the Scrambled context condition. Figures 1 and 2 show the interaction between list type and lure status as a function of lure type (critical lures in Figure 1 and weakly-related lures in Figure 2).

Figure 1. Mean ratings for primed and unprimed critical lures as a function of Context in Experiment 1.

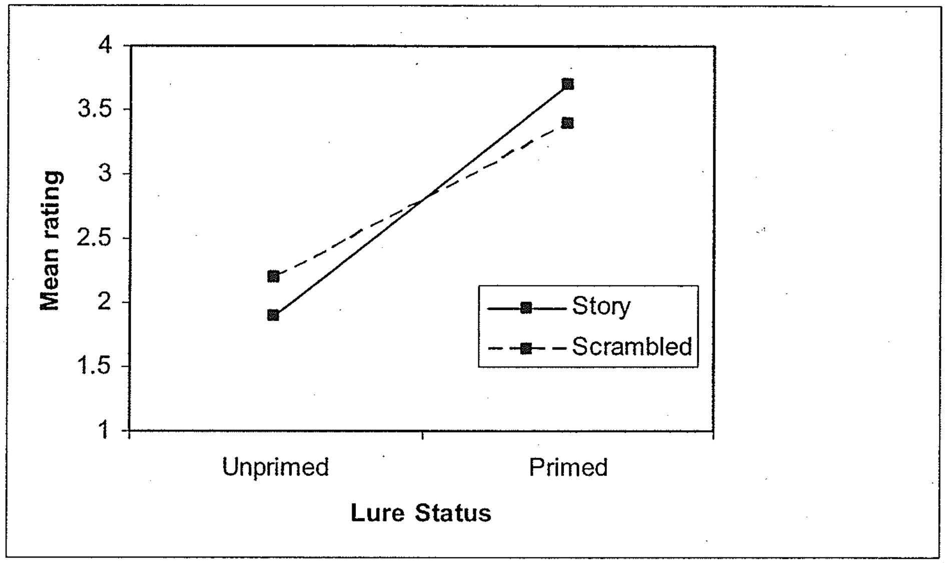
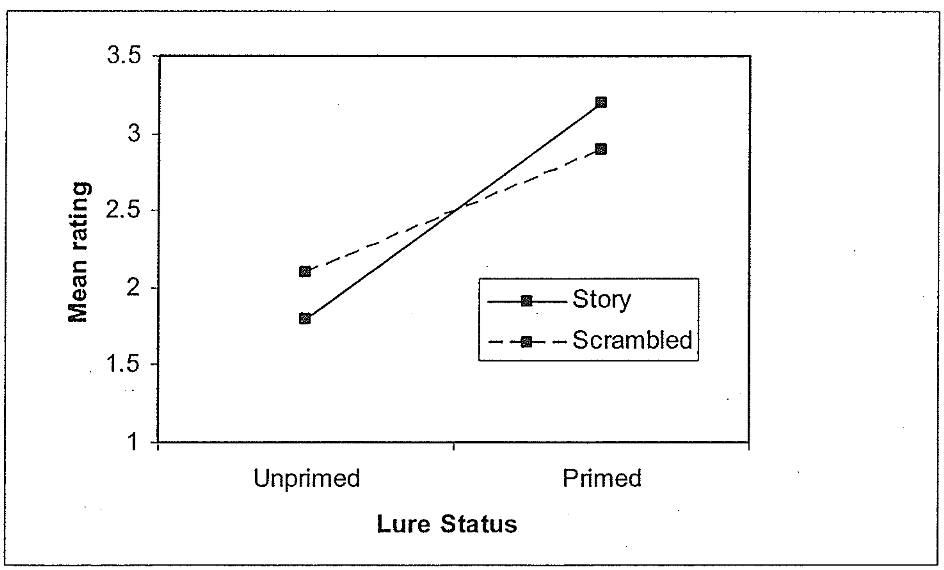


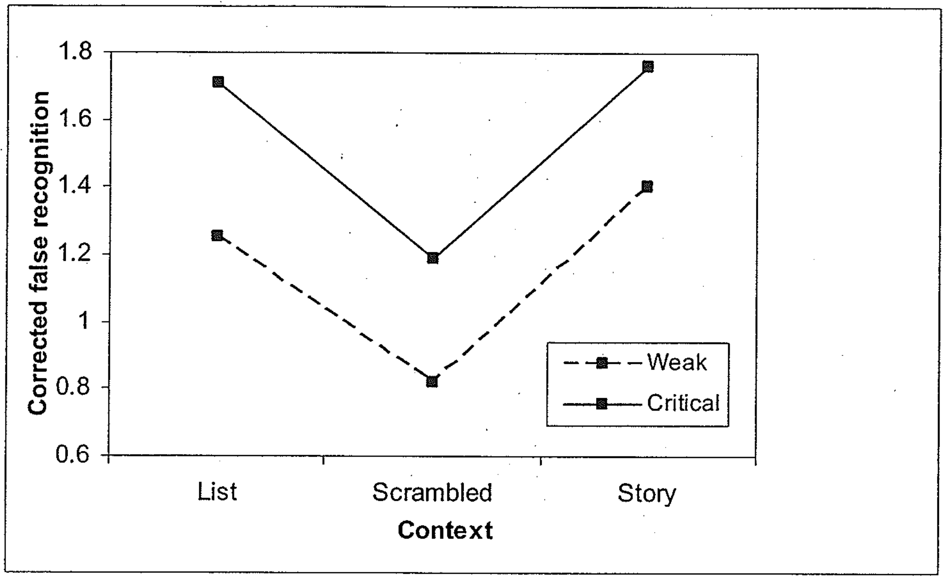
Figure 2. Mean ratings for primed and unprimed weakly-related lures as a function of Context in Experiment 1.



Given the opposite effect of Context on Primed (higher ratings for Story condition) and Unprimed (Higher ratings for Scrambled condition) Lures and to more clearly demonstrate the results described above, corrected recognition ratings were obtained by computing difference scores between the mean ratings for lures that belonged to studied lists (primed lures) and the lures that belonged to non-studied lists (unprimed lures). All subsequent data analyses will report these corrected ratings. To confirm that rating difference scores correlated well with other standard measures used in recognition memory studies, ratings were dichotomized as "Hits" (ratings of 3 or 4 for lures from studied lists) or "False Alarms" (ratings of 3 or 4 for lures from non-studied lists). These Hits and False Alarms were used to compute Hits Minus False Alarm scores, as used in many recognition memory studies (e.g., Keane, Orlando, & Verfaellie, 2006), as well as d-prime, another standard measure (Stanislaw & Todorov, 1999). The rating difference scores correlated highly both with Hits Minus False Alarms ($r = .921$) and with d' ($r = .911$). Given these correspondences and the common use of such difference scores for similar cognitive tasks (e.g., completion rates for primed and unprimed word fragments in implicit memory tasks and diverse priming effects on RTs in lexical decision and naming tasks), the difference in ratings (i.e., Corrected Recognition Ratings) were used in subsequent analyses.

Using these corrected recognition ratings for all three contexts (Story, List, and Scrambled), there was a main effect of lure type: $F(1, 128) = 26.034$, $MSE = .394$, $p < .001$ and no significant interaction between lure type and Context: $F(2, 128) = .154$, $MSE = .394$, $p = .858$, meaning that in all conditions, false recognition was significantly higher for critical than for weak lures. Controlling for lure type, there was a significant main effect of Context on false recognition: $F(2, 128) = 7.523$, $MSE = 1.080$, $p < .001$ (see Figure 3). Figure 3 demonstrates the effect of lure type on false recognition as well as the contrasts between conditions. Pairwise comparisons indicated that the List context differed from the Scrambled context, $F(1, 87) = 8.359$, $MSE = 1.186$, $p < .01$, but not from the Story context, $F(1, 85) = .432$, $MSE = 1.005$, $p = .513$.

Figure 3. Corrected false recognition of critical and weak lures as a function of Context in Experiment 1.



The results clearly show that when a number of extra words are randomly inserted into the standard list context, as in the scrambled context condition, the false memory effect is greatly reduced. But when these same words are re-arranged to form a coherent narrative, as in the story context condition, the false recognition effect returns to its former level. For this reason, the principal comparison, both earlier and using the corrected recognition ratings, is that between the story and scrambled context conditions. Again, when Story and Scrambled conditions were compared, there was a significant main effect of context on false recognition, F (1, 84) = 13.435, MSE = 1.046, p < .001, with the Story context producing significantly higher rates of false recognition than the Scrambled context. There was also a significant main effect of lure type, F (1, 84) = 16.071, MSE = .358, p < .001. Critical lures were significantly more often falsely identified as old than were weak lures in both Story and Scrambled conditions, with no evidence of an interaction between these factors, F (1, 84) = .010, MSE = .358, p = .922.

A similar benefit of context emerged when list items were analyzed. For corrected true recognition ratings, there was a main effect of Context with a greater degree of correct recognition for studied items in the Story condition ($M = 1.56$) than in the Scrambled condition ($M = 1.11$), $F(1, 84) = 11.586$, $MSE = .370$, $p = .001$. Furthermore, when studied items were controlled, and included as a co-variate, the story versus scrambled effect on false recognition became only marginally significant, $F(1, 83) = 3.296$, $MSE = .633$, $p = .073$. This strong relationship between true and false recognition suggests that the tendency to identify as old the non-presented lures depended on the tendency to also correctly recognize their highly associated list words as old, such that participants were essentially treating the lures as though they were actually presented, studied items.

The results of Experiment 1 support the predicted effect of story context; the DRM effect was stronger for Story than Scrambled contexts. When DRM stimuli are embedded in a story that relates to the overall theme of these stimuli, false recognition of their converging associates is clearly higher than is the case when the DRM stimuli are interspersed among the words from that story when those words are presented in a way that reduces or even eliminates the text structure of the story context. It is worth repeating here that the Story and Scrambled list contexts were identical with respect to the words processed by participants in addition to the total number of words studied, and that the crucial difference between the two conditions was the presence or absence of semanticity.

EXPERIMENT 2

Recall that one of the principal explanations for the DRM effect is associative activation, which refers to the activation of one concept stored in semantic memory due to the processing of other concepts found at the same conceptual level (Gallo, 2006). Most often associative links are based on semantic similarity but other factors, such as frequent co-occurrence in the language, can also contribute to the strength of connections in these semantically organized networks. A critical assumption of this model is that the processing of one word (e.g., bed) activates neighbouring concepts (e.g., sleep). When several associates of a word are processed, activation of that word is believed to be even stronger and more durable, as in the basic DRM paradigm. The hypothesis here is that embedding DRM list words into a coherent story that relates to the overall theme of those associated list words leads to even greater activation of those concepts due to thematic processing as a function of strengthening their connections, including increased activation and strengthening of concepts or words that are not presented but are strongly associated to presented items (i.e., critical lures).

The results of Experiment 1 provided evidence that the tendency to recognize a critical lure as having been studied earlier was greater when its semantic associates were processed in a Story context than when those same words were placed in the same story but randomly arranged (Scrambled context). The effect of scrambling the words in the story effectively served to reduce overall thematic processing. In turn, this led to decreased false recognition of critical lures.

The results of Experiment 1 focused primarily on the rate of false alarms to critical lures in the Story and Scrambled list context conditions since those two conditions provide the best comparison of any possible influence of text processing on the DRM effect. To better assess the role of text processing of DRM stimuli in enhancing false memory, all subsequent studies reported here maintained only the story and scrambled text conditions. Additional measures and manipulations were introduced to further investigate the principal question concerning the influence of text processing in enhancing false recognition.

One goal of Experiment 2 was to include a post-test questionnaire to assess participant awareness of text processing during encoding of DRM stimuli and its relationship to the probability of false recognition. When participants in the story context condition in Experiment 1 were presented with the four lists, they were not told that the items in each list conformed to a coherent text structure. It seems likely, however, that some, if not all, of the participants came upon this realization independently either during the presentation of the first list or at some later point. If text processing is driving the enhanced DRM effect relative to the scrambled text, as observed in Experiment 1, then awareness of text structure could facilitate the likelihood of false recognition later, even when the instructions explicitly called for participants to listen to and remember a list of words for a subsequent memory test. Under this assumption, participants who did not become aware that they were processing stories, or who were simply less aware, or perhaps became aware too late during the presentation of the four lists, would be less likely to show later false recognition due to the relative absence of text processing.

An absence of text processing was expected in the Scrambled context condition and presumably explains the lower rates of false recognition observed in Experiment 1. A questionnaire designed to tap the presence or absence of awareness during the initial encoding of the lists in both conditions will permit an assessment of the hypothesis that text processing is enhancing the DRM effect. In addition to measuring participant sensitivity to the text structure of the lists, a questionnaire can also assess whether participants identified a specific theme relating to each of the lists. It is possible that although some people may not become aware of the text structure of the lists, they may nonetheless become consciously aware of the thematic nature of the list. Additional questions can also probe the onset of awareness and/or theme identification during initial encoding of the lists and whether or not either or both of these were employed by participants during the recognition test.

Experiment 2 served not only to replicate the results found in Experiment 1 but also to provide a further test of the text effect in false recognition by manipulating the story and scrambled conditions within-participants as opposed to between-participants as was done in the first experiment. In conjunction with the questionnaire, this within-subject manipulation will provide a more sensitive measure of awareness and use of the textual elements of the story lists.

As in Experiment 1, each participant was presented with a series of lists with the same instructions to remember the words for a later test of some of the items. This time, however, all participants were presented with lists in both of the two formats: story and scrambled. Two lists of each type were combined into one presentation of four lists for each participant with order and set counterbalanced as in Experiment 1. Following the filler task, participants were presented with a recognition test again with two orders for counterbalancing purposes. A questionnaire was then administered to assess the degree of awareness or sensitivity to the text structure of each list and possible identification of a theme associated with each list (See Appendix C). Analyses will again focus on the number of critical items recognized as old versus new relative to a set of critical items from lists that were not studied. Questionnaire data were coded to measure sensitivity to text and theme identification and their relationship to false recognition of critical items. It was predicted that: A) participants would be more likely to falsely identify an item as "old" if that item was a critical associate of a list presented in a Story context than when an item was a critical associate of a list presented in a Scrambled context; B) awareness of the text structure for lists in the story format would be found for those lists presented within a text structure which would, in turn, be associated with a greater tendency toward overall false recognition; and C) participants who report awareness of the text structure of the lists would demonstrate more false memories than people who failed to report awareness or reported it later in the presentation session. Analogous differences between participants were predicted as a function of theme identification, independent of text awareness.

Method

Participants

One hundred Introductory Psychology students at the University of Winnipeg voluntarily participated in this experiment as part of a course requirement. As in Experiment 1, the only requirement for participation was that participants be fluent in English.

Design and Materials

The same two sets of four stories from Experiment 1 were used again for this study. New recordings of each story were made, however, so that the Story and Scrambled versions were mixed within the same presentation. For each of the eight DRM stories, there were two versions: Story and Scrambled, with different orders and versions of each story appearing in different conditions for counterbalancing purposes. In total, there were eight different conditions and each participant heard four lists of words, two story and two scrambled texts. The questionnaire consisted of several statements that participants indicated their agreement with on a scale from 1 to 5 (see Appendix C). These statements served to probe the degree to which participants recognized a theme in the lists of words that were presented to them earlier and whether or not they tended to recognize that the words in the lists they heard resembled a story.

Procedure

As in Experiment 1, participants were run in small groups in a small classroom. They were provided with the standard list instructions used in the first experiment, and given the expectation of a test of memory later on after the presentation of the four lists. Following this presentation, participants again worked on a 3-minute filler task involving mathematical questions, followed by instructions for the recognition test, as described in the first study. Following the recognition test, participants were given the questionnaire (Appendix B). To complete the questionnaire, participants were asked to think back to each of the four lists they had heard earlier in the experiment and provide ratings for each set of statements separately for each list they could remember.

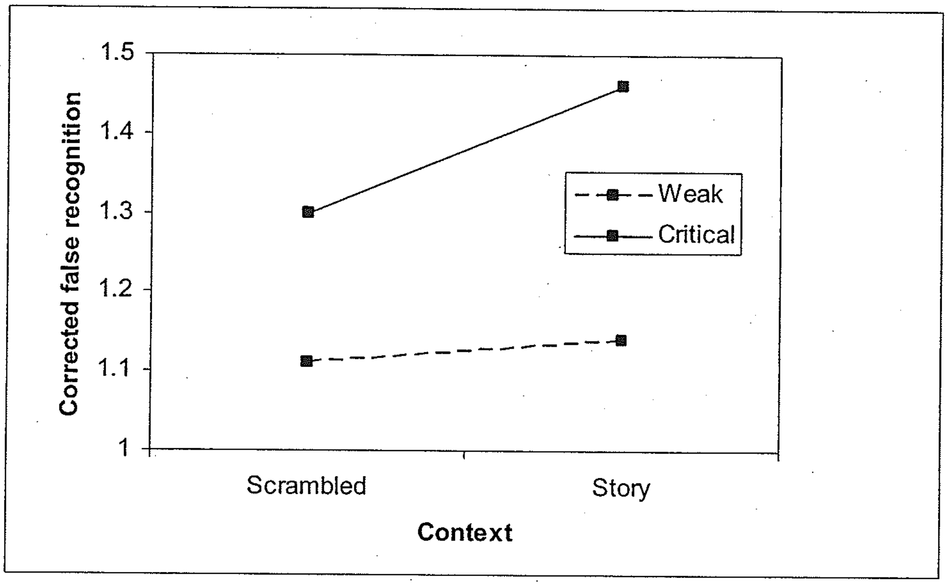
Results and Discussion

Table 3 reports mean recognition ratings and corrected recognition scores for lures and list items as a function of Context. A 2 x 2 analysis of variance was conducted with two within-subject variables: lure type (critical versus weak) and Context (story versus scrambled). The predicted difference between Story and Scrambled conditions was in the right direction, but not significantly so, F (1, 99) = 1.673, MSE = .810, p = .199. Although the interaction between lure type and Context was not significant, F (1, 99) = 1.192, MSE = .354, p = .277, the corrected false recognition rates for critical lures demonstrated the predicted Context effect, whereas the weak lures did not (see Figure 4). When critical lures were examined alone, the effect of their list associates being presented in either the Story or Scrambled contexts approached significance, F (1, 99) = 3.136, MSE = .383, p < .08, consistent with the predicted direction of a story context on false recognition relative to a scrambled one. Also, as in Experiment 1, false recognition of critical lures was significantly greater than it was for weak lures, F (1, 99) = 10.942, MSE = .484, p < .001.

Table 3. Mean recognition ratings and corrected recognition for primed and unprimed lures and list items as a function of Context in Experiment 2.

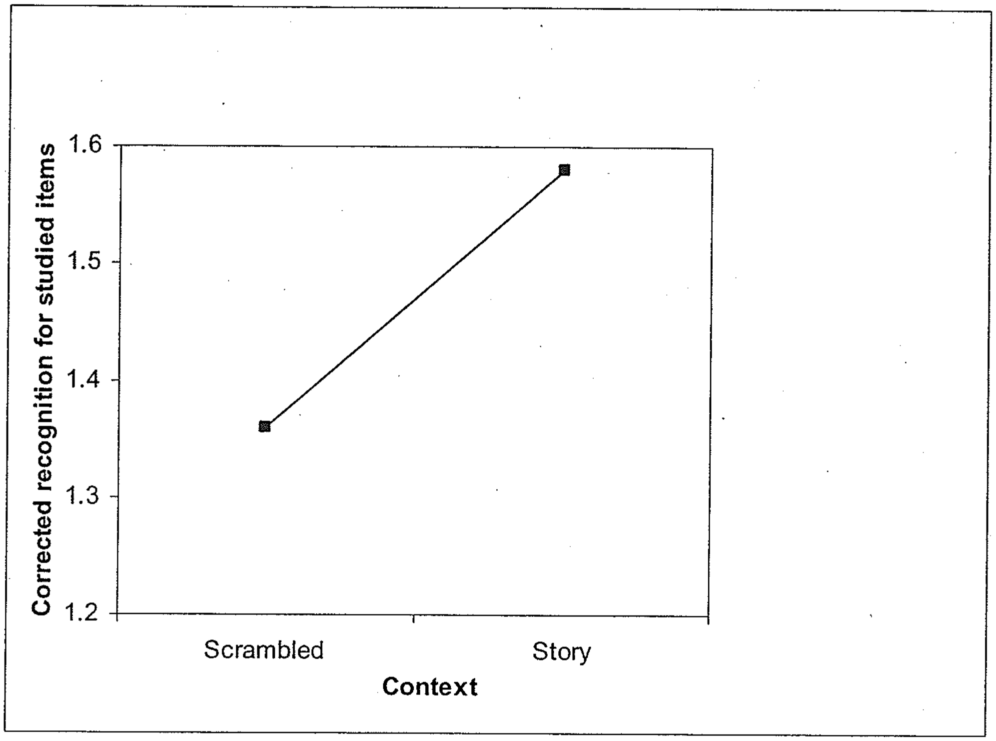
Item Type	Scrambled		Story
		Critical Lures	
Primed	3.40		3.55
Unprimed		2.09	
Corrected False Recognition	***1.30***		***1.46***
		Weak Lures	
Primed	2.92		2.94
Unprimed		1.80	
Corrected False Recognition	***1.12***		***1.14***
		List Items	
Studied	3.30		3.51
Nonstudied		1.93	
Corrected Recognition	*1.36*		*1.58*

Figure 4. Corrected false recognition of critical and weak lures as a function of Context in Experiment 2.



Corrected recognition ratings for list items were higher for Story than Scrambled Context, $F(1, 99) = 17.349$, $MSE = .133$, $p < .001$ (see Figure 5). Again, when ratings for studied items were included as a co-variate for false recognition of critical lures only, the near significant effect of Context reported above disappeared, $F(1, 96) = .006$, $MSE = .325$, $p = .940$. This finding is especially noteworthy, as will be discussed further later, because it is evidence that the false recognition of critical lures correlated strongly with the ability to remember actual items studied as old on the recognition memory test.

Figure 5. Corrected recognition of studied items as a function of Context in Experiment 2.



Analysis of the ratings from the questionnaire focused on ratings for statements two to seven, all of which concerned participant awareness during presentation of items. Ratings for the negatively worded statements 4 and 7 were reversed. Reliability measures of ratings were taken for the six items for each condition, producing an average Cronbach's Alpha of .855. Given the robust consistency across items, an aggregate measure of theme awareness was computed by averaging the six ratings for items 2 to 7. As predicted, ratings indicating theme identification and story awareness during presentation were higher for the Story context ($M = 4.51$) than for the Scrambled context ($M = 3.66$). This main effect of Context on ratings was highly significant, F $(1, 99) = 65.352$, $MSE = .548$, $p < .001$.

When ratings of theme awareness were included as a co-variate, the Story versus Scrambled context effect on false recognition of critical lures became insignificant, $F(1, 97) = .451$, $MSE =$ $.390$, $p = .504$, as did the Context effect for studied items, $F(1, 97) = 2.921$, $MSE = .129$, $p =$

.091. These results are consistent with the prediction that awareness of theme and story would be associated with the tendency toward false recognition. Given that correct recognition was strongly associated with false recognition, it is not surprising that higher ratings on the questionnaire were also associated with higher levels of true recognition.

EXPERIMENT 3

It was predicted that participants in Experiment 2 would report different degrees of awareness of the contextual nature of the lists in the story format, and this may determine greater or lesser tendencies toward false recognition of critical associates for those lists. The questionnaire results also provided some indication as to the point at which such awareness occurred in the initial processing of the lists. Subjective and individual differences in the level and timing of sensitivity to the text and/or theme identification are to be expected, since the instructions orient participants to list processing and do not explicitly mention text. Under the assumption that text processing enhances the DRM effect, orienting instructions that directly inform participants that they will be presented with a series of short stories should enhance the story effect on false recognition. The third experiment examined this prediction.

The two principal conditions from Experiments 1 and 2 (Story versus Scrambled context) were repeated here between-participants. The procedure was identical to that of the first experiment for two conditions; specifically, list instructions were used with Scrambled and Story lists. Two additional groups of participants were also presented with lists in the Story or Scrambled formats but with explicit story instructions replacing the list instructions. The resulting four groups were otherwise treated identically. The story instructions informed participants prior to the presentation of the four lists that they would be hearing four short stories, and that their memory for some of the words in each story would later be tested. All participants in the story context condition with story instructions would therefore begin with advanced knowledge of the textual nature of the to-be-remembered material.

Although the predicted effect of Story (versus List) instructions was clearest in the Story context condition, Experiment 3 also compared list and story instructions for the Scrambled condition. The story instructions for the scrambled format were worded so that the participants in that condition were prepared for the scrambled text. That is, participants were told that they would hear a set of four stories in which some words have been re-arranged in a random order. As shown in Experiment 1, presenting DRM stimuli in scrambled text with list instructions led to lower levels of false recognition than story text, presumably due to the absence of a semantic context. Given story instructions, however, participants may be able to use that knowledge to extract greater meaning and coherence from the list, raising the processing of the material somewhat closer to the text processing that is presumed to take place in the Story context condition. This should result in levels of false recognition somewhere in between that observed under both Story and Scrambled context conditions with list instructions. Such a result, if obtained, would provide further evidence for the role of text processing even in a situation where semanticity, due to disruption of syntax, was absent but nonetheless implied.

If story instructions enhance text processing, and text processing enhances the DRM effect, this instructional manipulation should produce levels of false recognition higher than that observed for text processing with list instructions. The prediction regarding the effects of story instructions is subject to one possible qualification. Explicitly informing participants prior to the presentation of the lists that they will be hearing stories could cause some participants to engage in greater semantic processing of the individual words. But deeper processing of the words in each of the stories may provide participants with more effective source monitoring during the recognition test, and therefore produce a higher number of correct rejections than false alarms. Dewhurst et al. (2007) offered a similar explanation for their failure to obtain higher levels of false recognition following the presentation of stories with embedded DRM stimuli in their older participants. Yet, given the methodological differences between that study and the present set of studies, the question of whether story instructions would further enhance false recognition of critical lures remained open. Experiment 3 therefore provided the additional empirical evidence necessary for making more definitive conclusions.

Finally, the questionnaire was employed again following the final recognition test for participants in all conditions. This allowed for a manipulation check in the story context condition with story instructions, as well as the potential effect of story instructions on theme identification and its relationship to false recognition. It was also useful to gauge the effect that story instructions in the scrambled format condition had on participant awareness of text and theme identification and the relationship of awareness to false recognition.

Method

Participants

One hundred and sixteen Introduction to Psychology students from the University of Manitoba voluntarily participated in this experiment in order to fulfill a course requirement.

Design and Materials

The stimuli from Experiment 1 were again used in Experiment 3, except that only the story and scrambled media files were used, so that there were two (story and scrambled) versions of each story in both sets of stories with two orders of the four stories in each of these versions, resulting in eight different conditions. That is, the eight original stories were divided into two sets for each of the these two conditions for counterbalancing purposes, and each provided a set of nonstudied items in the recognition test, as was done previously. The same recognition test and filler task from the first experiment was used again for this study, as was the questionnaire from Experiment 2.

Procedure

Participants were again run in small groups of ten to twelve and each group was randomly assigned to one of the eight conditions. The same procedure employed earlier was used in Experiment 3 with the exception of the orienting instructions provided to participants at the start of the study. For each of the eight conditions, one half of the groups were given List instructions, as was done in the previous two experiments and the other half of the groups were given story or Thematic instructions. In sum then, this resulted in a total of sixteen separate conditions. Following either of the two initial instructions, participants underwent the identical procedure as in Experiments 1 and 2, except with the addition of the questionnaire as described in study 2. Each experimental session lasted approximately 30-40 minutes.

Results and Discussion

A 2 x 2 x 2 analysis of variance was performed on the data collected from this study with one within-subjects variable: lure type (critical and weak), and two between-subjects variables: Context (Scrambled versus Story conditions) and orienting Instructions (List versus Thematic). Table 4 reports the mean ratings and corrected recognition for both critical and weak lures and list items as a function of Context and Instructions. For corrected false recognition ratings, the main effect of Context (Scrambled versus Story) on false alarms to lures was highly significant, $F(1, 112) = 10.297$, $MSE = 1.308$, $p < .01$. This replicates the finding from Experiment 1 that showed a robust effect of Context on false recognition.

Table 4. Mean ratings and corrected recognition for primed and unprimed lures and list items as a function of Context and Instructions in Experiment 3.

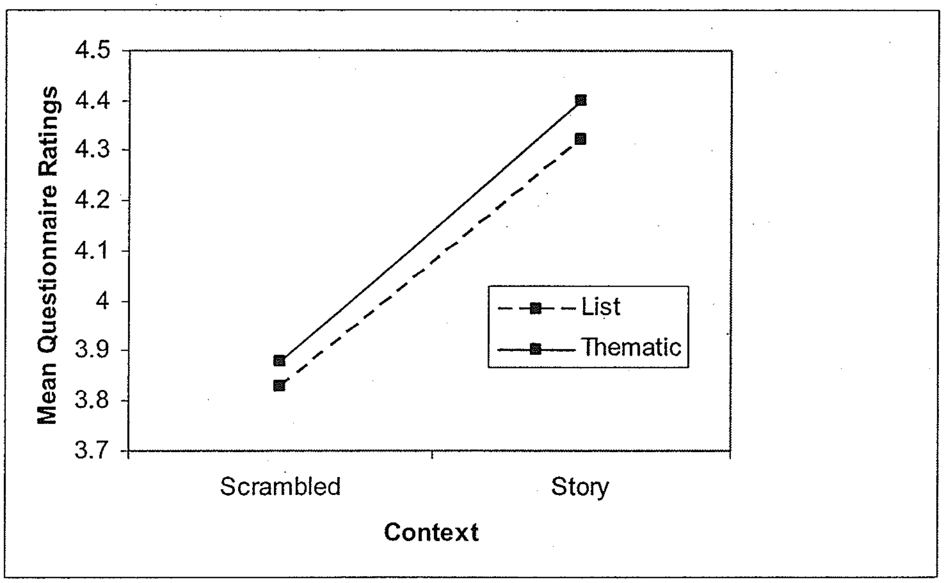
	Instructions			
	List		Thematic	
Item Type	Scrambled	Story	Scrambled	Story
	Critical Lures			
Primed	3.20	3.38	3.32	3.46
Unprimed	2.33	2.06	2.11	1.87
Corrected False Recognition	***0.82***	***1.31***	***1.21***	***1.59***
	Weak Lures			
Primed	2.70	2.90	2.85	3.17
Unprimed	2.19	1.96	2.05	1.73
Corrected False Recognition	***0.51***	***0.95***	***0.81***	***1.44***
	List Items			
Studied	3.10	3.28	3.24	3.60
Nonstudied	2.25	2.01	2.11	1.80
Corrected Recognition	*0.85*	*1.27*	*1.13*	*1.80*

Also found was a significant main effect of Instructions with primed lures more likely than unprimed lures to be identified as old under Thematic as opposed to List instructions, F (1, 112) = 5.818, MSE = 1.308, p < .05. There was no interaction between Context (Scrambled versus Story) and Instructions (List versus Thematic), F (1, 112) = .020, MSE = 1.308, p = .888. Thus, Thematic instructions raised rates of false recognition higher than did the standard List instructions in all conditions. In addition, as found in the first two studies, there was again a main effect of lure type with critical lures significantly more often identified as old than were weak lures, F (1, 112) = 14.481, MSE = .379, p < .001.

As with lures, similar significant main effects were found for studied items. Corrected recognition scores for presented items showed that studied items were rated old more often than nonstudied items in the Story (M = 1.56) than in the Scrambled condition (M = .99), F (1, 115) = 17.877, MSE = .479, p < .001. Whether or not participants received List (M = 1.06)or Thematic (M = 1.50) orienting instructions produced a similar main effect on studied versus nonstudied items, F (1, 115) = 9.858, MSE = .479, p < .01. Also, as with false recognition, there was no significant interaction between Context and Instructions, F (1, 115) = 1.013, MSE = .479, p = .316. The significant effects of Instructions and Context on false recognition, again, were eliminated when studied items were analyzed as a covariate: for Context, F (1, 111) = .011, MSE = .525, p = .915, and for Instructions, F (1, 111) = .001, MSE = .525, p = .970.

Ratings from the post-recognition test questionnaire again produced high reliabilities across stories, with an average Cronbach's Alpha of .856. Overall, higher ratings were obtained in the Story condition than in the Scrambled condition, F (1, 115) = 15.540, MSE = .477, p < .001 but there was no significant effect of instructions, F (1, 115) = .240, MSE = .477, p = .625 (see Figure 6). Thus, as in Experiment 2, participants in the Story condition (M = 4.37) were more likely than participants in the Scrambled condition ($M = 3.86$) to agree with statements indicating an identification of a theme during the presentation of the stories, as well as an awareness of their text structure.

Figure 6. Questionnaire ratings indicating theme identification and story awareness as a function of Context and Instructions in Experiment 3.



Also, as reported in the previous experiment, when ratings were included as a co-variate the significant Story versus Scrambled Context effect on false recognition was reduced, $F(1, 111) = 3.343$, $MSE = 1.160$, $p = .070$, as it was again also for correct recognition, $F(1, 111) = 6.668$, $MSE = .389$, $p = .011$. Thus, this finding provides additional evidence that the tendency to falsely remember nonpresented lures and correctly remember studied items was positively associated with both early identification of a theme and awareness of a story when participants were actually presented with lists of words that made up narrative text structures.

EXPERIMENT 4

Whereas the goal of the first three experiments was to investigate the principal hypothesis that text processing enhances the DRM effect relative to non-text conditions, the purpose of

Experiment 4 was to show a reduction in false recognition by intentionally disrupting text processing of textual material. Specifically, the aim of Experiment 4 was to reduce the text interpretation of the stories, not by scrambling the words within the story, but by manipulating levels of text-processing. Since it has already been shown that violation of normal syntactical arrangement of the words in a story disrupts text processing, there should be other ways in which to inhibit the processing of text, compared to a condition in which thematic story processing is enhanced.

Experiment 4 introduced a depth of processing manipulation involving only the story conditions with modified instructions to produce different levels of text processing. Two groups of participants were presented with four short stories for which memory of individual words from the stories were later tested as in previous studies. One group was instructed to rate the overall cohesiveness or coherence of each story, considering such factors as ease of understanding the story and its narrative structure, and overall comprehensibility. The second group rated the overall pronounceability of the words in each story, considering such factors as typical word length, and the ease with which individual words in the story were judged to be spoken or clearly enunciated.

Since the study instructions were intentional as in all previous studies, participants were informed that factors such as cohesiveness or pronounceability, depending on their condition, are characteristics that have been related to memory for text, which is why they were being asked to provide their corresponding ratings. Individuals were asked to indicate their ratings following the presentation of each story. Following the filler task and recognition test, the questionnaire was again used as a manipulation check to see if sensitivity to text and potential theme identification varied as a function of level of text-processing.

It was predicted that participants who read the stories for cohesiveness would be more likely to later falsely identify critical items as "old" compared to those who read the stories for pronounceability. Although this prediction was subject to the same qualification discussed for Experiment 3, it was expected that the results of this between-subject manipulation would mirror those of Experiment 1, which compared false recognition following presentation of Story versus Scrambled contexts. If the DRM story effect reported in that experiment is being driven by textual processes, then cohesiveness rating instructions should lead to higher rates of false recognition of strongly associated, but nonpresented words than word-level instructions. The outcome of this study should ultimately provide some evidence for that hypothesis or the possibility that there is an "upper limit" (as found in Experiment 1 perhaps) to the influence of story context on the DRM effect.

Method

Participants

A total of 84 students from a pool of Introductory Psychology classes at the University of Winnipeg volunteered to participate in this study as part of a course requirement. The only restriction was that English had to be the participant's first language.

Design and Materials

The same stimuli used in Experiments 1 and 3 were used again for this experiment, except that only the stories were employed in this study. This produced two story conditions, one from each set of stories, each with two different orders, giving a total of four story conditions.

Procedure

The same general procedure that was used in all three previous studies applied to Experiment 4. Participants were run in groups of 10-12 in a small classroom. All participants were told that this was a study of memory and that they were going to hear four lists of words, each of which made up a short story or narrative (thematic instructions from Experiment 3). Each of the participants had booklets, half of which included 'comprehensibility' instructions, and half of which included 'pronounceability' instructions on the first page. Specifically, participants in the comprehensibility conditions read the following instructions:

One factor related to memory for text is comprehensibility. Comprehensibility refers to how easy text is to understand. After each story, use the following scale to make a rating of the overall comprehensibility of the story you just heard. While listening to each story, pay careful attention to how easily you are able to understand it. If the story has a clear storyline or narrative, is coherent and cohesive, and is easy to comprehend, then give it a high rating of 6 or 7. If the story lacks these qualities and it is difficult to make sense of the meaning, then give it a low rating of 1 or 2. Use intermediate values for stories that are moderately difficult or easy to understand.

Similarly, participants in the pronounceability conditions were given a different set of orienting instructions:

One factor related to memory for text is pronounceability. Pronounceability refers to how easily the individual words in the text are spoken, articulated, or pronounced. While you are listening to the four stories, pay careful attention to the individual words you are hearing in the story. If the words in the story are generally words you consider to be easily spoken words then give an overall high rating of 6 or 7. If the words in the story lack this quality and are difficult to pronounce, then given the story a low rating of 1 or 2. Use intermediate values for stories that have words that are moderately difficult or easy to pronounce.

Once participants indicated they had read and understood the instructions, presentation of the stories followed. Participants rated each story during a short pause between each of the stories. Following the presentation of the final story, the procedure continued as described above in Experiments 2 and 3, as participants completed the distractor task, recognition memory test, and finally, the questionnaire.

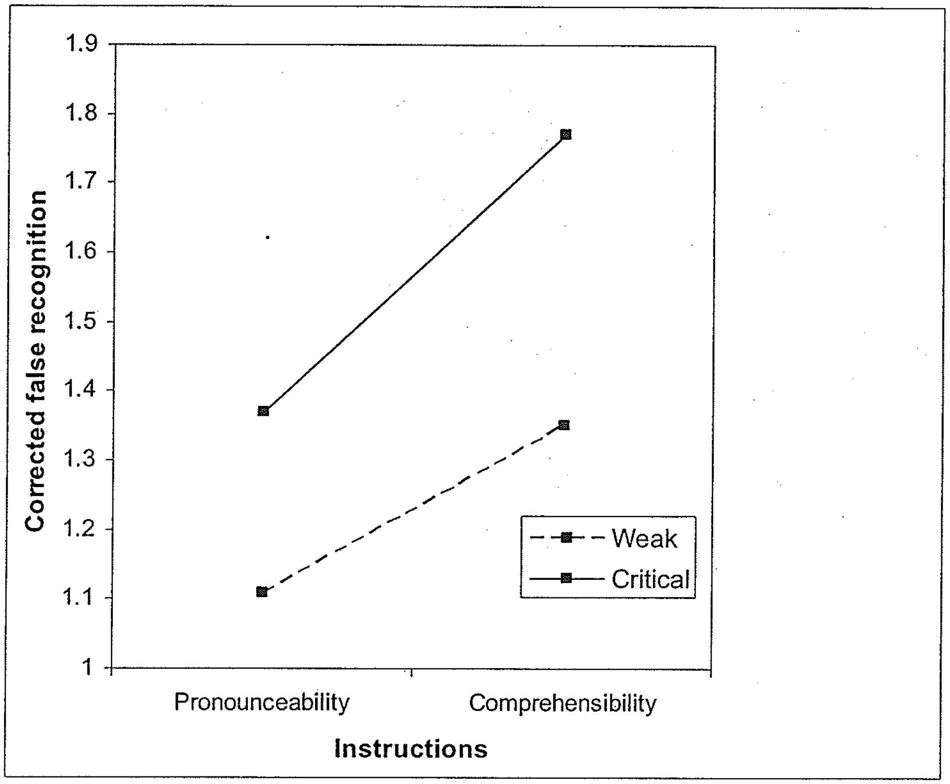
Results and Discussion

Table 5 shows mean recognition ratings and corrected recognition scores for lures and non-lures as a function of story processing instructions. Data was analyzed by a 2 x 2 analysis of variance with one within-subjects variable: lure type (weak versus critical) and one between-subjects variable: story processing instructions (comprehensibility vs. pronounceability). Averaged across both types of lures, the effect of Instructions on false recognition was marginally significant, $F(1, 81) = 3.173$, $MSE = 1.313$, $p = .079$. When analyzed separately, critical lures showed a marginally significant effect of instructions in the predicted direction, $F(1, 81) = 3.902$, $MSE = .842$, $p = .052$, whereas the predicted effect for weak lures did not approach significance by a two-tailed test, $F(1, 81) = 1.866$, $MSE = .779$, $p = .176$ (see Figure 7), although the difference was in the expected direction. Also, as reported in all previous experiments, there was once again a strong main effect of lure type, with critical lures significantly more likely to be falsely recognized as 'old' than weak lures, $F(1, 81) = 15.593$, $MSE = .303$, $p < .001$. The interaction between lure type and processing instructions was not significant, $F(1, 81) = .902$, $MSE = .303$, $p = .345$.

Table 5. Mean recognition ratings and corrected recognition for lures and list items as a function of story processing instructions in Experiment 4.

	Instructions	
Item Type	Pronounceability	Comprehensibility
	Critical Lures	
Primed	3.47	3.52
Unprimed	2.09	1.75
Corrected False Recognition	***1.38***	***1.77***
	Weak Lures	
Primed	2.94	2.95
Unprimed	1.82	1.59
Corrected False Recognition	***1.12***	***1.35***
	List Items	
Studied	3.43	3.53
Nonstudied	1.87	1.77
Corrected Recognition	*1.55*	*1.76*

Figure 7. Corrected false recognition of critical and weak lures as a function of story processing instructions in Experiment 4.



The weak effect of instructions on false recognition might be explained by the finding that studied items did not show a significant influence of the processing manipulation, with a corrected mean true recognition of 1.55 under the pronounceability instructions and 1.75 under the comprehensibility instructions, which did not differ significantly, F (1, 82) = 1.567, MSE = .535, p = .214. When controlling for studied items, the effect of instructions on the tendency to falsely recognize the critical lures lost significance, F (2, 81) = 2.226, MSE = .296, p = .140.

Questionnaire ratings produced a somewhat smaller average Cronbach's Alpha of .75 than in previous experiments. There was a main effect of instructions on the ratings, although it was in the opposite direction to that predicted. Specifically, pronounceability instructions produced higher overall ratings (M = 4.66) than comprehensibility instructions (M = 4.27), F (1, 83) =

16.983, $MSE = .197, p < .001$. Because of this reversal, including the ratings as a covariate actually *increased* the significance of the effect of processing Instructions on corrected false recognition, $F(1, 80) = 11.473$, $MSE = 1.116, p = .001$, and on true recognition, $F(1, 80) = 8.262$, $MSE = .453, p = .005$.

In sum, the results of Experiment 4 were generally in the direction predicted with one notable exception. As predicted, comprehensibility instructions led to higher levels of false recognition than pronounceability instructions. Also, consistent with the findings of the previous experiments, the false memory effect was especially pronounced for the critical lures compared to the weak lures. Unlike the previous two studies, however, ratings indicating theme identification and story awareness did not differ across conditions leading to differential false recognition, and curiously, were significantly higher in the shallow condition compared to the deep processing condition. The results of this study and the previous three experiments will be discussed further in the following section.

GENERAL DISCUSSION

In three experiments, false recognition of nonpresented lures was reduced for participants when lures were associated with words presented in a random context, as compared with participants who were presented with the same lures presented in a narrative context. This result occurred whether this variable was manipulated between-subjects (Experiments 1 and 3) or within-subjects (Experiment 2). This story effect clearly shows the important role for semantic processing of DRM-list items in the false recognition effect. The story effect as reported in this series of studies does not permit a direct comparison between a DRM story context and the standard list presentation as the differences between these two types of presentation are too large to make for any meaningful comparisons. Indeed, this was the basis of the concern discussed earlier with the results reported by Dewhurst et al. (2007) that were in fact based on such a direct comparison. The relative superiority of story context over the traditional list format was not the central question, however; that being the role of semantic processes underlying the false memory effect as elicited by a narrative, story context. To that end, eight standard DRM lists were used to construct eight DRM stories, out of which scrambled versions were created as the most reasonable and methodologically-sound means of testing for the potential benefits of embedding DRM list items in a semantically-coherent text structure.

Thus, when appropriately compared to a scrambled text, a story context, contrary to the conclusion drawn by Dewhurst et al. (2007), can reliably enhance the DRM illusion in adult participants and presumably, by implication, in older children as well. That the format in which DRM items were studied did not reliably influence levels of false recognition in older children in their study is consistent with the finding from Experiment 1 here where no difference in false recognition of lures was found when their list associates were presented in a list condition or a story condition. Notwithstanding the difficulties of comparing a list with a story condition discussed previously, an additional complication that arises out of doing so in Experiment 1 is the difference in presentation rate of studied items. In order to match the timing of the total presentation duration of the list associates in the list context with those in the scrambled and story contexts, pauses were inserted in place of the words (see Appendix A) that made up the stories, or their scrambled versions, significantly slowing the presentation rate of the studied items in the list condition. This gave participants in that condition extra processing time of the list items (minus the additional filler items present in the other two conditions) during encoding and may have elevated both true and false recognition to levels one might not observe otherwise. This argument is certainly consistent with past findings that have shown that when people are given more time to process list items in a DRM task, false recognition is higher (Arndt & Hirshman, 1998; Seamon et al., 1998). It is possible that under these conditions additional processing of the list words enhances the activation of the related lure. Taken together, these confounds highlight the difficulties inherent in comparing a standard list condition with a story condition and reinforce the rationale for the approach taken here to concentrate only on the story and scrambled conditions, where these potential sources of extra variability have been removed.

It is only with the story versus scrambled conditions, therefore, that the DRM story effect, as it is being called here, is most clearly defined and demonstrated. For example, in the *sleep* story, as shown in Appendix A, an additional 43 words (the number of additional items were approximately the same in all eight stories) were added to the list items that were presented under the list context. When these items were randomly inserted among the DRM list associates (the scrambled context) there was a significant drop in false recognition (see the left half of Figure 3). However, when these same words were arranged such that they formed a coherent text structure, a narrative (the story context), false recognition of the non-presented lures compared to the distractor lures, rose sharply again even slightly above, albeit not significantly due possibly to the reasons stated above, what was observed under the standard list format (see right half of Figure 3). This basic story effect was also replicated in two additional experiments (2 and 3).

An additional benefit of the present methodology is that any effect of the additional individual words comprising the story is tightly controlled. The exact same additional words occurred in the scrambled and story contexts, meaning that any difference in false recognition could not be attributed, for example, to additional priming of lures by individual presented words.

Primacy of the Critical Lure

When participants were presented with the DRM lists in both the story and scrambled formats in Experiment 2, the story effect on false recognition occurred only for critical lures, but not weak lures. Rather than diminishing the story effect, this finding highlights the general superiority of the critical lures over the weakly-related lures in producing the DRM illusion. Indeed, a consistent finding with respect to Lure type across all four experiments was that the standard critical lures were reliably stronger than the weak lures in eliciting 'old' judgments on the recognition tests. The fifteen list associates in each of the eight DRM lists generally constitute the first 15 items in the Russell and Jenkins (1954) word association norms. Following the procedure of Dewhurst et al. (2007), one of these 15 items was removed from the list and used as a second critical (weak) lure in the recognition tests. Specifically, for each of the eight lists, and their associated critical items: *cold*, *doctor*, *lion*, *sleep*, *smell*, *thief*, *music*, and *fruit*, the corresponding weak lures chosen were: *heat*, *patient*, *roar*, *nap*, *aroma*, *gun*, *melody*, and *apple*, respectively. The average serial position of each of these list associates in the standard lists is 8. Using items such as these from the DRM lists as secondary critical lures produced a weaker false memory effect in the current set of studies.

There are several possible reasons for the additional lures producing a weaker false memory effect than the standard lures. The most likely explanation is that the additional weak lures were simply less activated than the critical lures by the studied items. Much research has demonstrated that forward associative strength (FAS) is not as powerful a predictor of false recall or recognition as is backward associative strength (BAS) (Deese, 1959; Roediger et al., 2001). FAS is the likelihood that the critical lure elicits the items in the list as responses in a free association task, whereas BAS is the likelihood that the list items elicit the lure. It appears likely that the additional weak lures have lower BAS than the critical lures (i.e., the weak lures occur less often as associates to other list items). Further research with more lists and a greater sampling of alternative lures would resolve this question, as well as add to the research on the role of BAS in the DRM false memory effect.

Another possibility is that the weak effect of the weak lures is related to connectivity, or inter-item associative strength (IAS), which Deese (1959) defined as the "average relative frequency with which all items in a list tend to elicit all other items in the same list as free associates" (p. 235). As a list variable, however, connectivity has been found, along with FAS, not to correlate well with false recognition (r = .03, Roediger et al., 2001). It appears that lures gain their status because they are associated with each of the items in their respective lists, whereas individual items in each list are not necessarily associated with each of the other items in that same list. Again the use of alternative lures, as in the present study, provides an additional way to further test the role of these competing factors as it would be possible to select lures with specific properties (e.g., high or low BAS) from lists of varying connectivity.

Promoting a list associate to the level of a critical lure may also have produced another disadvantage in measuring overall false recognition. Robinson and Roediger (1997, Experiment 2) found that levels of false recall rose linearly as the number of associated items studied increased from 3, 6, 9, 12, and 15 items, even when unrelated filler items were added to equate the total length of each list at 15 items. They argued that this was evidence that the total backward associative strength, as opposed to mean BAS, of an individual list of associates is crucial in eliciting activation of that lists' critical item. Lowering the strength of each of the lists (stories) by removing one of its associates, especially if that associate was originally positioned relatively high in the list, may have weakened the reported results of all four of the experiments here.

Although this is a possibility, it is probably more likely that the removal of only one list item had a neglible effect on overall false recognition rates. Only future research can test for the effects of varying the number of list associates that are embedded in a story context. What does seem clear, for present purposes however, is that using items other than the standard critical lures as lures, is likely to diminish the false memory effect.

Story versus Scrambled Contexts

In Experiment 2, participants were presented with both story and scrambled contexts. Manipulating the context within-participants did not eliminate the story effect, but did weaken it somewhat. One drawback of this procedure was that each individual participant provided fewer responses for analysis, with only two critical items from each of the two stories, and the two scrambled stories. Another possible explanation for the reduced story effect is that exposure during encoding to both types of contexts, particularly the story contexts, encouraged semantic processing of the scrambled contexts and/or reconstructive processes of the scrambled stories during retrieval.

This may relate to the finding that scrambling stories can disrupt a readers' ability to comprehend those stories especially under restricted reading conditions (e.g., Kintsch, Mandel, & Kozminsky, 1977), as indicated by their summaries. Kintsch et al. (1977) compared the ratings of written summaries of a well-structured story with a poorly-structured story, read and written by a separate group of participants under free or restricted reading conditions. Half of the time the stories were presented normally and half of the time the stories were scrambled, the latter of which was achieved by rearranging the order of the paragraphs within the story. Only under free reading conditions of a well-structured story, regardless of whether it was presented normally or in scrambled form, did readers produce summaries that the raters were not able to discriminate, suggesting that the quality of both types of summaries were essentially equivalent. Kintsch et al. (1977) argued that for a scrambled, well-structured story (defined by the strength of the connections between the causal events in the story), a reader can use his or her schemata for stories and thereby reorganize the scrambled story so that ultimately the macrostructure of the text is no different than it would be for the same story presented normally.

A similar process may be at work when people read both normal and scrambled stories in the context of the present investigation, particularly Experiment 2. The stories used here were no doubt qualitatively different from those used by Kintsch et al. (1977), in that they were considerably shorter and simpler, to the point that they may not even qualify as stories, at least according to the theory of story comprehension as outlined by Kintsch et al. (1977). Also, of course, since these stories were relatively short, the scrambled versions constituted a reordering of the words within the story rather than a reordering of larger textual units (sentences, paragraphs). Regardless of these differences, the point here is that upon being presented with text, both in its proper and scrambled order, participants may have been able to use their story schemata as an aid to reorganize elements of the scrambled stories. In turn, this reorganization may have promoted semantic processing, at least to some degree, thereby leading to levels of false recognition similar to that which was found when processing nonscrambled stories.

Of course, it must be remembered that the story effect in Experiment 2, although weakened, was not eliminated, particularly for the standard, critical lures. Thus, the explanation proposed above, based on the account given by Kintsch et al. (1977) of their results, may still be relevant, but for the opposite reasons. One could assume that the text structures of the stories used in the present investigation lacked sufficient length and complexity that they did not lend themselves easily to an organization of their macrostructures. This might mean that the stories constructed out of the originals from Dewhurst et al. (2007) could be classified along the same lines as the poorly-structured stories, so defined, in Kintsch et al. (1977).

At the same time, however, the short passages presented as a coherent narrative were sufficiently story-like to promote the additional semantic processing underlying the story effect, as already proposed earlier. A secondary mechanism may have involved the use of a story schema on the part of the listener to aid that semantic processing, and therefore, comprehension. Although minimally sufficient in the story context condition, the absence of a strong macrostructure in the scrambled context condition may have inhibited the activation of a story schema, in addition to a reduction in semantic processing, thereby preventing a reorganization of its textual elements, and therefore, comprehension. This explanation, it should be noted, is perfectly consistent with the finding that participants in the story condition, but not in the scrambled condition, were much more likely to identify a theme when listening to the stories, in addition to indicating an awareness of its narrative text structure.

This account, based on the work of Kintsch et al. (1977), provides an attractive alternative framework for understanding the story versus scrambled context effect on false recognition, and is not necessarily incompatible with the mechanisms already proposed. At the very least, it suggests several possible avenues of future research that could investigate more systematically the effects of manipulating the text structure of the stories (e.g., longer versus shorter; complex versus simpler stories) used to produce the DRM story effect.

Thematic versus List Instructions

In Experiment 3, a deviation from the standard DRM procedure was introduced in that some participants were explicitly informed that they would be hearing a set of short stories, compared to the standard list instructions, used in Experiments 1 and 2, for which their memory for some of the words in the stories (or lists) would later be tested. It was predicted that thematic instructions would encourage the activation of a story schema prior to encoding. If the activation of story schemata occurred in Experiments 1 and 2, as the evidence strongly suggests, this would have had to occur spontaneously on the part of the listener, as participants were given list-based instructions with no expectation of a story. Although it is clear from the findings of the previous experiments that the story effect does not depend on theme-based instructions, it was nonetheless expected that the expectation of a story would enhance semantic processing relative to list-based instructions.

The results of this study generally conformed to the set of predictions regarding the influence of thematic instructions on the story effect. Thematic instructions prepared participants for the presentation of four lists of words that constituted four short stories, whereas the standard List instructions only indicated that participants would be presented with a series of four word lists. As shown in Table 2, in the Story Context condition with Thematic instructions people were more likely to falsely recognize critical lures that were not actually presented than they were when they had only received List instructions. One explanation for the benefit of Thematic instructions, as explained above, is that such instructions may have activated a story schema on the part of the listener prior to the presentation of the stories. In turn, the story schema may have facilitated the extraction of the meaning of the passage, resulting in the overall greater semantic processing presumed to underlie the story effect.

What is particularly interesting regarding the results of Experiment 3, however, is the effect of Thematic instructions in the Scrambled Context condition. As expected, there was a similar benefit of Thematic instructions when the stories were presented as scrambled text, relative to List instructions. Although both true and false recognition rates were higher in the Story Context condition compared to the Scrambled Context condition, replicating the results of Experiments 1 and 2, a similar influence on recognition was observed even with scrambled text, when they were preceded by Thematic, as opposed to, List instructions. Thus, it again seems appropriate to interpret this result, although speculative, in terms of a person's ability to reorganize the text during encoding guided by a story schema, enabling some added degree of extraction of its overall meaning. In turn, this would have engaged the listener in a level of semantic processing sufficient enough perhaps to produce the false memory effect in a situation where it has been found in the other studies reported here to be relatively weak (i.e., the Scrambled Context condition under standard List instructions). Although Thematic instructions may have been minimally sufficient under these conditions, it clearly did not reach the levels of false recognition as observed for actual text. This makes sense when considering that the ratings provided in the questionnaire did not show an effect of instructions, but were higher in the Story than in the Scrambled Context condition.

Blocking Effects

The results of this series of studies are both consistent and inconsistent with some findings of past studies, although any discussion of such comparisons of findings needs to be made cautiously given the unique manner in which DRM stimuli were presented here. For instance, the story effect is consistent with past findings concerning blocking effects (McDermott, 1996; Tussing & Greene, 1997), where greater levels of false recall and recognition have been obtained when list associates are blocked or grouped into themes at study than when they are presented in a mixed fashion. Blocking list associates at study is likely to encourage relational processing through its facilitation of the organization of study items, thereby encouraging associative activation, with this activation potentially summating across the study items. Due to both associative activation and greater use of organizational strategies during study, increases in both false and true remembering would be expected, as was the case in these and the current set of studies.

Of course, in the present case, list associates were blocked, albeit with a number of intervening words, in both the scrambled and story contexts but the standard blocking effect on false and veridical recognition was restricted to the story context, in which thematic consistency was maintained. Thus, the effects of blocking in the story condition may have depended upon the ability of participants to engage in the encoding of a thematic or gist representation of the list associates, as proposed by proponents of fuzzy trace theory (e.g., Brainerd et al., 2001). This could explain why presenting DRM stimuli in a narrative form, as opposed to its scrambled counterpart, led to higher rates of false remembering, even when both types of presentations of the DRM list associates were essentially presented in a blocked, rather than mixed, fashion. In this regard, an activation account of these results might be at a disadvantage presumably because false remembering should be unaffected by its presentation, provided the list associates were presented in blocks, as they were here. An explanation involving associative activation becomes less problematic, however, if one assumes that both blocking and thematic consistency, produced by a narrative presentation of the list associates, had additive effects on encouraging inter-item associations and semantic organization of those list items.

A related question for future research on the blocking effect would be to determine whether interspersing coherent or related text (as in the story condition) would interfere less with the false memory effect than interspersing random words (akin perhaps to the scrambled condition). That is, it may be that the story condition produced less interference with the standard blocked presentation, and hence led to higher levels of false recognition.

Veridical and False Recognition

Although no specific predictions were made with respect to studied items, it is clear from the current set of experiments that correct recognition of actually presented items mirrored the pattern for false recognition of nonpresented items, both showing a clear story versus scrambled context effect. The strong correlation between veridical and false recognition is contrary to what was reported in the multiple regression analyses across lists by Roediger et al. (2001) where the two were found to be significantly negatively correlated ($r = -.52$). That is, lists that elicit good veridical recall tended to produce lower rates of false recognition. Therefore, it seems that lists that produce good memory for studied items are less susceptible to falsely recalling or recognizing the critical lures of the list associates. Better encoding of list items during study may aid the source-monitoring process during retrieval such that the critical item can be better distinguished from the list items since it carries with it less item-specific information than do the studied items.

A contrasting pattern emerged in the present studies when the relationship was examined across subjects. In the current set of experiments false remembering of critical items depended on accurately recognizing studied items. This strongly suggests that people treated the nonpresented items as though they were actually presented. What might account for the apparent discrepancy between the current results and those reported by Roediger et al. (2001)? One explanation could be related to the function of embedding DRM-list items in a story context, which might also help elucidate the underlying mechanisms by which the DRM story effect is produced. This explanation implicates, greater semantic processing of the text presentations or relational processing of the items within the text as a function of placing DRM stimuli within a coherent, narrative text structure. Either mechanism likely led to increased activation of the critical items during encoding, and led to more source-monitoring errors during recognition. Indeed, past research has shown that similar experimental variables, such as increased semantic processing, can produce a positive correlation between veridical and false recall, discussed further in the next section.

Levels-of-Processing

Studies that have manipulated levels-of-processing (e.g., shallow versus deep processing of DRM lists) have typically found that deeper levels of processing has similar effects on both true and false remembering. That is, remembering more studied items under semantic conditions leads to more illusory memories (Toglia et al. 1999). These researchers dubbed this the ‘more is less’ effect because overall, remembering is less accurate because of the higher rate of intrusions on recall. Thapar and McDermott (2001, Experiment 2) found that deep levels of processing (pleasantness ratings) led to greater levels of true and false recognition than did shallow levels of processing (vowel or colour judgments). These results are in line with studies that have looked at the effects of blocking, discussed earlier, where it is likely that blocking results in a deeper level of encoding than a random presentation.

Overall, the results for variables that tend to encourage semantic processing are consistent with the current set of experiments in which lists of semantic associates that converge on a critical, nonpresented item were presented in a semantically-coherent narrative text structure.

Presenting either normal or scrambled versions of a narrative itself is a variable that naturally introduces a levels-of-processing manipulation without any additional, explicit processing instructions. In this way, a story context, can be understood to produce the DRM effect in a manner similar to past studies that have manipulated blocked versus a random presentation and depth of processing, which have produced similar results not only on false memory but also for true memory.

Experiment 4 was unique among the current set of studies in that only the story contexts were presented to participants but with additional encoding instructions designed to induce a "levels-of-text-processing" manipulation, with the goal of extending traditional depth of processing studies with lists comparing shallow and deep processing conditions. The shallow processing condition in Experiment 4, which focused participants' attention on the average pronounceability of the individual words in each of the four stories presented to them, was meant to serve as a condition similar to the Scrambled Context conditions of the first three experiments. The goal was to reduce or discourage semantic processing of the stories, and in turn, the false memory effect, and indeed, this is what was found. The false memory effect was stronger when participants were asked to pay attention to the overall comprehensibility of the narrative texts.

It was surprising, therefore, that the tendency to become aware of a story and identify a theme during encoding was not associated with the higher rate of false recognition in the deep processing condition, given that these were compelling aspects of false recognition in the first three experiments. Of course, a major difference with respect to those studies was that these subjective reports of awareness of story and theme identification were associated with false recognition following the presentation of story contexts when compared to the presentation of scrambled contexts, which, of course, were absent in Experiment 4. What was even more surprising was that higher questionnaire ratings were found in this study for participants in the shallow processing condition, where false recognition was absent.

Another difference between Experiment 4 and the previous studies was that correct recognition, like the questionnaire ratings, did not show an effect of the levels-of-text processing manipulation. Correct recognition of studied items, however, is typically better following semantic processing, but in this case, the studied items were presented as part of larger context, a narrative, and not simply as a list of words. Although not significant, deep processing instructions did favour correct recognition judgments and controlling statistically for correct recognition did weaken the effect of instructions on false recognition, consistent with the earlier studies.

In the first three experiments the positive association between true and false memory suggests that lure items were treated like studied items, implicating their conscious activation during encoding. During the subsequent recognition tests, this would have made the lures more difficult to distinguish from studied items. Under the activation/monitoring account, however, activation of lure items can occur either consciously or unconsciously. It is possible that the level of deep processing (comprehensibility) engaged in by participants may have caused the mental generation of the lure items during study to occur unconsciously, similar to Underwood's (1965) concept of an implicit associative response (IAR).

Perhaps it was the case that when given the comprehensibility instructions, the listener's task became overly semantic such that the ability to attend to the general factors associated with text comprehension, as instructed, came at the expense of actually comprehending the specific meaning elements of the narrative text to which they were attending. If so, this might explain why these participants did not report remembering a specific theme or an awareness of a story as strongly in the deep processing condition. At the same time, hearing multiple converging associates in the story could have been sufficient to cause unconscious associative activation of the lure item and later its apparent familiarity upon seeing the item during the recognition task.

So, while it might have been the case that the deep processing task was 'too deep' it may also have been the case that the shallow processing task was not sufficiently 'shallow' to promote the restrictive word-level processing intended for that condition. Given the relatively simple nature of these short passages, as mentioned earlier, perhaps participants in this condition were still able to extract meaning from the text, even while paying attention to factors associated with the individual words. On the questionnaire, therefore, one would expect overall high ratings with respect to an identification of a theme and awareness of a story structure. But because participants were focused on non-semantic features of the words in the story, including a number of unrelated items and function words, their memory representations of the list associates may have been too poor to lead to their correct identification later on the recognition test, which in turn, would have been associated with a reduced susceptibility to false recognition of the lure items. Conversely, under deep processing instructions, processing of extraneous words would have been minimal and relatively automatic, leading to better representations in memory of content words, which mostly included a high number of the DRM list items.

Implications for Theories of the DRM Effect

Whatever the exact cause(s) of some of the results of Experiment 4, it is clear that the novel merging of the DRM task, depth of processing, and story context, calls for more careful and systematic investigation in potential future research, particularly with respect to how these tasks and variables interact. The finding that false recognition was elevated under semantic processing of thematic material, however, is consistent with the general story effect reported in the first three experiments. It is from the results of the first three experiments reported here that the evidence for the story effect is most clearly supported.

It is worth repeating here that the apparent influence of a story context in producing the DRM memory illusion was principally based on a direct comparison of a similar context in which the narrative textual elements were removed. This method of eliciting the DRM illusion as a function of story context inherently complicates a direct comparison with past DRM studies, which have primarily been based on the traditional list-learning paradigm. As already stated, this was not the principal aim of this investigation. Rather, the chief goal of the present research was to examine more closely semantic processes believed to enhance the DRM effect via the introduction of a novel methodology that incorporated text-level processing. In turn, since this methodology represented an intended departure from past DRM studies, it permitted the closer examination of a second and related goal concerning the use of more naturalistic materials (in this case, narratives) in producing false memories while retaining traditional DRM stimuli and testing procedures.

As such, these set of experiments have produced some important and interesting empirical and theoretical contributions to the study of the DRM task and false memory. In particular, the present series of studies have added to the existing database of empirical findings on the DRM effect by contributing the investigation of a new, as yet largely unexplored, variable that moderates the false memory effect: story context. The results of these four studies have demonstrated clearly that the DRM effect is not limited to the standard list-learning paradigm that has primarily dominated the research over the last 14 years. With respect to the great amount of theoretical work that has been put forth to account for the DRM effect, any theory of its underlying mechanisms should now be expanded to include the influence of placing traditional DRM stimuli in a story context.

In many ways, the present research is a logical extension of many of the studies that have already been conducted and which were discussed in several places above. For example, the fact that blocked presentation of DRM lists leads to higher rates of false recall and recognition than random presentation suggests that preserving the theme or associative structure of the lists is crucial to the production of false memories of nonpresented, critical lures. Under such conditions, people are more likely to engage elaborative strategies during encoding. In the present set of experiments, not only was the thematic or associative nature of the lists preserved, it was explicitly provided, thereby leading to similar but perhaps more effective use of such encoding strategies.

Unfortunately, as often seems the case with DRM studies, many of the empirical results that arise from them can usually be accommodated by aspects of both of the two leadings theories of the DRM effect that have been the main focus throughout this discussion. The present investigation is certainly no exception. As we've seen, there are sometimes particular experimental variables that moderate false recall or recognition that are more easily explained by one or the other of these accounts. A story context variable might seem to lend itself more easily to an explanation by the fuzzy trace account given its emphasis on thematic consistency and the distinction between verbatim and gist processing (Brainerd et al. 2001). Neuschatz, Lampinen, Preston, Hawkins, and Toglia (2002) later extended this distinction to include what they referred to as global gist extraction. They proposed that individuals can encode gist representations based on the meaningful relationships among a series of items. So, for example, if a person is presented with the words, *bed*, *rest*, *nap*, *awake*, *dream*, and *snooze*, etc., not only will the meaning of each of the individual items be processed (local gist) but a person might also extract the global gist that the list contains words that are all related to the word *sleep*. When later presented with the word *sleep* on a recognition test, a person is likely to falsely recognize that word due to its thematic consistency. From this perspective, a story context that reinforces the meaningful relationships among several semantic associates embedded within that story would not only aid the global gist extraction process itself, but serve to direct the meaning of that extraction toward a theme associated with the critical nonpresented item, when the text is considered as a whole.

The present findings can also be couched in terms of the activation/monitoring account as described earlier. Associative activation assumes that people form a mental lexicon consisting of networks of semantically-related, or otherwise associated, words. When a word in a semantic network is presented, its activation is assumed to spread to other words in the network (Collins & Loftus, 1975). Typically, however, this activation is too short-lasting and temporary to produce false recall or recognition on a later memory test. But if multiple related words are presented that converge on a nonpresented critical item, activation of that critical item may lead to a longer lasting memory representation (Gallo, 2006). As with global gist extraction, a story context associatively related to its constituent lexical representations could be viewed as strengthening and maintaining this representation. Upon being presented with a nonpresented item during a recognition test, after its unconscious activation or conscious elaboration during study, an impoverished ability to distinguish an internally versus an externally generated event due to a source-monitoring error would lead to its subsequent false remembrance.

Either way, encouraging the text processing of the thematic elements of the DRM lists appears to take advantage of the mechanisms, whether associative activation or gist extraction, purported to underlie the production of false memory in the standard converging-associates task.

The current experiments therefore provide additional evidence for the theoretical mechanisms underlying the false memory effect by showing that text processing enhances the DRM effect through its capacity to strengthen those underlying mechanisms. It should also be noted here that given the departure from the standard DRM list-learning paradigm with the use of narrative text processing to elicit the false memory effect, the standard theoretical explanations of the DRM effect may require some additional assumptions. The story effect could also be explained by theories of discourse processing or specifically, models that have focused on story understanding, memory for factual material presented in texts, learning from texts, and the inferences involved in this process (e.g., van Dijk & Kintsch, 1983). In this view, text comprehension consists of forming mental representations (textbases and situation models) which connect the various elements of the text representation in network-like structures. Such an explanation is useful for understanding the finding that memory for lures was driven in large part by memory for list items, such that all items became part of a single memory complex and the textbase or situational levels of representation. At these levels, false memory recognition could be explained by the preservation of the semantic nature of the text or a person's own existing knowledge base without verbatim memory for the text.

More broadly speaking, a better understanding of the mechanisms believed to drive the DRM effect not only helps us to better explain the nature of false memories, but of memory in general, which has historically been based on the concept of association. Just as perceptual illusions are often used to demonstrate and understand ordinary perceptual processes, so too can memory illusions be used to further our understanding of ordinary memory. Thus, any variable, such as text processing, that has an important effect on false memory, by virtue of its impact on the same underlying mechanisms, can therefore be useful in understanding memory in general.

Conclusions and Future Directions

In addition to its theoretical implications, the story effect demonstrated here potentially opens numerous new avenues for empirical investigations relevant to the DRM effect. It would be predicted, for example, that the benefits of story context might be greater for some participants or for some DRM lists. Specifically, story manipulations via stimuli or instructions should be stronger for participants less susceptible to the false memory effect and weaker for those highly susceptible. Dewhurst et al. (2007) have confirmed the first prediction with children. Indeed, the evidence for the influence of contextual manipulations on increasing the false memory effect in younger populations continues to grow (Odegard, Holliday, Brainerd, & Reyna, 2008).

Using a paired associative learning paradigm, Odegard et al. (2008) paired DRM list items related to a critical item (e.g., *window*) with context words meant to direct attention toward the theme of the list (e.g., *shade-drapes*) or away from the theme of the list (*shade-tree*). On a subsequent paired associate recognition memory test, 11-year-olds were more likely to falsely accept unpresented pairs (e.g., *window-shade*) following study of theme-consistent context words than theme-inconsistent items. This context effect, however, was not observed among adult participants, presumably due to their greater ability to spontaneously generate inter-item associations among related DRM-list items. Odegard et al. (2008) concluded that the typical reduction in the false memory effect among younger children can be eliminated by promoting the processes that lead to the extraction of the global gist within a DRM list.

This interpretation is consistent with the findings reported by Dewhurst et al. (2007) and the explanation of the story context effect investigated in the present set of experiments. That is, theme-consistent narrative text processing of DRM-list items (Story Context) during study can have an analogous effect on adults' tendency toward false recognition, compared to when the ability to process the theme of a list is reduced (i.e., the Scrambled Context condition). In this way, scrambled text may be thought to produce the same attenuation in false recognition in adults in a manner similar to the way in which only lists or theme-inconsistent context words lead to a reduced susceptibility to false recognition in children.

Further studies could also examine the interaction of story context with other relevant individual or group differences, particularly for people who appear less susceptible to the DRM false memory illusion under standard list-learning conditions. With respect to materials, story contexts and thematic instructions may have stronger effects for lists of items that only weakly demonstrate the DRM effect than for materials that demonstrate robust DRM effects. It would be interesting to see, for example, whether DRM lists that have a relatively low BAS might be strengthened by a story context in their ability to elicit false remembering. Strengthening either the associative links between list items or enhancing the thematic nature of a weak DRM list via a narrative text structure might be a potentially useful method for investigating the conditions under which a story context could lead to elevated levels of false memory compared to list presentations, particularly in individuals who are reliably susceptible to false remembering following presentation of typically strong DRM lists.

It would also be worth examining whether or not the pattern of results due to story context reported here could be extended to false recall, given that the present set of findings apply only to false recognition. One could examine the rate of intrusions on a free recall task following the presentation of a series of DRM stories, perhaps with the added instructions at test to omit function words. Alternatively, after a single story, perhaps with a more natural presentation (either visual or auditory) than what was used here, participants could be instructed to reproduce the story in spoken or written format, with explicit instructions to orient participants toward verbatim reproductions, and look for the rate of intrusions in their recall protocols.

Experiments 3 and 4 allowed for some very preliminary results on the effect of context and other factors on recall. Following the rating task (which followed the recognition task), participants were asked to write down four items from each of the four originally studied lists. This was originally done as a check in the within-subjects design that ratings did indeed correspond with the list and story conditions, but this practice was continued in Experiments 3 and 4. Recall of lures was scored for those studies; scores could range from 0 to 4 for the critical lures and 0 to 4 for the weak lures (one of each from each of four lists). Recall was universally low for lures from non-studied lists despite exposure to these items in the recognition task. In fact means approached or reached 0 in the various conditions examined.

In Experiment 3, the effect of scrambled versus story conditions on incidental recall was present only for the critical lures and only for the list instructions. With list instructions (the only instructions used in Experiments 1 and 2), false recall of critical lures was higher for the story (M = 1.32) than the scrambled (M = 0.63) contexts (p = .007). With thematic instructions, false recall was similar for the story (M = 1.38) and the scrambled (M = 1.41) contexts (p = .930). If confirmed, this pattern suggests that either story context or thematic instructions can enhance the false memory recall effect over the standard non-story (scrambled) context with standard list instructions, but that the factors are not additive in their influence.

Only story contexts were used in Experiment 4, so it was not possible to examine the context effect in that study. Recall was again universally low for lures from nonstudied lists. There was a marginally significant instruction effect, but only for the weak lures. Specifically, false recall of weak lures was somewhat higher for semantic processing (M = .30) than for word-level processing ($M = .12$) instructions ($p = .077$). For critical lures there was no difference between semantic ($M = 1.14$) and word-level ($M = 1.27$) instructions ($p = .600$). Although both these effects, especially that from Experiment 4, need to be replicated in studies specifically designed to examine recall (e.g., use of more lists, explicit recall instructions during study), the findings suggest that the present model applies to recall as well as recognition.

Finally, in addition to building on previous research using the DRM paradigm, the present studies took the investigation of false memory as elicited by the DRM effect in a new direction. By extending the novel paradigm introduced by Dewhurst, a primary goal of this research was to look at the nature of the false memory effect when it is produced by more natural materials (i.e., stories) rather than simply lists of words.

Over and above the apparent practical advantage of using story contexts to elicit the DRM effect under certain conditions, is the theoretical importance of using story contexts as a means of creating false memories. The role of integrated text in eliciting distorted and false memories has a long history dating back to at least the early work of Bartlett (1932). The current investigations serve to bridge the DRM paradigm with everyday false memories, by incorporating materials that people more commonly encounter in their lives, such as stories and narratives. Moreover, these experiments are timely given that they merge nicely with current work by other researchers with similar objectives, which together may be leading to a 'second wave' of research into the nature of false memory.

The studies are timely as well because the traditional DRM task has been criticized for not being able to demonstrate how the mechanisms underlying the DRM effect may apply to an understanding of real world cases of false memory, involving the planting of entirely new events in memory (Freyd & Gleaves, 1996; Pezdek & Lam, 2007). Roediger and McDermott (1996) have argued, however, that their original findings (i.e., Roediger & McDermott, 1995) should not and cannot be directly generalized to situations in which certain therapeutic practices may lead to the creation of false memories. Miller and Gazzaniga (1998), however, made a stronger point when they argued that false memories produced by semantic associations may have less relevance than schema-based knowledge in understanding how false memories are created outside of the laboratory. Thus, one additional and important benefit of these newer approaches perhaps may be the ability to more closely align DRM lines of investigation with the practical application of understanding the formation of nonlaboratory-based false memories. The discussion of the story effect based on the current set of findings has focused primarily on its possible semantic origins, due to the use of traditional DRM stimuli. But since the story effect is a product of an experimental methodology employing story context, as has been argued here, a particular advantage of using narrative text processing in future research may lie in the relative ease with which any account of the theoretical explanations of its underlying mechanisms can be transferred between the possible role of semantic processes and more general knowledge structures that may underlie the production of experimentally-induced false memories and real-world cases of similar memory distortions.

In sum, the novel paradigm originally introduced by Dewhurst et al. (2007) for use with younger populations, and extended in this set of experiments for adult participants, involving embedding DRM-list stimuli into story contexts, has been shown to be a useful method for eliciting the DRM effect via narrative text processing. The story effect that this produces demonstrates that traditional DRM stimuli need not be limited to the simple list-learning paradigm that has dominated the empirical investigation and theoretical attention to the nature of this memory illusion since the seminal report of Roediger and McDermott (1995).

References

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

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

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology.* Cambridge, England: Cambridge University Press.

Brainerd, C. J., & Reyna, V. F. (1996). Mere memory testing creates false memories in children. *Developmental Psychology*, *32*, 467-478.

Brainerd, C. J., & Reyna, V. F. (1998). Fuzzy trace theory and children's false memories. *Journal of Experimental Child Psychology*, *71*, 81-129.

Brainerd, C. J., Reyna, V. F., & Forrest, T. J. (1998). Are young children susceptible to the false-memory illusion? *Child Development*, *73*, 1363-1377.

Brainerd, C. J., Wright, R., Reyna, V. F., & Mojardin, A. H. (2001). Conjoint recognition and phantom recollection. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 307-327.

Butler, K. M., McDaniel, M. A., Dornburg, C. C., Price, A. L., & Roediger, H. L. (2004). Age differences in veridical and false recall are not inevitable: The role of frontal lobe function. *Psychonomic Bulletin & Review*, *11*, 921-925.

Chan, J. C. K., & McDermott, K. B. (2006). Remembering pragmatic inferences. *Applied Cognitive Psychology*, *20*, 633-638.

Chan, J. C. K., McDermott, K. B., Watson, J. M., & Gallo, D. A. (2005). The importance of material-processing interactions in inducing false memories. *Memory & Cognition, 33*, 389-395.

Clancy, S., Schacter, D. L., McNally, R. J., & Pitman, R. K. (2000). False recognition in women reporting recovered memories of sexual abuse. *Psychological Science, 11*, 26-31.

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

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

Dewhurst, S. A., & Robinson, C. A. (2004). False memories in children: Evidence for a shift from phonological to semantic associations. *Psychological Science, 15*, 782-786.

Dewhurst, S. A., Barry, C., & Holmes, S. J. (2005). Exploring the false recognition of category exemplars: Effects of divided attention and explicit generation. *European Journal of Cognitive Psychology, 17*, 803-819.

Dewhurst, S. A., Barry, C., Swannell, E. S., Holmes, S. J., & Bathurst, G. L. (2007). The effect of divided attention on false memory depends on how memory is tested. *Memory & Cognition, 35*, 660-667.

Dewhurst, S. A., Holmes, S. J., Swannell, E. S., & Barry, C. (2008). Beyond the text: Illusions of recollection caused by script-based inferences. *European Journal of Cognitive Psychology, 20*, 367-386.

Dewhurst, S. A., Pursglove, R. C., & Lewis, C. (2007). Story contexts increase susceptibility to the DRM illusion in 5-year-olds. *Developmental Science, 10*, 374-378.

Dodd, M. D., & MacLeod, C. M. (2004). False recognition without intentional learning.

Psychonomic Bulletin & Review, 11, 137-142.

Freyd, J. J., & Gleaves, D. H. (1996). Remembering words not presented in a list: Implications for the recovered/false memory controversy? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 811-813.

Gallo, D. A. (2006). *Associative illusions of memory: False memory research in DRM and related task.* Psychology Press: New York and Hove.

Gallo, D. A., McDermott, K. B., Percer, J. M., & Roediger, H. L. (2001). Modality effects in false recall and recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 27*, 339-353.

Gallo, D. A., Roberts, M. J., & Seamon, J. G. (1997). Remembering words not presented in lists: Can we avoid creating false memories? *Memory & Cognition, 4*, 271-276.

Gallo, D. A., & Roediger, H. L. (2002). Variability among word lists in eliciting memory illusions: Evidence for associative activation and monitoring. *Journal of Memory and Language, 47*, 469-497.

Gallo, D. A., & Roediger, H. L. (2003). The effects of associations and aging on illusory recollection. *Memory & Cognition, 31*, 1036-1044.

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

Howe, M. L. (2005). Children (but not adults) can inhibit false memories. *Psychological Science, 16*, 927-931.

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

Keane, M. M., Orlando, F., & Verfaellie, M. (2006). Increasing the saliency of fluency cues reduces the recognition memory impairment in amnesia. Neuropsychologia, 44, 834-839.

Kintsch, W., Mandel, T. S., & Kozminsky, E. (1977). Summarizing scrambled stories. *Memory & Cognition*, *5*, 547-552.

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behaviour*, *13*, 585-589.

McCabe, D. P., & Smith, A. D. (2002). The effect of warnings on false memories in young and older adults. *Memory & Cognition*, *30*, 1065-1077.

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

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

McDermott, K. B. (2007). Inducing False Memories Through Associated Lists: A Window Onto Everyday False Memories? In James S. Nairnes (Ed.), *The foundations of remembering: Essays in honor of Henry L. Roediger, III* (pp. 297-311). New York: Psychology Press.

McDermott, K. B., & Chan, J. C. K. (2006). Effects of repetition on memory for pragmatic inferences. *Memory and Cognition*, *34*, 1273-1284.

McDermott, K. B., & Watson, J. M. (2001). The rise and fall of false recall. *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*, *25*, 1177.

Melo, B., Winocur, G., & Moscovitch, M. (1999). False recall and false recognition: An examination of the effects of selective and combined lesion to the medial temporal lobe/diencephalons and frontal lobe structures. *Cognitive Neuropsychology, 39*, 508-520.

Miller, M. B., & Gazzaniga, M. S. (1998). Creating false memories for visual scenes. *Neuropsychologia, 36*, 513-520.

Neuschatz, J. S., Lampinen, J., Preston, E. L., Hawkins, E. R., & Toglia, M. P. (2002). The effect of memory schemata on memory and the phenomenological experience of naturalistic situations. *Applied Cognitive Psychology, 16*, 687-708.

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.

Odegard, T. N., Holliday, R. E., Brainerd, C. J., & Reyna, V. F. (2008). Attention to global gist processing eliminates age effects in false memories. *Journal of Experimental Child Psychology, 99*, 96-113.

Park, L., Shobe, K. K., & Kihlstrom, J. F. (2005). Associative and categorical relations in the associative memory illusion. *Psychological Science, 16*, 792-797.

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 and Language, 35*, 261-285.

Perez-Mata, M. N., Read, J. D., & Diges, M. (2002). Effects of divided attention and word concreteness on correct recall and false memory reports. *Memory, 10*, 161-177.

Pezdek, K., & Lam, S. (2007). What research paradigms have cognitive psychologists used to study "False memory," and what are the implications of these choices? *Consciousness and Cognition, 16*, 2-17.

Read, J. D. (1996). From a passing thought to a false memory in 2 minutes: Confusing real and illusory events. *Psychonomic Bulletin & Review, 3*, 105-111.

Rhodes, M. G., & Anastasi, J. S. (2000). The effects of a levels-of-processing manipulation on false recall. *Psychonomic Bulletin & Review, 7*, 158-162.

Robinson, K. J., & Roediger, H. L. (1997). Associative processes in false recall and recognition, *Psychological Science, 8*, 231-237.

Roediger, H. L. (1996). Memory illusions. *Journal of Memory and Language, 35*, 76-100.

Roediger, H. L., & Gallo, D. A. (2004). Associative memory illusions. In R. F. Pohl (Ed.). *Cognitive Illusions: A Handbook of Fallacies and Biases In Thinking, Judgment and Memory* (pp. 309-326). Oxford: Oxford University Press.

Roediger, H. L., & 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., & McDermott, K. B. (1996). False perceptions of false memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 814-816.

Roediger, H. L., McDermott, K. B., & Robinson, K. J. (1998). The role of associative processes in creating false memories. In M. A. Conway, S. E. Gathercole, & C. Cornoldi (Eds.), *Theories of Memory II* (pp. 187-245). Hove, England: Psychology Press.

Roediger, H. L., Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review, 8*, 385-407.

Schacter, D. L., Verfaellie, M., & Pradere, D. (1996). The neuropsychology of memory illusions: False recall and recognition in amnesic patients. *Journal of Memory and Language, 35*, 319-334.

Seamon, J. G., Luo, C. R., & Gallo, D. A. (1998). Creating false memories of words with or without recognition of list items: Evidence for nonconscious processes. *Psychological Science*, *9*, 20-26.

Seamon, J. G., Goodkind, M. S., Dumey, A. D., Dick, E., Aufseeser, M. S., Strickland, S. E., Woulfin, J. R., & Fung, N. S. (2003). "If I didn't write it, why would I remember it?" Effects of encoding, attention, and practice on accurate and false memory. *Memory & Cognition*, *31*, 445-457.

Singer, M., & Remillard, G. (2008). Veridical and false memory for text: A multiprocess analysis. *Journal of Memory and Language*, *59*, 18-35.

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

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

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*, 137-149.

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 less is more. *Memory*, *7*, 233-256.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychologist*, *26*, 1-12.

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.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse processing*. New York: Academic Press.

Watson, J. M., Bunting, M. F., Poole, B. J., & Conway, A. R. A. (2005). Individual differences in susceptibility to false memory in the Deese-Roediger-McDermott paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 76-85.

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.

Watson, J. M., Balota, D. A., & Roediger, H. L. (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.

Winograd, E., Peluso, J. P., & Glover, T. A. (1998). Individual differences in susceptibility to memory illusions. *Applied Cognitive Psychology, 12*, S5-S27.

Appendix A

Example of the three list conditions in Experiment 1 for one story. The critical word is *sleep*. The additional list item removed from the norm list to serve as a second critical lure was *nap*.

For each of the eight critical words, there remained fourteen studied items from each list.

	Story Context	List Context	Scrambled Context
1	Cindy	.	still
2	lay	.	lay
3	in	.	heavily
4	bed	bed	bed
5	Although	.	Cindy
6	needing	.	needing
7	rest	rest	rest
8	she	.	began
9	remained	.	remained
10	awake	awake	awake
11	and	.	big
12	tired	tired	tired
13	She	.	slipped
14	wanted	.	and
15	to	.	getting
16	dream	dream	dream
17	and	.	in
18	not	.	to
19	wake	wake	wake
20	up	.	Cindy
21	She	.	and
22	began	.	lay
23	to	.	a
24	snooze	snooze	snooze
25	and	.	her
26	held	.	out
27	her	.	up
28	blanket	blanket	blanket
29	tight	.	a
30	Cindy	.	and
31	slipped	.	she
32	from	.	from
33	a	.	began
34	doze	doze	doze
35	to	.	not
36	deep	.	deep
37	slumber	slumber	slumber

Appendix A (continued)

	Story Context	List Context	Scrambled Context
38	and	.	she
39	began	.	up
40	to	.	to
41	snore	snore	snore
42	heavily	.	letting
43	She	.	tight
44	lay	.	she
45	in	.	wanted
46	peace	peace	peace
47	until	.	to
48	getting	.	she
49	up	.	to
50	and	.	and
51	letting	.	to
52	out	.	in
53	a	.	and
54	big	.	held
55	yawn	yawn	yawn
56	still	.	although
57	drowsy	drowsy	drowsy

Appendix B

The eight stories used in Experiments 1-4 with the fourteen DRM list associates underlined.

Below in parentheses, are each of the stories' critical lure items, weak lures, and total number of words.

Although hot inside, Jack saw snow outside the warm house. In winter, ice covered the wet ground. Entering the frigid outdoors, Jack felt the chilly weather. He expected to freeze as the air made him shiver. Jack imagined that arctic frost surrounded him.
(cold, heat, n = 43)

The nurse visited Sally who was sick. Her lawyer mother told Sally to take medicine for her health or visit a hospital. Sally hated them more than the dentist. Sally saw a physician when ill before but his office and stethoscope scared her. At a different clinic, a surgeon found a cure.
(doctor, patient, n = 52)

The tiger did not belong in a circus, but in the jungle. Here they need a tamer but they live in a den with cubs in Africa. Tom also imagined creatures with a mane in a cage. These feline creatures are fierce like bears. Tom remembered how they hunt with their pride.
(lion, roar, n = 52)

Cindy lay in bed. Although needing rest, she remained awake and tired. She wanted to dream and not wake up. She began to snooze and held her blanket tight. Cindy slipped from a doze to deep slumber and began to snore heavily. She lay in peace until getting up and letting out a big yawn, still drowsy.
(sleep, nap, n = 57)

John held his nose until he could not breathe, or even sniff. He could see and hear but covered his nostrils to avoid the whiff. A nice scent might mask the reek and stench. Imagining a fragrance of perfume, bath salts, and rose helped.
(smell, aroma, n = 44)

Susan knew it was wrong to steal. She once read about a robber. The crook was a burglar caught taking money by a cop. It was bad to rob so the man went to jail. People called him a villain for holding up a bank, just like a bandit. The criminal was punished for his crime.
(thief, gun, n = 56)

Striking a note produced sound from the piano. Brad began to sing along. On the radio, a band started. He heard a horn and imagined the instrument in a concert, maybe a famous symphony. Brad also wanted to try jazz with an orchestra. His teacher called it an art and helped him feel the rhythm.
(music, melody, n = 55)

Jenny and her dad picked up vegetables, and then an orange and a kiwi. Her dad called them citrus. Jenny put a ripe pear, banana, berry, and cherry in the basket. That was for juice to fill the salad bowl with a cocktail.
(fruit, apple, n = 43)

Appendix C
Processing Questionnaire

I would now like you to think back to the four lists you heard at the beginning of this experiment and answer some questions. Please answer as carefully as you can but do not worry if you cannot remember or are unable to provide an answer to a particular question for one or more of the lists you heard. Answer the following questions as best as you can.

Use the following rating scale to make your judgments. Questions that you would definitely Agree with should be rated 5, and question that you definitely Disagree with should be rated 1. Use values of 2, 3, or 4, to indicate moderate degrees of agreement or disagreement.

Definitely Disagree				Definitely Agree
1	2	3	4	5

We will start with the ***first list***. Think about that first list for a few moments, and then answer each of the following questions with respect to that list and that list only.

_____ 1. During the recognition test for this list, I based my judgments about whether an item had been studied or not solely on my memory for that specific item and not on whether it was related to other words.

_____ 2. As I was listening to the words presented in this list, I became aware that many of the words were all related to a common idea or theme.

_____ 3. I became aware very early in the presentation of this list (first 10 or so words) that many of the words were all related to a common idea or theme.

_____ 4. I never became aware during the presentation of this list that many of the words were all related to a common idea or theme.

_____ 5. As I was listening to the words presented in this list, I became aware that the words were connected in such a way that they formed a series of sentences that made up a short story.

_____ 6. I became aware very early in the presentation of this list (first 10 or so words) that the words were connected in such a way that they formed a series of sentences that made up a short story.

_____ 7. I never became aware during the presentation of this list that the words were connected in such a way that they formed a series of sentences that made up a short story.

_____ 8. During the recognition test for this list, I used my knowledge of a story or common theme to decide whether an item had been presented for study or not.

9. Please write below a brief description of the theme or general idea that relates many of the items presented on this list.

____