Robust False Memory Effects Under Conditions of 
Paired-Associate Learning

By

Erin Denise Sheard

A thesis submitted in conformity with the requirements
for the degree of Master of Arts
Graduate Department of Psychology
University of Toronto

© Copyright by Erin Denise Sheard (2000)
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.

L’auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L’auteur conserve la propriété du droit d’auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-50468-9
Abstract

Robust False Memory Effects Under Conditions of Paired-Associate Learning

Erin Denise Sheard, Master of Arts, 2000

Department of Psychology, University of Toronto

Substantial levels of false recall and recognition occur for lists of words semantically related to an unpresented word (Deese, 1959; Roediger & McDermott, 1995). Manipulations that have attenuated this “false memory effect” have reduced the probability of the presented words being linked semantically. We employed a paired-associate learning procedure to de-emphasize the relation between list items and to make individual list items more distinctive. Subjects learned 12 word associates either (a) as 12 individual words (standard condition), (b) as 12 words, each paired with a number (number-noun condition), or (c) as 6 two-word pairs (noun-noun condition). If pairing words with numbers minimizes semantic analysis of each item and reduces the probability of forming links between items, the false memory effect should be reduced. However, the results indicated increased false memory in the number-noun condition, relative to the standard and noun-noun conditions, which did not differ. Correlational analyses were used to explore the causes of the false memory effect.
Acknowledgements

I would like to express thanks to my supervisor, Colin M. MacLeod for his support, encouragement, and insight during all stages of the research process. I would especially like to extend my thanks for his valuable comments on earlier versions of this document. I would also like to thank my subsidiary advisor, Marilyn Smith for comments on an earlier versions of this document. Finally, I would like to thank Robert D. Scheifley for his unyielding support, encouragement, and assistance.
# Table of Contents

Title Page .................................................................................................................. i
Abstract ..................................................................................................................... ii
Acknowledgements .................................................................................................. iii
List of Appendices ..................................................................................................... v
List of Tables ............................................................................................................ vi
Introduction ............................................................................................................... 1
Method  
- Participants ........................................................................................................ 30
  - Design and Experimental Control ................................................................. 30
  - Materials .......................................................................................................... 31
  - Procedure ......................................................................................................... 32
Results ....................................................................................................................... 33
Discussion ............................................................................................................... 43
References .............................................................................................................. 60
List of Appendices

Appendix A  Instruction
Appendix B  Recognition Test
Appendix C  Paired-Associates Test (Noun-Noun Group)
Appendix D  Paired-Associates Test (Number-Noun Group)
List of Tables

Table 1  Mean proportion of correct recall and recognition as a function of encoding condition

Table 2  Mean proportion of correct recall and recognition as a function of list position and encoding condition

Table 3  Mean proportion of non-critical recall errors as a function of error type and encoding condition

Table 4  False memory: Mean proportion of recall and recognition of the critical unpresented lure as a function of encoding condition

Table 5  Mean recall position of incorrectly recalled critical lures as a function of encoding condition

Table 6  Mean proportion of non-critical recognition errors as a function of error type and encoding condition

Table 7  Mean proportion of paired-associates recalled as a function of encoding condition

Table 8  Correlations in test performance as a function of type of response and encoding condition
False Memory

Robust False Memory Effects Under Conditions of Paired-Associate Learning

False memories, or illusions of memory, have fascinated researchers studying memory for centuries. Although the term encompasses many ideas, a false memory can be defined as a report of events that deviates from the actual event in some way. In the laboratory, false memory is often studied by analyzing memory errors a subject makes following the presentation of some stimulus material (e.g., stories, pictures, sentences, word lists). The deviation, or error, can be one of commission, distortion, substitution, or confabulation. Interest in false memories has a long history (see Roediger, 1996; Schacter, 1995; also see Intons-Peterson & Best, 1998, for recent reviews), but a recent surge in public interest over issues such as recovered memory and eyewitness testimony has revived experimental interest in the topic (see Bruck & Ceci, 1997; Loftus, 1997; Lynn, Lock, Myers, & Payne, 1997; Payne, Neuschatz, Lampinen, & Lynn, 1997; Shobe & Kihlstrom, 1997, for reviews).

False Recollection in Narrative Reconstruction

Although curiosity and interest in memory distortions can be traced back to Aristotle (Intons-Peterson & Best, 1998; Schacter, 1995), Bartlett’s (1932) monograph “Remembering” is generally credited with beginning the modern experimental study of false memory. In his now famous experiment, Bartlett used a North American folk-tale entitled “The War of the Ghosts” to test subjects’ recall of a narrative over repeated attempts. The passage was read twice by each subject, prior to repeated testing occurring over at least two intervals (with no study opportunities between intervals), the first test occurring 15 minutes after reading the passage and subsequent tests occurring at various time intervals (up to a ten-
Bartlett based his selection of this particular narrative on several grounds: The story itself belonged to a culture different from that of the test subjects, there was a lack of order to the story, details of the story tended to arouse vivid visual imagery, and the conclusion employed a supernatural element. Although no statistical analyses of the data were provided, the results seemed to demonstrate a decline in accuracy over repeated reproductions, although the general form of the story (timing, style, and rhythm) was reproduced over repeated trials. In general, remembering was shown to be inaccurate more than it was accurate; in Bartlett’s words, accuracy was found to be “the rare exception and not the rule” (p. 93).

Another intriguing finding from Bartlett’s (1932) study was that many of the errors of commission were such that the subjects added information to bring the narrative into line with their own cultural schemas, a process Bartlett dubbed rationalization. This rationalization took the form of changing names, places, and events to coincide with current social norms, explaining away details inconsistent with their current schema (e.g., the black from the mouth of the dying man explained as his breath as he died), linking together events with added details not present in the original passage, and a loss of symbolism over recall attempts. Rationalization served to render the story more acceptable and understandable to the subjects - in other words, a rendition of the story that conformed to their own cultural expectations and beliefs. Bartlett’s book still has a powerful impact on scholars today, and his schema theory of memory is still used to account for many distortions of memory (Alba & Hasher, 1983).

Perhaps the most enduring legacy left by Bartlett (1932) is his distinction between reproductive memory and reconstructive memory. Reproductive memory, said to be
emphasized in rote recall of nonsense syllables or serial word lists, was abandoned by Bartlett in favour of more complex materials thought to emphasize reconstructive memory. Bartlett argued that for complex materials, such as stories, remembering is based on a general schema around which details of the event are built. The reconstructive process, as opposed to the reproductive process of veridical recall, was thought to be more likely to generate memory errors or memory distortions. Supporting this idea was the general finding of few errors of commission or distortion in free recall list learning paradigms compared to prose materials.

Bartlett (1932) saw this reconstructive remembering found using complex materials as more indicative of real-life memories than was remembering of less complex materials, which were assumed to emphasize mere rote memorization. Beginning in the 1970's, Bartlett's ideas regarding reconstructive memory really began to take hold in mainstream memory research and set the pace for studying errors of memory for the next two decades (Bruce & Winograd, 1998). Researchers for many years followed this path set by Bartlett, choosing to focus on complex stimuli such as sentences (e.g., Bransford & Franks, 1971), narratives or stories (e.g., Hasher & Griffin, 1978), and various types of visual stimuli (e.g., Posner & Keele, 1970) and pictures (e.g., Rosenberg & Simon, 1976). Similarly, the predominant attitude has been that serial list learning procedures do not allow for reconstructive memory processes, the very processes Bartlett asserted were the most interesting aspects of memory.

Although Bartlett used a free-recall paradigm, most false memory studies since then have utilized recognition paradigms, probably due to the difficulty in obtaining false memory effects on free recall tests. In general then, most false memory studies have focused on
memory for prose materials and employed recognition tests of memory. Alterations from this pattern have tended to produce few, or no, distortions of memory.

**False Recollection in List Learning**

One important deviation from the dominance of Bartlett's complex reconstructive material in false memory research is Underwood's (1965) study of false recognition using a serial list learning paradigm. Contained within the study list were critical words assumed to elicit an implicit associative response, a word associated with the presented word (e.g., the list word *bread* could elicit the unpresented word *butter*). For three classes of critical stimulus words, Underwood found that false recognition of the implicit associative response was much higher than false recognition of control words. Specifically, implicit associative responses to antonyms (e.g., *top* for *bottom* or *day* for *night*), converging associates (e.g., *bed* and *dream* for *sleep*), and superordinates (e.g., *maple*, *oak*, *elm*, and *birch* for the superordinate *tree*) resulted in false recognition. Note that the presentation of an antonym three times, but not once, resulted in false recognition, emphasizing frequency of presentation of the associate as an important variable.

Underwood's results seem to offer great potential for false memory research using less complex stimuli, but there have been few replications of his findings. Anisfeld and Knapp (1968) did replicate Underwood's (1965) findings using a list learning paradigm with common associative responses and synonyms as the critical lure words. They found higher rates of false recognition errors to word pairs that were bi-directionally related or where the associative link was in a forward direction (from the presented word to the lure), compared to where the associative link was in a backward direction (from the lure to the presented word). However, they noted that the lower number of false recognitions when words were
associated in a backward direction could be due to the lower number of presentations of backward direction critical lures (once, compared to twice for the forward direction critical lures). Underwood (1965) also speculated about the necessity of a bi-directional relation between the word associates, hypothesizing that the lack of a false recognition effect for sense impressions (e.g., doughnut, dome and globe for round) could be related to the unidirectionality of the association. In contrast to Underwood and Anisfeld and Knapp, other demonstrations of false-recognition in list learning paradigms have shown fairly modest effects (e.g., MacLeod & Nelson, 1976), with some studies failing to elicit a false recognition effect at all (e.g., Gillund & Shiffrin, 1984). As Roediger and McDermott (1995) state, the rarity of robust false recognition effects in list-learning paradigms did little to counteract Bartlett’s (1932) assertion that complex stimuli are needed to demonstrate the powerful effects of false memory.

The Deese-Roediger-McDermott False Memory Paradigm

Another notable exception to the assertion that complex stimuli are necessary to elicit robust false memory effects is the list learning paradigm developed by Deese (1959) and revived by Roediger and McDermott (1995). The paradigm, now dubbed the DRM (pronounced “dream”) paradigm, has, in recent years, become the dominant paradigm used to study false recognition and false recall (Bruce & Winograd, 1998; Stadler, Roediger, & McDermott, 1999). In contrast to the relatively modest false memory effects in past list learning paradigms, and particularly in free recall tests, Deese was able to elicit false recall errors with high frequency. In fact, Deese’s results are of particular interest because he employed two procedures typically assumed to be unlikely to yield false recall effects or memory errors - list learning and free recall. Deese, noting that list recall was often
contaminated by so-called "unpredictable extra-list intrusions," sought to develop a method for predicting the occurrence of the extra-list intrusions by showing that they are often associatively related to words presented in the list.

Using a word-association technique, Deese (1959) developed 36 lists consisting of 12 words each. The twelve words for each list were the primary word associates for a given critical stimulus not presented, referred to as the critical lure (e.g., sour, candy, sugar, bitter, good, taste, tooth, nice, honey, soda, chocolate, and heart for the critical lure sweet). Free recall of the critical lure ranged from 0% (for butterfly) to 44% (for sleep), a finding that parallels the likelihood that the critical lure appears as a response in a free association task. Lists containing words more likely to elicit the critical lure in a free association task (higher average frequency), were more likely to elicit false recall of the critical lure in a false recall task. In fact, the correlation between false recall and mean association value was .873. In the context of Bartlett's (1932) reconstruction-reproduction hypothesis, Deese's results showed that elaboration in memory can occur in a simple list learning paradigm. In this instance, elaboration involved the addition to the event memory of items associated with the study words. Although this study potentially provided a new paradigm that could easily elicit false recall effects, it remained largely unnoticed for many years (Bruce & Winograd, 1998), until Roediger and McDermott (1995) replicated and extended Deese's findings.

Roediger and McDermott (1995) reproduced Deese's (1959) findings using six lists that had produced high levels of false recall in the original Deese experiment. Again, the 12 highest associates for each critical lure were included in each list. The presentation of each list was followed by an immediate free recall test, and after the presentation of all six lists the subjects were given a recognition test. Overall, mean probability of recall for a studied word
was .65, with marked recency and primacy effects. The critical lure was recalled with a mean probability of .40, a probability comparable to recall of a word presented in the middle of the study list. Other non-related intrusions occurred with a probability of .14, indicating that guessing probably could not account for the high percentage of subjects recalling critical lures. Extending Deese's results to a recognition test, Roediger and McDermott found that the critical lures were called old at about the same rate that the studied items were called old, a finding that parallels Underwood's (1965) results using converging associates. Again, the false alarm rate for critical lures was much higher than the false alarm rate for other weakly related lures and unrelated lures.

Extending the findings of their first experiment, Roediger and McDermott designed a second experiment that included an expanded version of the lists (15 words each instead of 12) and a condition in which the false recognition test did not follow a recall test, permitting them to assess the effect of repeated testing. They also sought to determine the false alarm rates for critical lures when the associated list was not presented (which is not higher than the false alarm rate for other unrelated lures). Finally, they explored subjects' judgments about their phenomenological experience while recognizing the critical lures, using Tulving's (1985) remember-know procedure. In this procedure, subjects are asked to judge each item as old or new, and for each item judged to be old they are asked to elaborate by indicating whether they remembered or just knew that the item was presented during the study trial. An item is said to be remembered if the subject can mentally re-experience the presentation of the word, perhaps by remembering where in the list the word was presented, the word it preceded or followed, what they thought of when they heard the word, what they were doing when they heard the word, or any physical/perceptual characteristics associated with the
presentation of the word. In contrast, a know judgment means that the subject believes the item was presented but cannot mentally relive the experience. In general, know judgments are more common than remember judgments. Although controversial, there is some evidence that remember-know judgments do not merely represent differences in familiarity criterion, but actually represent two different components of recognition memory (Rajaram, 1993). It is worth noting, however, that not everyone advocates use of the remember-know procedure as an accurate way of assessing memory experiences because of the potential problem of familiarity criterion (see, e.g., Chandler, 1994).

Overall, Roediger and McDermott's (1995) results from Experiment 2 showed an even more dramatic rate of false recall than in Experiment 1. The probability of falsely recalling the critical lure was .55 (compared to .40 in Experiment 1), a difference that may be attributable to the expanded list length or to the expanded number of lists (6 to 16). The probability of recall of the critical lure was significantly greater, although marginally, than the probability of recalling a word from the middle of the list (.47 compared to .55). False recognition after a free recall test was greater than false recognition that did not follow a free recall test (followed arithmetic problems instead), indicating that prior free recall enhances the effect in later recognition. Remember judgments for critical lures were also higher in the recall-then-recognition condition (.72) compared to the arithmetic-then-recognition condition (.53), although remember judgments for studied items were approximately the same across the two conditions. Their high rate of remember judgments indicates that there is not just a feeling of familiarity for the critical lure, but that some vivid memory exists for the experience of the presentation of the critical lure that was in fact never presented.

Roediger and McDermott (1995) found powerful false recognition and false recall
effects that parallel effects found using prose or other complex materials. This has important implications for Bartlett's reconstructive-reproductive memory distinction. The results of Deese (1959) and Roediger and McDermott indicate that all memory is constructive. Roediger and McDermott propose that differences in false recall effects between paradigms are quantitative, not qualitative, in that some paradigms are more likely to elicit stronger false memories than others are. Replication of Deese's findings indicated that serial list-learning of word associates could be a very powerful method to study false memory. Part of the power of this paradigm is its usefulness in studying false free recall, rather than always employing measures of recognition memory or cued recall.

**Further Investigations of the False Memory Paradigm**

Numerous studies have replicated Roediger and McDermott's 1995 findings (e.g., McDermott, 1996; Payne, Elie, Blackwell, & Neuschatz, 1996; Read, 1996; Schacter, Verfaellie, & Pradere, 1996) and have extended these findings by further testing the robustness of the Deese (1959) serial word-list paradigm. Experiments using this paradigm have focused on manipulating a variety of different independent variables and observing their effects on rates of false recall and false recognition. Some of these studies have looked at the phenomenological aspect of the false memories. Of course, when examining phenomenological data derived from subject self-reports, caution should be used: Such reports seem quite powerful, but there is often a vast difference between what subjects report is happening and what really is happening. In fact, subjects can be surprisingly poor at describing what they are experiencing (Nisbett & Wilson, 1977). However, subjects can also provide valuable insight into what they are thinking and feeling, and this should not be disregarded (e.g., Davison, Navarre, & Vogel, 1995; Ericsson & Simon, 1980; Hayes, 1989).
By asking subjects to rate their responses, and later to indicate whether they remembered the events, Roediger and McDermott (1995) showed that not only are subjects confident that the critical lures were presented at study, but they also claim to actually remember the presentation of the critical lures with high frequency. The fact that subjects assign remember, rather than know, judgments to the critical lures indicates that the words not only seem familiar to the subjects, but that they have a memory of the word being presented at study. Examining the nature of this finding further, Payne et al. (1996, Experiment 3) found that when words were presented visually by two different sources, subjects were quite willing to make a source attribution for the unpresented critical lure 87% of the time. This finding is particularly dramatic given that the subjects were explicitly cautioned about the difficulty of the task and were instructed not to make guesses about the nature of the source. Apparently, the effect was so compelling that some subjects were unwilling to believe the word had never been presented, even after the videotape was replayed back to them. Accuracy checks indicated that when the associated list was presented by a particular source (e.g., the female) the critical lure was judged as having been spoken by that confederate only 53% of the time. Lack of accuracy, however, did not change the participants' strong assertion of remembering the source.

Bransford and Franks (1971) examined confidence ratings in their earlier study and found that subjects were more confident at recognizing sentences that contained the most semantic relations, even though these sentences contained more semantic relations than did any of the actually studied sentences. In fact, the fewer semantic associations the sentence contained, the less confident the subjects were that the sentence had been presented. Holmes, Waters, and Rajaram (1998) replicated this finding and also showed higher
remember false alarms for sentences with more semantic relations or episodic content. Taken together, these results indicate that subjects have strong feelings for false memories, and that in some instances they are more confident of their false memories than their true ones. Accuracy, it seems, plays a very small role in confidence. Subjects are often certain of their memories even when those memories contain dramatic and numerous errors (Neisser & Harsch, 1992). However, it is important to note that although the critical lures do elicit strong phenomenological feelings and high levels of confidence that these feelings are memories, there are differences in the qualitative and neurological experience of studied words and critical lures.

Mather, Henkel, and Johnson (1997) postulate that remember judgments are assigned to critical lures because of the vivid qualities associated with the word, but they warn that these vivid qualities differ between true and false memories. Specifically, using a memory characteristics questionnaire, they found that true memories were associated with greater auditory detail and more feelings and reactions. However, subjects still falsely recognized critical lures at a high rate, indicating that semantic information may over-ride other phenomenal details. Norman and Schacter (1997) also explored the characteristics of false memories. They asked subjects to elaborate on remember judgments by writing down exactly what they remembered about the words’ presentation. Although asking for elaboration did not reduce the rate of remember false alarms, there was a difference in the details used to describe real versus false memories. In general, subjects tended to describe memories of critical lures in terms of related words or thoughts associated with related words. In contrast, memories for studied words were more likely to refer to the circumstances surrounding the words’ presentation at study, such as emotional reactions,
context, or sensory characteristics of the word. A second experiment provided quantitative evidence for this finding using yes/no questions designed to determine the nature of the memory for the critical lure and for studied words. However, given the conviction with which subjects assert their memories as real, it seems that elaboration has little influence on the phenomenological experience of false memories, despite the differences noted.

Johnson, Nolde, Mather, Kounios, Schacter, and Curran (1997) caution that the phenomenological similarity of true and false memories depends on the test or encoding conditions examined. In other words, under certain conditions true and false memories are likely to seem more similar than under other conditions. Support for this assertion comes from ERP studies showing that brain waves for true and false memories depend on the test condition (similar for the random condition, different for the blocked condition). It seems that distinguishing the two memories based on phenomenological reports may be a difficult task. A second method of distinguishing them may be to look at differences and similarities in their effects on other measures of performance (e.g., priming).

The false memory effect can lead to memories so similar to those of the word actually studied that they even generate priming on perceptual implicit memory tests (e.g., word stem completion, word fragment completion), although to a lesser degree than do studied words. McDermott (1997) showed that priming on a conceptual implicit test of word association was equivalent for critical lures to priming associated with presented words. Perceptual priming effects were found, however the level of priming was lower for the critical lures than for studied words. This finding indicates that conscious awareness of the lure occurs during study, a finding that seems consistent with other phenomenological evidence such as source attributions and remember judgments. Interestingly, although the critical lure behaves like a
studied word in priming tasks, the critical lure does not always behave as it would if it had been studied. When a critical lure is included as a study item, recall of that item is double the recall of the same item when just the associative list is presented without the critical lure, although the conditions produce identical levels of conceptual priming (McDermott, 1997).

The Persistence of False Memories

False memories created using the DRM paradigm have been shown to be very resistant to experimental manipulations designed to eliminate the effect. Test delays of up to two days do not affect probabilities of false recall; in fact, false recall rates actually exceeded hit rates after a two day delay (McDermott, 1996, Experiment 1). Similarly, a 24-hour delay on recognition tests had no effect on false recognition rates, but hit rates (recognition of previously studied items) decreased significantly over the interval (Payne et al., 1996, Experiment 1). The subjects’ reported experience of recollecting the critical lures also did not decrease over a 24-hour interval (Payne et al., 1996, Experiment 1). Posner and Keele (1970) asserted in their study that loss of specific instances, despite maintenance of the general schema over time, indicates that abstraction of the general schema occurs during encoding and not at retrieval, a notion consistent with false recall and recognition in the DRM paradigm.

Successive test trials also fail to diminish the effect, with higher rates of false recognition following a free recall test (Roediger & McDermott, 1995), and rates of false recall increasing across repeated recall tests following a single study trial (McDermott, 1996, Experiment 1; Payne et al., Experiment 2). The persistence of the effect across test trials was clearly not the result of a criterion shift, because similar results were obtained in free and forced recall conditions over successive test trials (Payne et al., 1996, Experiment 2). If an
increase in false recall over test trials was due to a criterion shift, then the effect would only be obtained in a free recall condition because in a forced recall condition the criterion is assumed to remain unchanged (see also Erdelyi, Finks, & Feigin-Pfau, 1989, Experiment 1). In contrast, multiple study-test trials can reduce false recall of critical lures, but the effect is not eliminated (McDermott, 1996, Experiment 2). McDermott reduced false recall from a probability of .57 to .32 or from .30 to .20 (depending on condition) over five study-test trials, but these probabilities were still higher than those for unrelated intrusions, although not dramatically so.

Closely related to the number of tests or study trials is the finding that false recognition increases and then decreases as a function of the number of times a particular word associate is presented during the test trial (Hall & Kozloff, 1970). Specifically, false recognition increased up to the third presentation of the critical stimulus word, then declined after five and seven presentations (the maximum tested), explained in terms of a two-stage decision making model focusing on the perceived situational frequency of the word. In other words, each subject has a frequency criterion and words that meet this criterion are judged old, whereas words that do not meet this criterion are judged new. However, varying word repetition does not always influence false recognition (Shiffrin, Huber, & Marinelli, 1995; Tussing & Greene, 1997). Shiffrin et al. found that false alarms were related to the number of associates presented in a study trial and not to the number of times each associate was presented. In a second study by Hall and Kozloff (1973), comparing multiple presentations of the same associate with a single presentation of more than one converging associate, they showed there was a higher rate of false recognition in the converging associate condition compared to the multiple presentation condition.
In general then, the strength of the false memory effect is not only related to the number of test trials, study trials, or presentations of a particular word, but is tied to the number of associates within a given study trial. Overall, the more associates for a given word, the stronger the false memory effect for that word serving as a critical lure. Underwood (1965) found that an antonym repeated three times produced a false recognition effect for its implicit associative response, but the same antonym presented only once did not. Roediger and McDermott (1995) found a stronger false recall effect in Experiment 2, which used 15 associates, than in Experiment 1, which used only 12 associates. Payne et al. (1996, Experiment 3) hypothesized that their lower than usual false recall rate (33%) may have been because they used only 8 word associates per list, rather than the usual 10-15. In their first experiment, using 15 associates, they obtained a false recall rate of 45%. Taken together, these results imply a possible influence of the number of associates on false recall rates.

Robinson and Roediger (1997) tested this hypothesis by manipulating the number of associates per word list (holding list length constant using filler words, and without holding list length constant). They found that false recall and false recognition were tied to the total number of associates, not the mean strength of association (mean strength is stronger for shorter lists because only the most highly associated words are included). In general, as the number of associates increased, the false recall and false recognition rates increased, from .03 for a 3-word list to .31 for a 15-word list. The use of filler words to equate list length in Experiment 2 did not reduce false recall or false recognition rates, but did depress overall recall of studied words (the list-length effect). For discussions of list-strength effects, which refers to the strength of items contained in a list (in this case strength refers to associative
power) see Ratcliff, Clark, and Shiffrin (1990a; 1990b), and Shiffrin et al. (1995).

On a similar note, how the words are encoded can be just as important as how they are presented. Manipulating the level of processing of associate word lists yields fairly robust false memory effects across all levels of processing (Read, 1996; Tussing & Greene, 1997). Read found no difference between elaborative rehearsal (.73) and maintenance rehearsal groups (.76), although both were significantly higher than the serial learning group (.50), which still produced a dramatic false recall effect. Tussing and Greene extended Read's work and found that false recognition was significantly lower in all three of their incidental learning conditions (letter-counting, vowel-judgment, and pleasantness rating) compared to baseline measures (established using a standard recognition test following study of the word lists), however no significant difference was found between the different levels of processing. They hypothesized that rehearsal may play an important role in false memory, or that the processing tasks served as a distraction from the processes involved in false memory formation; however, firm conclusions were not possible without further analysis. Overall, false memory does seem to be robust across levels of processing manipulations, but greater effects in elaborative and maintenance rehearsal groups indicates that semantic analysis may increase the false memory effect.

Can False Memories Be Eliminated?

In general, false memories have been shown to be very compelling, robust, enduring phenomena. In fact, of all the experimental manipulations focusing on false recall in the DRM list learning paradigm, none has successfully eliminated the effect. There are, however, some notable instances of dramatic reductions in false recall and false recognition in some experimental paradigms. Studies reviewed earlier indicated a diminished effect of
false memory in serial learning (memorizing word order) tasks (Read, 1996), over multiple presentations of the study word (Hall & Kozloff, 1970; McDermott, 1996; Tussing & Greene, 1997), with a small number of converging associates or short lists (Robinson & Roediger, 1997), or where there was a one-way relation between associates (Anisfeld & Knapp, 1968; Underwood, 1965).

Perhaps one of the most compelling and widely replicated findings of diminished effect is the significant difference in false recall when word lists are blocked (the associates are all grouped together, with different lists presented consecutively), compared to random (in which word associates from several lists are presented in a mixed, or random, order). Without exception, higher levels of false recall and false recognition are obtained following blocked presentation of words compared to random presentation (Mather et al., 1997; McDermott, 1996; Tussing & Greene, 1997). Mather et al. also found that false memories created in the blocked condition were more resistant to scrutiny, and therefore were more likely to be labeled old in an old-new judgment, even when asked to elaborate on old decisions. In blocked conditions, source attribution judgments for study words are more accurate than in random conditions. Similarly, source attributions for critical lures are more accurate, in the sense that the source is assumed to be the source that presented all of the word associates in the corresponding list (Mather et al., 1997).

The evidence also indicates that there are differences in false memories created in random and blocked conditions, with those in the blocked condition assuming more characteristics of true memories. Therefore, not only are more false alarms characteristic of blocked presentation, but these memory errors are also more compelling and real in subjective terms. Given that the probability of eliciting the critical lure should be the same
for each word associate independent of the words presented before or after that particular word associate, we might, a priori, expect no difference between random and blocked presentation. The fact that there is a difference emphasizes the importance of a semantic connection between study words above and beyond activation of the critical lure for each individual word. Semantic links between words seem to play an important role in false recall and false recognition of the critical lure.

False recognition is also reduced when distinctive perceptual details of items, rather than just semantic details, are encoded at study. Using line drawings to represent word associates, Israel and Schacter (1997) found a marked reduction in false recognition rates compared to rates obtained using words during encoding (both words and pictures were accompanied with an auditory label to eliminate confusion over the nature of the drawings). Presenting pictures improves memory for item specific details, resulting in fewer conceptual errors for semantically related words. In other words, increasing focus on perceptual details seems to diminish emphasis on semantic details, thus reducing false recognition of semantically related associates. However, although reduced, the false recognition effect was not eliminated in this study: Significant false recognition rates still occurred after pictorial encoding.

Similarly, presentation modality can affect false memory rates. Visual presentation of words reduced the rate of false recall and false recognition compared to auditory presentation (Smith & Hunt, 1998), just as pictorial encoding reduced false recognition. Specifically, in the visual condition, false recognition and false recall rates were well below the hit rates for studied items, whereas in the auditory condition the hit rates for presented words and the false alarm rates for critical lures were equivalent. Smith and Hunt propose
that visual presentation allows for a better means of discriminating true from false memories than auditory presentation allows, resulting in a decrease in false memory rates when words are presented visually. In other words, visual encoding allows for better discrimination at test than does auditory encoding, perhaps because visual presentation of words (or pictures) makes those words more distinctive and therefore easier to recognize as old.

Perceptual information and visual presentation can reduce false memory, as can forewarning subjects about the memory illusion. Gallo, Roberts, and Seamon (1997) warned subjects about falsely recognizing critical lures (including a false recognition demonstration), urged subjects to minimize errors, or gave no warning at all (except control instructions). The proportion of falsely recognized critical lures diminished as a result of forewarning (.81 in uninformed group, .74 in cautious group, and .46 in forewarned group), although the effect was not eliminated and false recognition for critical lures was still significantly higher than for unrelated lures (.46 versus .14). This effect was replicated in an in-class demonstration by the same authors. A post-experiment questionnaire revealed that the instructions at tests (forewarning, cautioning, uninformed) influenced the type of retrieval strategy used, with subjects in the forewarned group attempting to identify the critical lures. However, even though the forewarned group tried to determine the critical lures, they were not inoculated against the false memory effect.

Using a modified DRM paradigm, McDermott and Roediger (1998) asked subjects to determine whether the critical lure was present in the word list (it was in fact present in half of the lists). Some subjects were told about the false recognition phenomenon and were given an example to illustrate the effect and other subjects were given no warning at all. Even when subjects were warned in advance, they still could not accurately identify the
presence or absence of the critical lure. However, the warning instruction did attenuate the false recognition effect compared to the no-warning condition. McDermott and Roediger found that subjects had difficulty performing even under conditions of forewarning, stating that instructions had little (but some) effect on performance.

In contrast, Jacoby and Whitehouse (1989) found a significant effect of warning on false recognition by presenting context words prior to presentation of the test word. The subjects were either told of the presence of the context word (aware condition - conscious perception) or were not told about the context word (unaware condition - unconscious perception). The context word either matched (same) or did not match (different) the test word. In the unaware condition, unconscious perception of a context word increased false recognition if the context word and the test word matched, and decreased recognition if the context word and the test word did not match. In contrast, when the subjects were informed about the presence of context words, they were more likely to disregard a matching context word as opposed to an unmatched context word. In this study, knowledge of the context word helped subjects to disregard matched context words.

Knowledge can also reduce false recognition for semantically related words when information about the study word is accessible during the recognition test, dubbed the false-recognition reversal effect (Brainerd, Reyna, & Kneer, 1995). Specifically, under certain conditions (such as an increased level of repetition or an increased age of subjects), related distractors (including associates, category names, rhymes, and same-category exemplars) are easier to reject than unrelated distractors. This finding of false-recognition reversal proves to be a very robust phenomenon (Brainerd et al., 1995).

On a related note, knowledge of the list theme may have a profound impact on false
memory. Hasher and Griffin (1978) found that when giving prose passages to subjects, if they changed the label given to the theme of the story at recall from the theme given at study (prose passages were devised so that either theme was a suitable label), subjects relied less on the theme to generate recall of the story. Consequently, those in an inconsistent theme condition were less likely to erroneously include inaccurate details (memory errors) in their story recall. Again, this demonstrates how profoundly knowledge can influence false memory effects. On a similar note, the same passage, given two different titles, elicits two different patterns of recall, patterns consistent with the theme established by the title given (Bransford & Johnson, 1972). Furthermore, in 1962 Tulving demonstrated that during free recall subjects would spontaneously organize lists of randomly presented words into groups based on theme. When given a theme (or when subjects generate their own theme), subjects rely on knowledge of that theme to aid in recall. In the context of the DRM paradigm, knowledge of the theme may aid in recall of studied words, and in the false recall of non-studied theme-related words such as the critical lure (especially given that the theme of the list can often be defined in terms of the critical lure).

Age also appears to be an important variable in false memory. Brainerd et al. (1995) found false recognition reversal for older children and young adults, but not for younger children. Younger children still tended to accept related distractors at a higher rate than unrelated distractors, although they did exhibit reversal when manipulations increased use of verbatim retrieval or when online interference (i.e., talker variability) was reduced. Other investigations of older versus younger adults have revealed that older adults may be more susceptible to false memory effects than younger adults and that the qualitative difference in true and false memories is somewhat less pronounced in older subjects (Norman & Schacter,
As well, older adults appear to be less able than younger adults to suppress false memory after multiple study trials, consistent with the idea that older adults rely more on gist-based information rather than on detailed information of individual items (Kensinger & Schacter, 1999). However, older adults do show suppression of false recognition after pictorial encoding, just as younger adults do (Schacter, Israel, & Racine, 1999). Note, though, that despite suppression after pictorial encoding, both older and younger adults often still falsely recognize critical lures, and the effect is still more pronounced in older adults (Koutstaal & Schacter, 1997). Although no one age group is immune to false memory effects, it does appear that susceptibility increases for particular cohorts (see Schacter, Norman, & Koutstaal, 1998, for a review).

False recall and false recognition effects are also diminished in certain patient populations, most notably for amnesic patients (Schacter et al., 1996). Using the DRM paradigm, Schacter et al. found that amnesic patients had lower levels of false recognition for critical lures, but higher levels for unrelated lures. In an extension of this study, amnesic patients still showed less false recognition than did controls for semantically related words, and also for perceptually related words (Schacter, Verfaellie, & Anes, 1997). Implicated in the different patterns of false recognition between controls and amnesic patients is damage to the temporal lobes associated with amnesia. Temporal lobes may play a role in storing or retrieving the semantic information associated with false memories (see Schacter, 1997, for a review). In general, recent cognitive neuropsychological research has helped to advance understanding of false memory by focusing on the differences and similarities in the processing of true and false memories (Schacter et al., 1998). For instance, Schacter et al. (1996) found differences in blood flow between veridical and illusory recognitions, with an
additional region of increased blood flow for veridical memories. Neurological work with patients and neuroimaging studies on non-patient populations is opening another door to aid in the understanding of false memories created using the DRM paradigm.

Age, knowledge, and encoding conditions all influence false memory to varying degrees. However, despite numerous experimental manipulations trying to show otherwise, false recall and false recognition as shown in the DRM paradigm (and in other false memory paradigms) has proven to be relatively impervious to elimination. Of particular note is the robustness of false recall and false recognition in the DRM serial word list paradigm, given that list learning promotes verbatim recall, the word lists and retention intervals are generally short, guessing is discouraged, and free recall tests are usually found to be impervious to errors of intrusion (McDermott, 1996; Roediger, 1996). How can one explain the persistence of false memories? There are several possible sources of the powerful and robust effect of false memory for critical lures in the DRM paradigm.

Why Do False Memories Arise and Persist?

Underwood (1965) explained the false memory effect in terms of implicit associative responses (IAR), assumed to be either the most frequently produced associate for a given word or a word learned as a response to a given stimulus. Underwood proposed that the IAR is unintentionally, but consciously, activated during study and may be encoded during study along with the actual studied words. At the time of recall, subjects may suffer a type of source confusion or reality monitoring error (defined as the inability to distinguish between true memories and imagined events; see, e.g., Johnson, Hashtroudi, & Lindsay, 1993; Johnson & Raye, 1981) for the critical lure, unsure of whether they heard or thought about the word, and as a result the critical lure behaves in much the same way as a studied word.
Another interpretation of Underwood’s theory could be that activation of the critical lure is truly implicit, but given the vividness of phenomenological memories for the critical lure, it seems unlikely an implicit model based on spreading activation could elicit such compelling recollections (Roediger & McDermott, 1995). Although Roediger and McDermott do not apply a specific model to their data, they do assert that associative models can explain false recall and false recognition effects.

Another conceptual framework used to account for false memory effects in the DRM paradigm is the fuzzy trace account (Brainerd et al., 1993; Reyna, 1998). In this view, memory can be based either on a gist trace that preserves the meaning of an event, or on a verbatim trace that preserves specific information about the event. Both traces, gist and verbatim, are assumed to be encoded in parallel (but independently) and can be retrieved independently. The two representations are assumed to differ in specificity and in forgetting rates (with gist memories forgotten more slowly). In the case of the DRM paradigm, the event is the word association list, the gist trace would be assumed to preserve the meaning or theme of the list (often defined in terms of the critical lure), and the verbatim trace would preserve information about each word contained in the list. For a brief review of the application of fuzzy trace theory to false recall and false recognition studies, see Payne et al. (1996).

Similar to the idea of fuzzy trace theory is the notion that memory for stimulus parts is independent of memory for the stimulus whole, or global structure (Reinitz, Verfaellie, & Milberg, 1996). In this view, memory distortions are seen as binding problems, defined as an inability to form associations between elements (Reinitz et al., 1996) that interferes with long term memory processes (Kroll, Knight, Metcalfe, Wolf, & Tulving, 1996). Looking at
recombined stimuli (stimuli formed by combining two previously presented stimuli), Reinitz et al. (1996) demonstrated that people would often falsely recognize the recombined stimulus. More interesting is the finding that amnesics with structural damage to areas presumed to perform the binding function are more likely to falsely recognize recombined lures than are normals (Kroll et al., 1996; Reinitz et al., 1996). In other words, subjects who could not establish the appropriate relations between the stimulus parts were more susceptible to false recognition. Binding is closely related to fuzzy trace theory in that the gist representation is essentially the binding of different events into a coherent whole. In fact, the two ideas are not incompatible and appear to describe essentially the same event using different terminology.

Attributional theories, similar to source monitoring approaches, are also applied to the false recall - false recognition phenomena. Attribution theory, as advocated by Jacoby and colleagues (e.g., Kelley & Jacoby, 1996) focuses on the relation between fluency of memory retrieval and the probability of believing that the memory is real (the easier that the event comes to mind, the more likely it is to be judged real). Similarly, the more vivid the memory, the more likely it is to be judged real. In terms of the word association paradigm, because critical lures come to mind readily and are accompanied by vivid memories, they are judged as old on recognition tests and recalled with high frequency on recall tests.

The strength of these memory theories is that they are based on constructiveness accounts of memory. Although Bartlett (1932) did not conceive of serial list learning as a constructive memory process, Deese (1959) and Roediger and McDermott (1995) have shown otherwise. Given the nature of false memory findings, the evidence for dissociations between false and true memories, and the persistence of false memories (even in
straightforward list-learning paradigms), it is evident that veridical views of memory that treat memory errors as simple retrieval or storage failures are not adequate to account for false memory findings. However, there is controversy over which of the constructivist theories best explains the different demonstrations of false memory. Although that question will not specifically be addressed in this study, the results will be analyzed with respect to the differing views. Furthermore, given the breadth of memory illusions, it is unlikely that a single theory can account for all of the different types of memory illusions, including the diversity of findings within each type of illusion. Schacter et al. (1998) go even further and assert that it would be extremely difficult to “tease apart implicit associative response and pattern separation failure [fuzzy trace theory] accounts of false recognition” (p. 296). Although not parsimonious, it does seem probable that future studies trying to uncover the explanation behind false memory phenomena will conclude that multiple explanations are necessary to account for the wide range of false memory phenomena cited in the literature.

**Logic of the Present Study**

Based on the past research, it seems that there are several givens in the list-learning false memory effect. It seems apparent that this effect found in the DRM paradigm is dependent on either the formation of a critical word associate for the study list as a whole (assumed to be the critical lure) or the formation of some sort of memory representation (conscious or unconscious) for the general theme of the list (a gist representation). Either way, a semantic association is made between the words in the list. Supporting this notion is the finding that subjects tend to recall words in clusters organized by theme, when multiple lists are presented in one study trial (Payne et al., 1996; see also Bousfield, 1953; Tulving, 1962). Organizing lists indicates that subjects may abstract the theme of the list and use this
theme as a retrieval cue. This semantic link is assumed to be crucial for the production of the critical lure on free recall tests and for the recognition of the critical lure on recognition tests.

It follows that by preventing subjects from forming implicit associative responses with the critical lure or from encoding the semantic gist of the word lists, it may be possible to eliminate, or greatly reduce, the false recall effect. Underwood (1965) stated that IARs can be the result of either generated associations to the stimulus word or learned relations between the stimulus word and another response (as seen in paired-associate tasks). Based on this logic, the present study will attempt to eliminate the semantic link between the word associates in the DRM paradigm by explicitly giving subjects another link to learn, using a paired associates task (for example, linking the word associate tired with a random two-digit number).

Evidence for the importance of semantic information can be found in studies with amnesic patients. Schacter et al. (1996) hypothesize that patients are less susceptible to false memory than are controls because patients fail to retain the associative or semantic information that links the word associations with the critical lure. Similarly, conditions that increase the chance that the word associates will be encoded in terms of a common theme - such as semantic analysis of items, blocked presentation, or an increased number of associates - also increase the number of critical false recalls and false recognitions (Mather et al., 1997). In fact, when subjects are instructed to make recognition judgments based on consistency with list themes (rather than on actual presentation of the word) critical distractors are accepted at higher rates than are the actually presented words (Brainerd & Reyna, 1998). In keeping with this logic, conditions that decrease the chance that
information will be encoded in terms of a general theme should decrease the number of false
recalls and recognitions. The present experiment is designed to test this very question, by
inducing conditions in which the probability that a subject will encode items in terms of a
general schema is greatly reduced.

Specifically, in this study we attempted to reduce the probability that a subject would
form a semantic link between the study words (the word associates) by explicitly providing
them with a paired associate to learn. In other words, each word associate was matched with
either a low association number or another word associate from the same list. Subjects in the
control group, who were not given an explicit associate for each study word, were expected
to perform in a similar manner to subjects in Roediger and McDermott (1995). Simply, the
standard group was expected to demonstrate marked false recall and false recognition effects.
In contrast, the first experimental group was presented with each word associate linked to a
low association number. Because the subjects were told to focus on learning the number-
word pair, the expectation was that the focus on learning the pair would decrease semantic
analysis of the word associate and would consequently reduce false recall and false
recognition effects. Tussing and Greene (1997) hypothesized that incidental learning may
have decreased false memory effects in their experiment because the incidental acts that the
subjects engaged in distracted them enough to reduce semantic analysis. The same theory
may hold true in this experiment, with learning of the number-noun pairs serving as the
distractor (even though the learning was not incidental in this case).

In the second experimental group, subjects learned the word associates in pairs.
Again, the focus was on learning the pairs, not just the individual words. There were two
possible outcomes of this manipulation. False recall and false recognition might have
remained at standard group levels or even increased because the semantic link between the word associates was made more obvious. Subjects that may not have encoded the words in terms of a general semantic link might have been more likely to do so when the relation was made explicit. However, it is also conceivable that false recall and false recognition would decrease because subjects were already being provided with an appropriate associative response for each item, making further semantic analysis that would lead to generation of the critical lure unlikely. Underwood (1965) stated that implicit associative responses could either be generated by the subjects or provided by the experimenter, as in a paired associates learning task. If subjects still generated an associative response to the pairs of word associates, despite already having been provided with an explicit link, Underwood’s two types of associative responses may not be mutually exclusive, in the sense that providing a response undermines generation of another associative response. However, if subjects failed to generate a second associative response, this may suggest that the two conditions are mutually exclusive.

In general, this experiment is designed to test the hypothesis that false recall and false recognition effects in the DRM paradigm are dependent on the formation of a semantic link between word associates. The general prediction is that diminishing semantic analysis of the word associates will decrease false memory effects (as demonstrated in the number-noun paired associated task), whereas making the semantic link more explicit could potentially increase false memory effects (as seen in the pairing of word associates) compared to the standard group.

The practical implications of this research are controversial. Freyd and Gleaves (1996) caution against generalizing false memory findings using laboratory paradigms, such
as the DRM paradigm, to current false memory controversies such as recovered memories. They indicate that the nature of the material (non-emotional, non-sexual) and the difference between laboratory findings and real-life events limit the appropriateness of generalizing these results to other applied areas of memory research. However, as Roediger and McDermott (1996) indicate, the ability to create such a powerful memory illusion with great ease can generate research that will allow us to better understand how human memory works. Applied uses of this paradigm have already begun (e.g., Schacter et al., 1996; 1998) and potentially great insight may be gained upon closer study of this memory phenomenon, just as research focusing on perceptual illusions has led to growth in understanding of our perceptual processes.

Method

Participants.

Participants were 531 University of Toronto at Scarborough undergraduates participating for bonus credit in their introductory psychology course. There were 190 participants in the standard group, 183 participants in the number-noun group, and 158 participants in the noun-noun group.

Design and Experimental Control.

The study was a between-subjects-after-only design in which subjects were assigned to one of three conditions (standard, number-noun pairs, or noun-noun pairs). The standard group was merely presented with each word associate one at a time (e.g., bed, rest, awake, tired ...), as in the reviewed literature. The number-noun group was presented with each word paired with a two-digit number of low association value (e.g., 17-bed, 26-rest, 77-awake, 55-tired ...), and the noun-noun group was presented with six pairs of words a (e.g.,
mented. The low-association two-digit numbers were taken from MacLeod (1988).

Materials.

Two study lists were used (sleep and needle), each composed of the top twelve associates for that particular list. The word associates were taken from the lists modified by Roediger and McDermott (1995, Appendix A). The needle list consisted of the associates thread, pin, eye, sewing, sharp, point, prick, thimble, haystack, thorn, hurt, and injection. The sleep list consisted of the words bed, rest, awake, tired, dream, wake, snore, blanket, doze, slumber, snooze, and nap. The lists and paired associates were presented auditorily, one at a time. The presentation order of the words and paired associates was constant for all subjects, with the strongest word associates for the critical lure presented first, following common practice for this procedure (Roediger & McDermott, 1995). The sleep list always preceded the needle list in presentation.

The response booklets contained materials for the assessment tests (recall and recognition only for the standard group; recall, recognition and cued-recall for the experimental groups). The free recall test simply required the participants to indicate their responses on a blank sheet of paper. The participants were each provided with two sheets of blank paper for this purpose. The recognition test consisted of studied words from position one, three, five, seven, nine, and eleven, plus the critical lure, two weakly related lures and two unrelated lures, for a total of 11 items. Following Roediger and McDermott (1995), the weakly related lures were drawn from position 13 or lower in the word association norms. The paired-associates test took the form of a cued-recall test in which the subjects were provided with the stimulus component for each paired associate and were asked to provide
the response component.

Procedure.

Participants were tested in groups during a session lasting approximately 20 minutes. Initially they were told they would be hearing lists of words that would be used in a later test of memory. Subjects in the two paired-associate groups were instructed that their memory for the pairs would be tested later in the experiment. Instructions to these subjects emphasized the importance of learning the number-noun or noun-noun pairs, rather than just learning the individual words (see Appendix A for complete instructions). Each list was presented only once. The rate of presentation for the words in the standard condition was approximately one word every 1.5 s., the pairs were presented at a rate of approximately one pair every 2 s. The presentation rate of the number-noun pairs and the noun-noun pairs was determined in a pilot study in which presentation time was manipulated to equate level of recall for studied words across the three groups.

In general, the participants were presented with the study list, following which they completed a series of tests (recall, recognition, and cued-recall, if appropriate). After completion of the tests, the entire process was repeated with a second list. Specifically, following presentation of each list, all subjects were asked to recall the words previously presented, in any order they wished. Subjects in the number-noun group were told not to worry about recalling the pairs, but to focus on recalling the words in any order, regardless of what number the word was paired with. Subjects in the noun-noun condition were told to recall as many words as they could, without worrying about recalling the words in pairs.

Following the free recall test, which had a three-minute time limit, the subjects were given the recognition test (see Appendix B). The subjects were told that they would see a
series of words on a sheet and they were to circle the words they recognized from the earlier study list. Subjects were given a three-minute time limit for completion of this task. The recognition test for all groups was identical.

Finally, following the recognition test, the two paired-associate groups were given a paired-associate test (see Appendices C and D). The subjects were asked to recall each studied stimulus-response pair, writing down the pairs exactly as they were studied. Subjects were provided with the stimulus and were asked to provide the response.

For each test given (recall, recognition, cued-recall if appropriate), subjects were reminded of the importance of only writing down/circling words from the study list; they were instructed not to guess. Upon completion of the second set of tests, subjects were debriefed as to the nature of the experiment and the experimental session ended.

Results

An alpha level of .05 was adopted for all statistical tests. All data were included in the analyses. All post-hoc analyses were performed using Tukey’s LSD procedure.

Free Recall Test

Correct Recall

For each subject, a recall score was generated as the proportion of words correctly recalled from both presented lists (a perfect score of 1.0 reflects the correct recall of all 24 presented words, 12 words from each list). Recall scores were also broken down based on the words’ position within the list, from the first third, the second third, or the final third. In general, higher proportions reflect better recall of the presented words. An analysis of variance showed a significant difference in overall correct recall between groups F(2, 528)=32.30, MSE=.01648. Pairwise comparisons revealed significant differences between
the groups, with the standard group demonstrating the highest levels of correct recall at .50, followed by the noun-noun group at .47, and then the number-noun group at .39 (see Table 1).

An analysis of recall by word position showed a significant difference between groups in recall from the first third of the word lists, $F(2, 528) = 72.37$, $MSE = .04089$. Again, the standard group recalled the greatest proportion of words from the first third of the lists, followed by the noun-noun group, and then the number-noun group (see Table 2). The difference in correct recall from the middle third approached significance, $F(2, 528) = 2.85$, $MSE = .03861$, $p = .057$. Although not definitive, post hoc analyses revealed significant difference favouring the standard group over the number-noun group and the noun-noun group over the number-noun group. The difference between the standard group and the noun-noun group was not significant (see Table 2). Finally, an analysis of variance on recall from the final third revealed a significant overall difference among the groups, $F(2, 528) = 4.53$, $MSE = .03791$, with the noun-noun group recalling significantly more words than either the standard group or the number-noun group. The difference between the standard group and the number-noun group was not significant (see Table 2).

Overall, recall performance varies significantly across groups, with higher levels of recall demonstrated by the standard and noun-noun groups and lower levels of recall demonstrated by the number-noun group. Poorer recall of the number-noun group is evident for each third of the list. The difference between the standard group and the noun-noun group is related to the section of the list from which the words were recalled, with the standard group outperforming the noun-noun group in the first third and the noun-noun group outperforming the standard group in the final third.
Recall Intrusions

The proportion of recall intrusions was also calculated for each subject, categorized by the type of recall intrusion made: non-critical (intrusions other than the critical lure) or critical (an intrusion of the critical lure). These recall intrusions, critical and non-critical, are expressed as proportions, with higher proportions indicating more intrusions. The overall $F$ value for between-group differences in the recall of non-critical intrusions was not significant, $p > .05$; however, post hoc comparisons revealed that the number-noun group made significantly more non-critical intrusions than did the noun-noun group (see Table 3). Breaking down the non-critical intrusions by type, we see that many of the recall errors can be classified as either phonological intrusions (words that sound like the presented words) or non-critical semantic intrusions (words other than the critical lure that fit the general theme of the lists). Overall, there was no difference between groups in the proportion of errors that can be defined as phonological, $p > .05$ (see Table 3). The overall $F$ test for group differences in the proportion of intrusions that can be defined as semantic was also non-significant ($p > .05$); however, post hoc analyses revealed significantly more non-critical semantic intrusions in the number-noun group compared to the standard group (see Table 3).

In terms of intrusions of the critical lure, there was no effect of group on recall of the critical lure overall (i.e., collapsed across lists), $F < 1$ (see Table 4). However, when recall of the lure *sleep* is assessed independently of recall of the lure *needle*, there is a surprising dissociation. Although, there was no significant difference between groups in recall of the lure *sleep*, $F < 1$ (see Table 4), there was a significant difference between groups in recall of the lure *needle*, $F(2,528) = 4.45$, MSE = .247. For recall of the lure *needle*, subjects in the number-noun group made significantly more false recalls than did subjects in the standard
group (see Table 4). In general, then, there was a tendency for subjects in the number-noun group to falsely recall both critical and non-critical items to a greater extent than subjects in either of the other two groups.

When subjects did recall the critical lure, there was a difference in the recall position of that lure between the groups. An analysis of variance for recall position of the lure *sleep* resulted in a non-significant $F$ (although the value approached significance), $F(2,528)=2.61$, $MSe=4.536$, $p=.076$. Although the overall $F$ was not significant, post hoc analyses revealed that subjects in the number-noun group recalled the lure *sleep* significantly earlier in their output than did subjects in the standard condition (see Table 5). For recall of the lure *needle*, there was a significant between-group difference, $F(2,528)=12.24$, $MSe=4.627$, with the number-noun group recalling the lure *needle* significantly earlier in the output than was the case for either the standard group or the noun-noun group (see Table 5). There was not, however, a significant relation between recall position of the word *sleep* and recall position of the word *needle*, $r=.07$, $p<.05$. In other words, recall position of the critical lure was effected by experimental group, but was not related to recall position of other critical lures (i.e., early recall of one lure does not suggest early recall of other lures). Overall, after averaging the recall positions of all critical lures, there was a significant between-groups difference, $F(2,528)=10.12$, $MSe=3.929$, again with the number-noun group recalling the lures earlier than the standard and noun-noun groups (the words were recalled in positions 3.91, 4.88, and 4.91, respectively). Note that this difference in average recall position for the critical lure is mostly a function of the large between-group differences found using the *needle* list.

*Recognition Test*
Correct Recognition

Following from the free recall analysis, for each subject a recognition score was generated as the proportion of words correctly recognized from all presented lists (a perfect score of 1.0 would reflect the correct recognition of all 12 presented words included in the recognition tests, 6 words from each test). Recognition scores were also broken down based on the word’s position within the list, from the first third, the second third, or the final third. In general, higher proportions reflect more accurate recognition of the presented words. An analysis of variance revealed a significant difference in correct recognition across the groups, $F(2,528)=19.58$, $MSe=.02359$. Paralleling the free recall results, the standard group demonstrated the highest levels of correct recognition at .79, followed by the noun-noun group at .75, and then the number-noun group at .69 (see Table 1).

Recognition of presented words from the first third of the word list also differed significantly across groups, $F(2,528)=31.18$, $MSe=.04144$. Again, the highest levels of correct recognition were demonstrated by subjects in the standard group, followed by the noun-noun group, and then the number-noun group (see Table 2). The recognition hit rate for middle third did not significantly differ across groups overall, $F(2,538)=2.64$, $MSe=.05421$, although the difference approached significance, $p=.07$. However, post hoc analyses reveal significantly greater recall of words from the middle third by the standard group compared to the number-noun group (see Table 2). Recognition of words from the final third also differed significantly across groups, $F(2,528)=5.570$, $MSe=.06581$. Recognition by both the standard group and the noun-noun group was significantly higher than recognition by the number-noun group (see Table 2). In sum, the same pattern was evident for correct recognition as for correct recall. Both the standard group and the noun-
noun group outperformed the number-noun group, and there was seemingly little difference between the standard group and the noun-noun group.

**False Alarms in Recognition**

There are three types of possible errors on the recognition test: false alarms to unrelated lures, false alarms to weakly related lures, and false alarms to the critical lures. Proportions were calculated for each type of error based on the number of possible errors of that type that could be made (e.g., there were 2 unrelated lures per recognition test which creates the potential for 4 false recognitions of unrelated lures over the two tests, thus a proportion of 1.0 would indicate that 4 such false alarms were made). As with the false recalls, higher proportions indicate that a greater number of errors/false alarms were made. There was a significant between-group difference in the false recognition of unrelated lures, $F(2,528)=3.49, MSe=.008254$. The number-noun group made more false alarms to unrelated lures than did either the standard or the noun-noun group (see Table 6). However, there was no group difference in the false recognition of weakly related lures, $F<1$ (see Table 6). Of central interest, there was a significant between-group difference in recognition of the critical lure, $F(2,528)=3.83, MSe=.107$. Both the noun-noun group and the number-noun group made more false recognitions of critical lures than did the standard group (see Table 4). Looking at the needle list and the sleep list independently, it is clear that this difference arises from the needle list. There was no difference between groups in recognition of the lure sleep, $F<1$, but there was a significant difference in recognition of the lure needle, $F(2,528)=13.24, MSe=.168$. For the needle list only, false recognition of the lure was higher for the noun-noun and number-noun groups than for the standard group. Again, these results parallel the results of the free recall analysis: Between group differences in recall of the lure
seems specific to the **needle** list. Although the list differences are interesting, it is important to note that the **needle** list always followed the **sleep** list, therefore, obtained list differences could be due to either the nature of the lists or to the presentation position of the list.

**Paired-Associates Test**

The two experimental groups, the number-noun group and the noun-noun group, also completed a cued-recall paired associates task. Proportion of paired-associates correctly recalled was computed as a function of the number of possible correct paired associates (6 for the noun-noun group and 12 for the number-noun group). In general, there was a significant difference between the two groups in number of recall attempts, \( F(1,339)=178.05, \) \( MS_e=.02518. \) Subjects in the noun-noun group made significantly more recall attempts than did subjects in the number-noun group. There was also a significant difference between the two groups in the proportion of pairs correctly recalled, with the noun-noun group completing proportionately more pairs correctly than the number-noun group, \( F(1,339)=82.70, \) \( MS_e=.01224 \) (see Table 7). Looking at the errors made in recalling the pairs, there was no significant difference between the two groups in use of non-critical intrusions to complete the pairs, \( F<1. \) However, the noun-noun group used critical lures to complete the pairs significantly more often than did the number-noun group did, \( F(1,339)=60.53, \) \( MS_e=.02560 \) (see Table 7). Overall, the noun-noun group made more accurate recall attempts, but they were also more likely to falsely pair the non-presented critical lure with the presented items.

**Correlational Analyses**

The preceding analyses showed reduced correct recall and recognition for the number-noun group, and yet the error rates for this group were higher than those for the other
two groups. Analyses of the relation between correct recall/recognition and non-critical
intrusions and between correct recall/recognition and critical intrusions should help to clarify
this relation. For all correlational analyses Fisher's $t$ to $z$ transformation was used to
determine the significance of $r$. Overall (collapsed across all groups), an $r$ value of .08, .11,
or .14 is significant at an alpha level of .05, .01, or .001, respectively. For the standard group
only an $r$ value of .14, .19, or .24 is significant at an alpha level of .05, .01, or .001,
respectively. Similarly, for the number-noun group, an $r$ value of .15, .19, or .24 is
significant at an alpha level of .05, .01, or .001, respectively. Finally, for the noun-noun
group, an $r$ value of .16, .20, or .26 is significant at an alpha level of .05, .01, or .001,
respectively.

The correlation between correct recall and the number of recall intrusions was -0.51, a
correlation which is significant. For each group, as the incidence of correct recall increased,
the number of non-critical intrusions decreased (see Table 8). Correct recall was also related
to recall of the critical lure; although the relation was weak, $r=.25$, it was significant. It
appears that there was an increased tendency toward recall of the critical lure as correct recall
increased, a relation which is significant for each individual group (see Table 8). Note,
however, that when recall of the critical lure is broken down by the two specific items, the
relation between correct recall and recall of the critical lure becomes slightly weaker ($r=.18$
for correct recall and recall of the lure sleep, $r=.20$ for correct recall and recall of the lure
needle). These correlations were both still significant, but the relation was weak and
information from one variable did little to aid in prediction of the other variable. There was
also an overall weak, but significant, negative relation between recall of the critical lure and
other non-critical recall intrusions, $r=-.18$. This negative relation is true for each group, but
is not significant in the case of the standard group (see Table 8).

Recall of the critical lures themselves were also positively and significantly, but weakly correlated, $r=.14$, a pattern which holds true for each group, although is not significant for the noun-noun group (see Table 8). There was a slight tendency to recall needle if sleep has been recalled, but the relation was not a strong one. Overall then, the strongest correlation was the relation between correct recall and non-critical recall intrusions, which suggests that more accurate correct recall is related to fewer non-critical intrusions (a finding that is not unexpected because those subjects who are more accurate at recall are also expected to make fewer mistakes). Interestingly, although better performance in terms of correct responses is highly correlated with fewer non-critical errors, correct performance is also correlated with false memory for the critical lure. It seems that accurate performance reflects a decline in some types of errors (non-critical), but a tendency towards an increase in other types of errors (errors related to the critical lure). Other correlations, although significant, were modest at best and lacked useful predictive power because they accounted for only a small amount of the variability in the relation.

Significant correlations also appeared between the different recognition scores. Overall, correct recognition was significantly, but weakly, negatively related to false recognitions of unrelated lures, $r=-.16$, a pattern true for each group (but not significantly true in the case of the number-noun group), and was not related to false recognitions of weakly related lures, $r=.00$, a finding which also holds for each group (see Table 8). There was, however, a significant positive relation between correct recognitions and false recognitions of the critical lure, $r=.27$, and this is true for each group (see Table 8). This relation was true for both critical lures, for the lure sleep, $r=.23$, and for the lure needle.
\( r = .17 \). Again though, these correlations were modest. Recognition of the critical lure \textit{sleep} and the critical lure \textit{needle} was also correlated; although the relation was not strong, it was significant overall, \( r = .12 \). However, this strong correlation stems from the performance of the standard group, the other groups did not demonstrate a significant relation between recognition of the lures (see Table 8). As with the recall scores, correlations between accurate and inaccurate recognition were significant but very weak, and are not useful for predicting performance on recognition tests.

Not unexpectedly, there were strong relations between recall and recognition scores. There was a significant positive relation between correct recall and correct recognition overall, \( r = .55 \), and within each group (see Table 8). Similarly, there is a significant relation between recall of critical lures and false alarms to critical lures overall, \( r = .49 \), and again this relation holds within each group (see Table 8). For the critical lures, the correlation was evident for both the lure \textit{needle} and the lure \textit{sleep} independently. The relation between recall of the lure \textit{needle} and recognition of the lure \textit{needle} was significant, \( r = .53 \), as was the relation between recall of the lure \textit{sleep} and recognition of the lure \textit{sleep}, \( r = .47 \). Recall errors were also significantly related to recognition errors (although the relations were weak). There was a relation between non-critical recall intrusions and recognition of unrelated lures overall, \( r = .25 \), true for both the standard and number-noun groups (the positive relation for the noun-noun group was not significant). A relation also existed between recall of non-critical intrusions and recognition of weakly related lures, \( r = .21 \), however, this relation only holds for the number-noun group, the positive relation for the other two groups was not significant (see Table 8). In general, recall scores were predictive of recognition scores, although the nature of the relation varied slightly between groups.
Performance on the paired-associates test was also correlated with performance on the earlier recall and recognition tests. There was a significant relation between correct recall of word associates and correct recall of paired associates, \( r = .39 \), and between correct recognition of presented words and correct recall of paired-associates, \( r = .29 \). Overall, high performance on one test was predictive of high performance on other tests.

**Discussion**

**The Standard Group**

The obtained results for the standard group demonstrated levels of recall comparable to, but slightly lower than, past literature. For instance, the standard group in the present experiment recalled presented words correctly with a mean probability of .50, compared to Roediger and McDermott's (1995, Experiment 1) finding of .65 and Read's (1996, Experiment 1) finding of .64. Similarly, the recognition levels of the standard group were much lower than levels cited in past literature. The present standard group had a hit rate of .79, compared to Roediger and McDermott's (1995, Experiment 1) hit rate of .86. Although slightly lower than usual, the present standard group demonstrated the same broad pattern of hit rates as did comparable past groups. Notably, the recognition hit rate was much higher than the proportion of words correctly recalled. This result is not unexpected given that the recognition test always followed the free recall test, a condition which has been shown to inflate recognition scores (Roediger & McDermott, 1995).

Note also that a marked primacy effect was obtained, but a recency effect was not apparent (the mean proportions were .71, .40, and .39 for the first, second, and final third of the list, respectively). This finding does not parallel the marked recency and primacy effects obtained by Roediger and McDermott (1995). The likely cause of this discrepancy is in the
test instructions. During the free recall test, Roediger and McDermott (1995) instructed subjects to write down the last few words first; we did not give these same instructions. The subjects were merely asked to recall the words in whatever order they could think of them. It is possible that Roediger and McDermott (1995) fostered a recency effect through these instructions. The recognition scores for the present standard group showed the same pattern—an apparent primacy effect, but no recency effect (the mean proportions were .90, .79, and .68 for the first, second, and thirds of the list, respectively). The same pattern of results should be expected on the recognition test given that it followed the free recall test. Even though Roediger and McDermott did not explicitly give recency-effect inducing instructions during the recognition test, the artificial recency effect created during the free-recall test followed through to the recognition test. In the present study, the lack of a recency effect was consistent between tests.

Overall, the results demonstrated by the standard group are typical, with levels of correct recall and correct recognition slightly lower than, but comparable to, past literature reports. In terms of the serial position of words recognized and recalled, only a primacy effect occurred, there was no evidence of a recency effect, a finding that can be easily explained by variations in test instructions. In terms of correct recall and recognition, there was little difference between the present standard group and comparable groups in previous studies.

In terms of errors made by the standard group, non-critical intrusions occurred with a very low frequency (in approximately 6% of the lists), and over 20% of these non-critical intrusions involved words that either sounded like (12.25%) or were semantically related to (11.03%) the words in the list. This rate of non-critical intrusions is considerably lower than
that found in Roediger and McDermott's (1995) study: Their subjects falsely recalled non-
critical words approximately 14% of the time. In the present experiment, the critical lure
appeared in approximately 42% of the lists, comparable to Roediger and McDermott's
(1995) false recall rates of 40% (Experiment 1) and 55% (Experiment 2). This study utilized
12 word associates and the false recall effect of 42% is virtually identical to Roediger and
McDermott's finding of 40% when they used only 12 associates (in Experiment 2 they used
15 associates, and speculated that the increased rate of false recall was a result of the
increased number of word associates). In some ways, these comparable levels of false recall
of the critical lure are unexpected given the difference in correct recall between the present
study and Roediger and McDermott's study (the present recall levels were lower). This
suggests there is not a necessary relation between correct recall and false recall, a finding
supported by the present correlational analyses and, in an analysis of the relation between
failed to find a consistent or reliable relation between true and false recall and in the present
study only a very weak relation between the two factors was found.

Looking at errors made on the recognition test, the critical lure was falsely
recognized in approximately 68% of the lists (versus 84% in Roediger & McDermott, 1995).
In fact, the false recognition rate of 68% is only marginally lower than the obtained hit rate
of 79% (Roediger & McDermott, 1995, also found comparable false alarm and hit rates), but
is much lower than the false alarm rate cited by Roediger and McDermott. False recognition
of unrelated lures was low (less than 1%), as was false recognition of weakly related lures
(approximately 14%). These false recognition results are comparable to prior studies (e.g.,
Roediger & McDermott, 1995). Following past interpretations (e.g., Roediger &
McDermott, 1995), a low proportion of non-critical errors lends credibility to the proposal that false memories of the critical lure are unique and are not merely a reflection of guessing or simple mistakes.

The question remains: Why does the present standard group seem to differ in some ways from apparently comparable past groups? Overall, although correct recall and recognition was only slightly lower compared to previous studies, an interesting pattern of errors emerged. Despite reduced correct recall, the proportion of false recalls paralleled past studies, but the proportion of false recognitions was lower than in past studies. Given that there is no clear relation between correct recall and false recall or between correct recognition and false recognition, the present results are likely not cause for concern. In other words, the fact that there is no clear relation between hits and false alarms (Stadler et al., 1999) means that variations in the hit rate-false alarm pattern can be expected, and these normal fluctuations can likely accommodate the present findings.

A second explanation is in the number of word lists used in the present study. For instance, in their Experiment 1, Roediger and McDermott (1995) presented subjects with six lists; in their Experiment 2, they used 16 lists. The present results may simply be more unstable given that the results cited here are averaged over only two lists. Read (1996, Experiment 1) used only one list (the sleep list) in his first experiment and his results differed in some ways from Roediger and McDermott’s. In fact, Stadler et al. (1999) found that lists can differ dramatically in their effectiveness. It might be the case that using a restrictive sampling of lists results in a consequent alteration of the pattern of results obtained. Supporting this idea is the dramatic finding of differing results between the two lists utilized in the present study. Different patterns of results between the needle list and the sleep list
were found in several analyses. Potentially, the present results differ from prior literature simply because of the number of and the nature of the lists we choose.

In the present study two lists shown to produce dramatic false recall and false recognition effects were selected to determine whether these false memories could be eliminated, but the present findings revealed different patterns of results for each list. Interestingly, the use of lists shown to produce powerful false memory effects should have resulted in higher than normal levels of false recall and false recognition, but equivalent levels of false recall and reduced levels of false recognition were found. Why did we not obtain inflated levels of false alarms by using two lists shown to create powerful false memory effects? It is possible that the reduced levels of recall and recognition overall attenuated a potential increase in false recall and false recognitions. Although Stadler et al. (1999) found the hit rates and false alarm rates are not correlated, in the present study the correlation was in fact significant, albeit very weak. In the present study, there was a general tendency for unrelated false alarms to decline and related false alarms to increase as correct recall performance increases. The lack of a definitive relation between accurate and false recall/ recognition makes it hard to interpret the intricate relations found here and to assess why they differ from past studies. Future studies are needed that more clearly delineate the relation between true and false recall, which potentially could differ for each word list used. General principals in this realm may be hard to extract. In all likelihood, if any true relation does exist, it is likely to be true under a limited number of circumstances.

In general, despite the small differences in the performance of the present standard group compared to equivalent groups from previous studies, the present results fit the general DRM paradigm pattern: Typically accurate performance on simple word list memory tasks
is marred by high numbers of false alarms for words associatively related to words actually presented. Small variations in the present findings compared to the original Roediger and McDermott (1995) revival study are not reason for concern. Deviations from the original pattern of results are not unusual and have been cited in many follow-up studies (e.g., Read, 1996).

**The Number-Noun Group**

With respect to the experimental questions of interest, the performance of the number-noun group did differ from that of the standard group, but not in the direction expected. First, overall levels of correct recall were lower in the number-noun group, probably an indication of the list-length effect (Murdock, 1961). Subjects in the number-noun group had twice as many items to learn as did those in the other two groups (all of the word associates plus the 12 low-association numbers, compared to just the 12 word associates), plus they had to learn to link up the pairs. Although a pilot study was conducted to determine a presentation rate that would equate recall performance between groups, the experimental number-noun group still showed reduced recall performance (in the pilot study, recall rates were equated across groups at a 2-s presentation rate, however, this equivalence did not follow through to the experimental groups). One possible explanation is the change in presentation mode that occurred between the pilot study and the experiment. During the pilot study the words were presented via audiotape over a loud-speaker system. In contrast, mechanical difficulties prevented the use of the same system during the actual testing and, as a result, the words were still presented auditorily, but this time they were spoken out loud by the experimenter. During the experiment, the presentation rate, although well-practiced at 2 s, may have lacked the precision of the pilot study. The end result might have been lower
recall scores for the number-noun group. As expected, the lower recall scores followed through as lower recognition scores and then as lower paired associate scores. For all tests given, the other two groups out-performed the number-noun group in terms of accuracy.

The noun-noun group and the standard group also outperformed the number-noun group in terms of accuracy in a second way: The number-noun group was far more susceptible to the false memory effect than were either of the other groups. In the recall test, the number-noun group falsely recalled non-critical, non-presented words more often than did the noun-noun group, and they made more non-critical, but semantically related intrusions than did the standard group. Furthermore, the subjects in the number-noun group more often falsely recalled the critical lure needle than did subjects from the other two groups (there was no between-group difference in recall of the lure sleep). As well, subjects in the number-noun group tended to recall critical lures earlier in their output than did subjects in the noun-noun or standard groups. These findings are interesting because not only did the number-noun group recall fewer words correctly, they also made errors at a rate equivalent to or exceeding the rate of errors made by the other two groups.

This finding of equivalent and, in some cases, higher than normal error rates is a finding that is inconsistent with the original hypothesis. The expectation was that the number-noun group would be less susceptible to the false memory effect because learning of the number-word pairs, a difficult task, would supercede semantic analysis of the individual word associates. In other words, we expected that the task of learning the word pairs would consume most of the study time, leaving no time for forming semantic links between words, a factor that seems to be necessary in the formation of false memories for lists of word associates (Tussing & Greene, 1997). Tussing and Greene (1997) hypothesized that
processing tasks could distract from the processing involved in false memory formation, processes that are assumed to occur during encoding. In their experiment, incidental learning tasks served as a distracting processing task that resulted in greatly reduced levels of false memory formation. Similar results using the paired-associate task were expected.

Furthermore, the higher than expected false recall rates are even more unexpected given that pairing numbers with the word associates serves to differentiate the associates at the time of study, precisely when the critical lure is thought to be formed. Schacter, Norman, and Koutstaal (1998) hypothesized that false memories arise because features of the different events, in this case the word associates, are too similar to allow for adequate pattern separation and so the critical lure, which is not sufficiently distinct, is mistaken for a presented word. Following this logic, manipulations which serve to differentiate the study items, such as pairing the associates with distinctive numbers, should make the events distinct enough that adequate pattern separation is possible and the critical lure, being without a distinctive number, is not mistaken for a presented word. A similar idea has been put forward by Read (1996), after he found that giving subjects instructions to remember the order of presentation decreased false recall of critical lures. He argued that this occurred because subtle differences in the processing of presented items at study can make the items more distinctive at retrieval, thus eliminating the conditions necessary for false memory. Note the similarity these arguments have to IAR and fuzzy trace theories. In sum, past research has all pointed to an expected decline in false memory formation in the number-noun group, a finding that did not materialize in this study. There are several possible explanations for this unanticipated result.

First, the subjects in the number-noun group had more information available to them
than any of the other subjects, but for some reason they did not take advantage of this extra information. Potentially the numbers could have been used to render the presented words more distinctive, resulting in a decrease in false memory formation (if distinctiveness theories are correct). Possibly the subjects did not utilize this information because they did not have time. Paired associates are difficult to learn, especially after one presentation, and the presentation was at a fairly quick pace. Similarly, the subjects were asked to learn 12 paired-associates, a number reaching the upper limits of paired-associate learning. It could be that the subjects were unable, under the conditions given, to learn the paired-associates, resulting in a greater than normal reliance on list theme. Specifically, the number-noun subjects not only had more to learn, but the learning of the paired-associates led to greater processing demands. Memory over load might account for the performance displayed by the number-noun group (Seamon, Luo, & Gallo, 1998). A previous finding consistent with this notion (but explained in terms of deployment of attention) showed that subjects could accurately remember complex sentences only if given sufficient time to adequately encode the sentences. In contrast, without adequate encoding time, even simple sentences were difficult to reproduce accurately and subjects tended to demonstrate memory for only the general gist of the sentence (McDaniel, 1981). In other words, without sufficient study time, subjects relied on the gist of the sentence and failed to accurately reproduce the sentences. In all likelihood, the subjects in this study simply could not meet all of the demands placed on their memory in such a short time and so, during recall, had to rely on general themes rather than individual items because the individual items had been too difficult to encode.

In terms of the distinctiveness/pattern separation theory put forward by Schacter and his colleagues (1998), the events actually became more similar, not dissimilar, simply
because of the task difficulty. In the end, it seems that the number-noun group could only retain the common semantic features of the list, not the distinctive details of the individual items. This idea is supported by the finding of lower than normal levels of correct recall (indicating that the words were not well-learned) and the poor performance on the paired associates task (pairing of the stimulus with the critical lure and other non-critical intrusions). We can see from the paired-associates test the pairs were not well-learned (in fact, less than 10% of the word pairs were correctly recalled). The same arguments can be made to explain performance on the recognition test, which follows the same pattern as the recall test.

To sum-up the performance of the number-noun group, substantial false memory effects were found despite the expectation that the paired-associate manipulation would reduce the false memory effect. In all likelihood, by increasing task demands without substantially compensating for the increased demands, we created a situation in which the subjects relied on the theme of the list to a greater, not a lesser, extent. The result – poorer overall performance and a greater susceptibility to false memories.

The Noun-Noun Group

With respect to the second experimental question of interest, performance of the noun-noun group did not differ significantly from that of the standard group. In general, we found similar rates for correct recognition (although the portion of the list the words were recalled from differed slightly), with the noun-noun group correctly recalling approximately 47% of the presented words and the standard group recalling approximately 50% of the presented words (although the difference was significant, it is not substantial). The noun-noun and standard groups also had similar error rates for non-critical intrusions (6% and
5.6%, respectively) and for critical intrusions (46% and 42%, respectively). Similarly, there
was no difference in output position of the critical lure between the two groups (position 4.9
and 4.8 for the noun-noun and standard groups, respectively). Results from the recognition
test for the standard group and the noun-noun group were also very similar, with similar rates
of correct recognition (approximately 79% and 75% for the standard group and the noun-
noun group, respectively), which, although significantly different, are not dramatically so.
Although recognition of unrelated lures and weakly related lures did not differ between the
standard group and the noun-noun group, the noun-noun group did falsely recall the critical
lure needle more often than the standard group (approximately 87% vs. 65% of the time).
This difference in false recognition of the lure needle (but not sleep) is the only difference
between the two groups. This difference can possibly be attributed to the pairing of
associates, which served to make the semantic link between the words more obvious, and
consequently, made the words on the list less distinctive. Specifically, the needle list
traditionally produces lower levels of false recall than the sleep list (Roediger & McDermott,
1995), so pairing may have had a greater effect on the needle list than on the sleep list.

In terms of the hypothesis, the noun-noun group did not substantially differ from the
standard group. There was the possibility that false recall and false recognition rates might
have become inflated by pairing word associates together because the act of pairing the
associates could serve to make the relation between the words more obvious (and there was
some evidence of this with the needle list). In other words, pairing the words artificially
created a link between the words that some subjects might not have generated on their own.
However, the forced link did not substantially impact false memory rates (with the exception
of false recognition of the lure needle). One possibility is that because the false memory
rates were already so high, particularly for the sleep list, it might have been hard to increase the rates further (the ceiling effect). Simply put, most of the subjects were forming the links on their own and making the link more explicit for the few subjects who may not have been generating the link did not impact the overall results.

Alternatively, we anticipated a possible decline in false memory rates because the pairing of word associates provided subjects with a semantically appropriate associative response for each word, possibly eliminating further analysis that would generate the critical lure. Again, the results did not support this hypothesis. It seems that providing an alternative associative response does little to suppress generation of the critical lure (a finding paralleled in the number-noun group). Pairing word associates has little impact on true or false recall rates, indicating that there is nothing inherent in the act of pairing that alters rates (therefore the findings in the number-noun group cannot be explained as simply an effect of the pairing process). It seems that presenting the same words in an altered manner (in paired rather than individual form) has little effect on the overall results. As an aside, it is notable that subjects were better at recalling pairs in the noun-noun condition than in the number-noun condition. Although list-length plays a role in this effect, this finding is not unlike an earlier study which found that contextually related word pairs are easier to remember than pairs that are not contextually related (Fisher & Craik, 1977).

This finding of no difference between the standard group and the number-noun group also argues against the distinctiveness theory because, again, pairing the word associates serves to make them more distinctive (the response component of the pair acts as the distinctive feature). Possibly, because the response component of the pair still fits the general theme of the list, it cannot be considered a distinguishing characteristic. This
supposition is supported by the finding that subjects falsely paired the stimulus with critical lures approximately 19% of the time when they made a recall error. If the act of pairing made the associates more distinctive, not only would there be reduced levels of false recalls and false recognitions, but there would be few critical lure intrusions during the paired associates test. In sum, the noun-noun group and the standard group demonstrated approximately equivalent performance across all tasks: Pairing the word associates did not affect the results in a substantial way.

The Overall Picture

Overall, we did not manage to successfully eliminate, or even to reduce, the false memory effect. The number-noun group actually demonstrated higher than normal levels of false recall and false recognition, a finding consistent with the idea that they were more likely to encode the list in a general theme, rather than as individual, distinct items. The standard group, as expected, also seemed to encode the list in terms of a general theme, but there is more evidence of veridical recall in this group, indicating that they were able to access information about individual items to a greater extent than was the case for the number-noun group. Performance by the noun-noun group parallels that of the standard group. These findings are consistent with fuzzy trace theory, in which memories are believed to be stored in two independent, but parallel traces: A gist trace and a verbatim trace. In the context of these results, subjects in the number-noun group, because of their higher memory load, were forced to rely on the gist trace to a greater extent than were subjects in the noun-noun group or the standard group, resulting in more unrelated errors and more false recalls and false recognitions to the critical lure. This is not to say that distinctiveness theory fails: In fact, distinctiveness theory is consistent with fuzzy trace
theory in that, when items are not distinctive, subjects are more likely to rely on the gist-based trace because the verbatim trace is less useful when there is too much overlap between the to-be-remembered items (see, e.g., Einstein & Hunt, 1980). We were unable to make the words sufficiently distinctive, and thus, were not able to induce an act of memory retrieval supported by the verbatim trace to a greater extent than the gist trace. In fact, by making the task too difficult, subjects "gave up" (a reaction visible to the experimenter by their moans and exasperated sighs), and during subsequent tests their verbatim trace was incomplete or unreliable, resulting in their gist trace becoming the dominant player in recall and recognition. Evidence from other experiments supports the contention that veridical recall is not necessary for false recall (e.g., Seamon et al., 1998; Stadler et al., 1999), and in fact, the two types of memory (true and false) appear to be unrelated, or at most, very modestly related.

To conclude, the results, although not entirely expected, still shed some light on the differing theories of false memory formation. As illustrated, the theories are not always inconsistent in their predictions, and in all likelihood multiple theories are necessary to describe false memory phenomena adequately. In the context of this study we have focussed on fuzzy trace theory and distinctiveness theory as a means of interpreting the results, but this is not to say that other theories cannot account for these findings as well. We had originally hypothesized that reducing semantic analysis of the word associates or changing the course of the analysis (by creating the pairs) would reduce the false recall effect. Semantic analysis is related to fuzzy trace theory in the sense that the gist trace is essentially the semantic theme that is extracted when the word associates are semantically analyzed. The manipulation, which was intended to reduce semantic analysis, also had the effect of
altering the distinctiveness of the word associates, which allowed us to discuss the results in terms of distinctiveness theories of false memory formation. In general, the results can be explained by gist and fuzzy trace theories and by distinctiveness theories, so we maintain that multiple explanations are possible, and perhaps, necessary.

*Implications*

The results reported in this study parallel other findings that demonstrate robust false memory effects in simple list-learning paradigms utilizing both free recall and recognition tests. To reiterate what Freyd and Gleaves (1996) said, the practical implications of this work are contentious at best. Although these results do illustrate the ease with which false memories can be created in laboratory settings, they do not mirror the real-life situations in which false memories can possibly have profound effects (e.g., in the recovering of repressed memories or in eyewitness testimonies). Studies such as this one are often used as evidence of memory’s’ fallibility; however, it is hard to extrapolate from paradigms such as this general principals which can be applied to memory formation in emotionally-charged, complex situations. At best, we can say that by understanding false memories at a basic level we can then take that knowledge and try to apply it at more complex levels (such as during emotionally laden events). For instance, if distinctiveness theory holds, the ideas behind distinctiveness and pattern separation could then be tested to determine their relevance during more complex memory tasks. What we can take away from studies such as this one is that false memories can easily be created in some contexts and that these memories appear to be quite robust and impervious to elimination.

*Future Research*

Although significantly reduced levels of false recall and false recognition in the
number-noun group were not found, this area of research is worthy of further exploration. For the reasons cited earlier, there is cause to believe that the subjects did not adequately learn the paired associates, a problem which may have resulted in increases, rather than decreases, in false memories. A longer learning time is necessary before paired associates can be ruled out as a possible method of reducing false memories by making the events more distinctive and by providing ready-made associative responses.

A second cause for concern with the number-noun group is that the numbers were presented as the stimulus and the words as the response (so that the numbers could serve as cues on the paired associates test). As a result, during the study trial, we were actually providing an associative response for each number, rather than supplying a new (non-critical) associative response for each word associate. Given this, based on Underwood’s theory of IAR’s, there would have been no mechanism to prevent subjects from generating an IAR to the nouns because they were not provided with an IAR in the study phase. To assess the effects of this, a second number-noun group, in which the noun is presented as the stimulus and the number as the response (i.e., a noun-number group), is needed to see if providing an IAR for each word associate reduces encoding in terms of a general theme and makes the word associates more distinctive.

In sum, future research in this area should focus on presentation length and presentation order as potential factors influencing false memories. Increasing study time and reversing presentation order of the number-noun paired associates is expected to render the words more distinctive and consequently to reduce false memories by eliminating or decreasing reliance on the gist-based, semantically generated, memory trace. Although the findings did not support the hypotheses, areas of concern have been pinpointed and these
potential sources can be isolated in future research initiatives.
References


Brainerd, C. J., & Reyna, V. F. (1998). When things that were never experienced are easier to remember than things that were. *Psychological Science, 9*, 484-489.


**Psychological Bulletin.** **114**, 3-28.


Appendix A

Instructions

Standard Group

Study Instructions
You will be presented with a series of words, each for about 1.5 seconds. Your task is to study these words and to learn them for a later test of your memory for them. The words will only be presented once each, so be sure to pay close attention. After the presentation of all of the words, I will give you further instructions.
Do you have any questions?

Recall Test
Your task is to try to remember all of the words that were just presented. Write these words down on the blank sheet in front of you in whatever order you think of them. Please write down a word only if you remember it being presented; that is, please do not guess. You will have three minutes to complete this task.
Do you have any questions?
You may begin now.

Recognition Test
Please turn to page X when I say begin. On this page you will see a list of words. Your task is to circle any word that you recognize as being a word that you just studied. Again, please do not guess - only circle a word if you are sure that it was presented earlier. You will have three minutes to complete this task.
Do you have any questions?
You may begin now.

Number-Noun Group

Study Instructions
You will be presented with a series of number-word pairs, each pair for about 2 seconds. Your task is to study these number-word pairs and to learn them for a later test of your memory for them. The pairs will only be presented once each, so be sure to pay close attention. After the presentation of all of the number-word pairs, I will give you further instructions.
Do you have any questions?

Recall Test
Your task is to try to remember all of the words that were just presented. Do not worry about recalling the numbers; for now, just try to remember the words that were presented, regardless of the number the word was paired with. Write these words down on the blank sheet in front of you in whatever order you think of them. Please write down a word only if
you remember it being presented; that is, please do not guess. You will have three minutes
to complete this task.
Do you have any questions?
You may begin now.

Recognition Test
Please turn to page X when I say begin. On this page you will see a list of words. Your task
is to circle any word that you recognize as being a word that you just studied. Again, please
do not guess - only circle a word if you are sure that it was presented earlier. You will have
three minutes to complete this task.
Do you have any questions?
You may begin now.

Paired-Associates Test
Please turn to page X when I say begin. On this page, you will see a list of numbers. Your
task is to try to provide the studied word that was paired with each of the numbers on the
sheet. Please do not guess; only provide a word if you are sure that it was paired with that
specific number during the study session.
Do you have any questions?
You may begin now.

Noun-Noun Group

Study Instructions
You will again be presented with a series of words, each for about 2 seconds. Your task is to
study these words and to learn them for a later test of your memory for them. The words will
only be presented once each, so be sure to pay close attention. After the presentation of all
of the words, I will give you further instructions.
Do you have any questions?

Recall Test
Please turn to page X when I say begin. Your task is to try to remember all of the words that
were just presented. Write these words down in whatever order you think of them. Please
write down a word only if you remember it being presented; that is, please do not guess.
You will have three minutes to complete this task.
Do you have any questions?
You may begin now.

Recognition Test
Please turn to page X when I say begin. On this page you will see a list of words. Your task
is to circle any word that you recognize as being a word that you just studied. Again, please
do not guess - only circle a word if you are sure that it was presented earlier. You will have
three minutes to complete this task.
Do you have any questions?
You may begin now.
Paired-Associates Test
Please turn to page X when I say begin. On this page, you will see a list of words. Your task is to try to provide the studied word that was paired with each of the words on the sheet. Please do not guess; only provide a word if you are sure that it was paired with that specific word during the study session.
Do you have any questions?
You may begin now.
## Appendix B

### Recognition Test

<table>
<thead>
<tr>
<th><strong>Sleep List</strong></th>
<th><strong>Needle List</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>peace *</td>
<td>yarn *</td>
</tr>
<tr>
<td>cabbage +</td>
<td>tennis +</td>
</tr>
<tr>
<td>snooze</td>
<td>prick</td>
</tr>
<tr>
<td>awake</td>
<td>eye</td>
</tr>
<tr>
<td>doze</td>
<td>haystack</td>
</tr>
<tr>
<td>sleep</td>
<td>needle</td>
</tr>
<tr>
<td>dream</td>
<td>sharp</td>
</tr>
<tr>
<td>snore</td>
<td>hurt</td>
</tr>
<tr>
<td>drowsy *</td>
<td>knitting *</td>
</tr>
<tr>
<td>beach +</td>
<td>religion +</td>
</tr>
<tr>
<td>bed</td>
<td>thread</td>
</tr>
</tbody>
</table>

**Note.** * denotes the weakly related lures and + denotes the unrelated lures.

The critical lure is shown in **bold**.
Appendix C

Paired Associates Test - Noun-Noun Group

<table>
<thead>
<tr>
<th>Sleep List</th>
<th>Needle List</th>
</tr>
</thead>
<tbody>
<tr>
<td>bed -</td>
<td>thread -</td>
</tr>
<tr>
<td>awake -</td>
<td>eye -</td>
</tr>
<tr>
<td>dream -</td>
<td>sharp -</td>
</tr>
<tr>
<td>snooze -</td>
<td>prick -</td>
</tr>
<tr>
<td>doze -</td>
<td>haystack -</td>
</tr>
<tr>
<td>snore -</td>
<td>hurt -</td>
</tr>
</tbody>
</table>

**Note.** The words presented in the paired-associates test are the stimulus component of the stimulus-response pair.
Appendix D

Paired Associates Test - Number-Noun Group

<table>
<thead>
<tr>
<th>Sleep List</th>
<th>Needle List</th>
</tr>
</thead>
<tbody>
<tr>
<td>17 -</td>
<td>23 -</td>
</tr>
<tr>
<td>76 -</td>
<td>55 -</td>
</tr>
<tr>
<td>35 -</td>
<td>80 -</td>
</tr>
<tr>
<td>48 -</td>
<td>40 -</td>
</tr>
<tr>
<td>94 -</td>
<td>81 -</td>
</tr>
<tr>
<td>77 -</td>
<td>56 -</td>
</tr>
<tr>
<td>26 -</td>
<td>63 -</td>
</tr>
<tr>
<td>32 -</td>
<td>30 -</td>
</tr>
<tr>
<td>70 -</td>
<td>95 -</td>
</tr>
<tr>
<td>97 -</td>
<td>89 -</td>
</tr>
<tr>
<td>28 -</td>
<td>72 -</td>
</tr>
<tr>
<td>42 -</td>
<td>44 -</td>
</tr>
</tbody>
</table>

*Note.* The numbers presented in the paired-associates test are the stimulus component of the stimulus-response pair.
Table 1

**Mean Proportion of Correct Recall and Recognition as a Function of Encoding Condition**

<table>
<thead>
<tr>
<th>Encoding Condition</th>
<th>Standard</th>
<th>Noun-Noun</th>
<th>Number-Noun</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recall</td>
<td>.50</td>
<td>.47</td>
<td>.39</td>
</tr>
<tr>
<td></td>
<td>(.13)</td>
<td>(.13)</td>
<td>(.12)</td>
</tr>
<tr>
<td>Recognition</td>
<td>.79</td>
<td>.75</td>
<td>.69</td>
</tr>
<tr>
<td></td>
<td>(.15)</td>
<td>(.16)</td>
<td>(.15)</td>
</tr>
</tbody>
</table>

**Note.** Standard deviations are shown below the means.
Table 2

Mean Proportion of Correct Recall and Recognition as a Function of List Position (First, Second, or Third Portion) and Encoding Condition

<table>
<thead>
<tr>
<th>Encoding Condition</th>
<th>Standard</th>
<th>Noun-Noun</th>
<th>Number-Noun</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt;</td>
<td>3&lt;sup&gt;rd&lt;/sup&gt;</td>
</tr>
<tr>
<td>Recall</td>
<td>.71</td>
<td>.40</td>
<td>.39</td>
</tr>
<tr>
<td></td>
<td>(.19)</td>
<td>(.19)</td>
<td>(.19)</td>
</tr>
<tr>
<td>Recognition</td>
<td>.90</td>
<td>.79</td>
<td>.69</td>
</tr>
<tr>
<td></td>
<td>(.16)</td>
<td>(.23)</td>
<td>(.26)</td>
</tr>
</tbody>
</table>

Note. Standard deviations are shown below the means.
Table 3

Mean Proportion of Non-Critical Recall Errors as a Function of Error Type and Encoding Condition

<table>
<thead>
<tr>
<th></th>
<th>Encoding Condition</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Standard</td>
<td>Noun-Noun</td>
<td>Number-Noun</td>
<td></td>
</tr>
<tr>
<td>Intrusions</td>
<td>Overall</td>
<td>.06</td>
<td>.05</td>
<td>.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.12)</td>
<td>(.08)</td>
<td>(.13)</td>
</tr>
<tr>
<td></td>
<td>Phonological</td>
<td>.12</td>
<td>.08</td>
<td>.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.30)</td>
<td>(.24)</td>
<td>(.28)</td>
</tr>
<tr>
<td></td>
<td>Semantic</td>
<td>.11</td>
<td>.14</td>
<td>.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.31)</td>
<td>(.32)</td>
<td>(.39)</td>
</tr>
</tbody>
</table>

Note. Standard deviations are shown below the means.
Table 4

**False Memory: Mean Proportion of Recall and Recognition of the Critical Unpresented Lure as a Function of Encoding Condition**

<table>
<thead>
<tr>
<th></th>
<th>Encoding Condition</th>
<th>Standard</th>
<th>Noun-Noun</th>
<th>Number-Noun</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Recall</strong></td>
<td>Overall</td>
<td>.43</td>
<td>.46</td>
<td>.48</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.38)</td>
<td>(.54)</td>
<td>(.37)</td>
</tr>
<tr>
<td></td>
<td>Sleep List</td>
<td>.42</td>
<td>.35</td>
<td>.37</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.49)</td>
<td>(.48)</td>
<td>(.48)</td>
</tr>
<tr>
<td></td>
<td>Needle List</td>
<td>.44</td>
<td>.52</td>
<td>.59</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.50)</td>
<td>(.50)</td>
<td>(.49)</td>
</tr>
<tr>
<td><strong>Recognition</strong></td>
<td>Overall</td>
<td>.68</td>
<td>.76</td>
<td>.76</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.37)</td>
<td>(.30)</td>
<td>(.31)</td>
</tr>
<tr>
<td></td>
<td>Sleep List</td>
<td>.69</td>
<td>.65</td>
<td>.71</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.46)</td>
<td>(.48)</td>
<td>(.46)</td>
</tr>
<tr>
<td></td>
<td>Needle List</td>
<td>.65</td>
<td>.87</td>
<td>.81</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.48)</td>
<td>(.34)</td>
<td>(.39)</td>
</tr>
</tbody>
</table>

*Note.* Standard deviations are shown below the means.
Table 5

Mean Recall Position of Incorrectly Recalled Critical Lures as a Function of Encoding Condition

<table>
<thead>
<tr>
<th>Encoding Condition</th>
<th>Standard</th>
<th>Noun-Noun</th>
<th>Number-Noun</th>
</tr>
</thead>
<tbody>
<tr>
<td>Needle</td>
<td>4.67</td>
<td>5.26</td>
<td>3.73</td>
</tr>
<tr>
<td></td>
<td>(2.35)</td>
<td>(2.04)</td>
<td>(2.07)</td>
</tr>
<tr>
<td>Sleep</td>
<td>5.13</td>
<td>4.49</td>
<td>4.38</td>
</tr>
<tr>
<td></td>
<td>(2.32)</td>
<td>(1.93)</td>
<td>(2.04)</td>
</tr>
<tr>
<td>Average</td>
<td>4.88</td>
<td>4.91</td>
<td>3.91</td>
</tr>
<tr>
<td></td>
<td>(2.16)</td>
<td>(1.89)</td>
<td>(1.88)</td>
</tr>
</tbody>
</table>

Note. Standard deviations are shown below the means.
Table 6

Mean Proportion of Non-Critical Recognition Errors as a Function of Error Type and Encoding Condition

<table>
<thead>
<tr>
<th></th>
<th>Encoding Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Standard</td>
</tr>
<tr>
<td>Unrelated Lures</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>(.06)</td>
</tr>
<tr>
<td>Weakly Related Lures</td>
<td>.14</td>
</tr>
<tr>
<td></td>
<td>(.28)</td>
</tr>
</tbody>
</table>

**Note.** Standard deviations are shown below the means.
Table 7

Mean Proportion of Paired-Associates Recalled (Correctly and Incorrectly) as a Function of Encoding Condition

<table>
<thead>
<tr>
<th>Encoding Condition</th>
<th>Noun-Noun</th>
<th>Number-Noun</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct Recall</td>
<td>.19</td>
<td>.08</td>
</tr>
<tr>
<td></td>
<td>(.14)</td>
<td>(.07)</td>
</tr>
<tr>
<td>Non-critical Intrusions</td>
<td>.03</td>
<td>.04</td>
</tr>
<tr>
<td></td>
<td>(.08)</td>
<td>(.13)</td>
</tr>
<tr>
<td>Critical Intrusions</td>
<td>.19</td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td>(.18)</td>
<td>(.14)</td>
</tr>
</tbody>
</table>

Note.  Standard deviations are shown below the means.
Table 8

**Correlations in Test Performance as a Function of Type of Response and Encoding**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Encoding Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Overall</td>
</tr>
<tr>
<td><strong>Recall</strong></td>
<td></td>
</tr>
<tr>
<td>Correct recall w/ non-critical intrusions</td>
<td>-.50 ***</td>
</tr>
<tr>
<td>Correct recall w/ critical intrusions</td>
<td>.25 ***</td>
</tr>
<tr>
<td>Non-critical intrusions w/ critical intrusions</td>
<td>-.18 ***</td>
</tr>
<tr>
<td>Recall of needle w/ sleep</td>
<td>.14 ***</td>
</tr>
<tr>
<td><strong>Recognitions</strong></td>
<td></td>
</tr>
<tr>
<td>Correct recognition w/ unrelated lures</td>
<td>-.16 ***</td>
</tr>
<tr>
<td>Correct recognition w/ weakly related lures</td>
<td>.00</td>
</tr>
<tr>
<td>Correct recognition w/ critical lures</td>
<td>.27 ***</td>
</tr>
<tr>
<td>Recognition of needle w/ sleep</td>
<td>.12 **</td>
</tr>
<tr>
<td><strong>Recall with Recognition</strong></td>
<td></td>
</tr>
<tr>
<td>Correct recall w/ correct recognition</td>
<td>.55 ***</td>
</tr>
<tr>
<td>Critical intrusions (recall w/ recognition)</td>
<td>.49 ***</td>
</tr>
<tr>
<td>Non-critical intrusions (recall w/ recognition – unrelated)</td>
<td>.25 ***</td>
</tr>
<tr>
<td>Non-critical intrusions (recall w/ recognition - weakly related)</td>
<td>.21 ***</td>
</tr>
</tbody>
</table>

**Note.**  * denotes p.<.05, ** denotes p.<.01, *** denotes p.<.001.