2-1-2012

# Will YOU have a false memory? : investigating individual differences in semantic false memories

Susan Stevens-Adams

Susan Stevens-Adams

*Candidate*

Psychology

*Department*

This dissertation is approved, and it is acceptable in quality and form for publication:

*Approved by the Dissertation Committee:*

Dr. Karin Butler, Chairperson

Dr. Tim Goldsmith

Dr. Eric Ruthruff

Dr. Keith Hutchison

Dr. Laura Matzen

**WILL *YOU* HAVE A FALSE MEMORY? INVESTIGATING INDIVIDUAL DIFFERENCES IN SEMANTIC FALSE MEMORIES**

**BY**


**SUSAN M. STEVENS-ADAMS**

B.A., Psychology, New Mexico State University, 2003

M.S., Psychology, University of New Mexico, 2006


**DISSERTATION**

**Submitted in Partial Fulfillment of the**

**Requirements for the Degree of**

**Doctor of Philosophy**

**Psychology**

**The University of New Mexico**

**Albuquerque, New Mexico**

**December 2011**

## ACKNOWLEDGMENTS

I would like to express my gratitude to my advisor and committee chair, Karin Butler, Ph.D. Her never ending patience, support and encouragement are greatly appreciated and will never be forgotten. I would also like to thank my committee members, Tim Goldsmith, Ph.D., Eric Ruthruff, Ph.D., Keith Hutchison, Ph.D. and Laura Matzen, Ph.D. for their insight and vital contributions toward my dissertation. Special thanks to Dr. Hutchison for agreeing to be on my committee on a moment's notice.

I would like to thank the undergraduate students who have helped me collect data for my dissertation: Carly Camplain, George Custodio, Bryan Franks, Stephanie Jackson, John Mooney, Kristopher Phillips, Alanna Shea, Sarah Shrum, Aaron Wagner and Randy Walters. These students helped me out tremendously and I am grateful to each of them.

Last but not least, I must thank my family and friends for their support, especially my husband Brian and my parents. Their unconditional love, support, and patience were pivotal in my success. This dissertation would not have been possible without their love and encouragement.

**WILL *YOU* HAVE A FALSE MEMORY? INVESTIGATING INDIVIDUAL DIFFERENCES IN SEMANTIC FALSE MEMORIES**

**by**

**Susan Stevens-Adams**

B.A., Psychology, New Mexico State University, 2003

M.S., Psychology, University of New Mexico, 2006

Ph.D., Psychology, University of New Mexico, 2011

**ABSTRACT**

Four experiments extended the false memory research by investigating false memory performance at the individual level (e.g., Stevens, unpublished Master's Thesis, 2006). The experiments used the DRM paradigm which presents lists of associated studied words (e.g., *hot, winter*) and then tests whether nonpresented theme word(s) (e.g., *cold*) is (are) falsely remembered. Previous research (e.g., fuzzy-trace theory) has proposed that the semantic relationships between words on the DRM lists are influential in false memory performance. In the experiments, each participant rated the semantic relatedness of word pairs drawn from associated sets. The first experiment investigated whether participants' semantic ratings would predict their false memory performance. For the other three experiments, the ratings were later used in the study-test procedure, such that the theme words on the recognition test were preceded by words either deemed strongly related or weakly related to the theme word. The final experiment also explored whether creating study lists from semantic ratings influenced false memory. The results, expected to support the fuzzy-trace theory of false memory, provided more support for the activation/monitoring theory.

# TABLE OF CONTENTS

## LIST OF FIGURES

# LIST OF TABLES

# CHAPTER 1: INTRODUCTION

False remembering is remembering events that never occurred or remembering events rather differently than they actually occurred. Bartlett (1932) was one of the first researchers to study false memory and found that his participants exhibited distortions in their memories of a folktale. Several years later, Deese (1959) found that words that had a high occurrence of being falsely recalled were also highly associated to the other words on the list. Roediger and McDermott (1995) extended Deese's paradigm and found a high likelihood of both false recall and false recognition of the nonpresented theme word; they reported that in some cases the likelihood of false recognition approached the likelihood of correct recognition. Numerous studies that have expanded this work have used what has been termed the DRM (Deese-Roediger-McDermott) paradigm in which participants study a list of words that are highly related to a nonpresented theme word. The conditions that influence the likelihood of falsely recalling or recognizing this theme word are of most interest. Many researchers have replicated Roediger and McDermott's findings in numerous settings and with different populations (see Gallo, Roberts & Seamon, 1997; Lampinen, Neuschatz, & Payne, 1999; McDermott, 1996; McKelvie, 1999; Read, 1996; Robinson & Roediger, 1997; Seamon, Luo & Gallo, 1998).

One theory posited for why semantic false memory occurs is the activation/monitoring theory (see Roediger, Balota & Watson, 2001). The theory is based on spreading activation, the idea that each concept in memory is represented by a node in a network and the degree of association between pairs of concepts is represented by the distance between their corresponding nodes in the network (Collins & Loftus, 1975). In the DRM paradigm, the nodes for studied words are activated at study and activation spreads to nonstudied, but related, word nodes (i.e., the nonpresented theme word). On a subsequent memory test, the previously activated nonstudied word node is falsely remembered as having been studied if it reached a critical level of activation during study and test. Spreading activation has been supported by evidence of priming (Markman, 2002; Meyer & Schvaneveldt, 1971) and children's knowledge of certain concepts (e.g., Chi & Koeske, 1983).

Previous research has found a relationship among level of association between words and the likelihood of false memory, which has been argued to support the activation/monitoring theory of false memory. McEvoy, Nelson and Komatsu (1999) and Roediger, Watson, McDermott and Gallo (2001) found that the relatedness of words affected the rate of false recall and recognition; that is, the stronger the connection strength (namely, using list items that produced the nonpresented theme word with the highest probability), the higher the rates of false recall and recognition, which supports the spreading activation theory. McEvoy et al. (1999) utilized the DRM paradigm but varied both the strength (high or low) and density (high [list words that share many associate connections] or low [list words that share few associate connections]) of the connectivity of the words on the lists. They found that the likelihood of false recognition for the nonpresented theme words was higher when the strength of the connectivity of the words was high and when the words were more densely connected. They proposed that the stronger and more densely connected words caused more activation of the nonpresented theme words, thus supporting the spreading activation theory of false recognition.

It is important to explain how associations are obtained and to distinguish two ways of acquiring these association relationships between the list words and their respective theme word. In obtaining associations, participants are given one word (the cue) and asked to indicate the first word that comes to mind (the target; Nelson et al., 1998). Target responses are then ranked based on frequency; this frequency is used as a measure of the strength of association. This method results in two types of associations between studied and theme words. In the DRM paradigm, in one case, the theme item is the cue and the studied items are considered the targets. This has been referred to as the forward association strength (FAS) from the theme item to the studied item. The second type of association, when the studied items are cues and their theme item is the target (also called backward association strength; BAS), is more important for semantic false memory. McEvoy et al. (1999) and Roediger, Watson et al. (2001) used BAS to construct their study lists.

Many other researchers (e.g., Hicks & Hancock, 2002; Huff & Hutchison, 2011; Hutchison & Balota, 2005; McEvoy et al., 1999; Roediger, Balota & Watson, 2001; Roediger & McDermott, 1995; Roediger, Watson et al., 2001) have argued that spreading activation and, accordingly, the activation/monitoring theory, may explain the occurrence of false memories. In one study, Robinson and Roediger (1997) varied the number of words studied per list in a DRM paradigm. They found that false recall of the nonpresented theme words increased as the list length increased. Roediger, Balota et al. (2001) argued that these results are consistent with a spreading activation account, which "predicts that activation should summate for critical items in a way that is similar to the summation of activation in semantic priming experiments" (pg. 105). Thus, during a false memory experiment, the activation of the studied items extends to the nonpresented theme words leading the individual to believe that the nonpresented word had, in fact, been studied.

Hicks and Hancock (2002) investigated source attribution in false memories. In their experiments, the participants were presented with DRM lists in which either a male or female speaker presented the entirety of the list or a male speaker presented half of the items on the list and a female speaker presented the other half of the items on the list. The list words were arranged such that their BAS was from high to low. For the condition in which both a male and female speaker presented the list items, one speaker presented the low BAS words and the other speaker presented the high BAS words. The researchers found that participants were most likely to attribute the falsely recalled theme word to the speaker who presented the high BAS study list words. They argued that the results were consistent with the activation/monitoring account of false memory since the level of association had an influence on the attribution of false memories (Hicks & Hancock, 2002).

Another theory for why false memories occur is the fuzzy-trace dual-process theory (Brainerd, Reyna & Kneer, 1995). This theory states that there are essentially two processes occurring during the encoding of a memory: a verbatim-based process and a gist based-process. Verbatim traces include perceptual details of an experience that allow people to distinguish memories from one another. Gist information includes commonalities among experiences that are part of an episode. The theory states that false remembering of related but nonpresented words arise from the degree to which they match the gist traces. False memory of these nonpresented theme words can be decreased when retrieval of verbatim traces of the studied associates occurs. According to this theory, the semantic relationship between the words on the false memory lists is what determines whether a related nonpresented word is falsely remembered or not. In their study, Brainerd, Yang, Reyna, Howe and Mills (2008) analyzed the characteristics of words on typical DRM lists and found that false recall, false recognition and BAS loaded with measures of semantic relatedness like meaningfulness and familiarity. The researchers argued that these results best support the fuzzy-trace theory, since the theory assumes that semantic relationships are what drives the false memory effect.

Two studies, however, have found that semantic characteristics are not necessarily predictive of false memory. Huff and Hutchison (2011) performed a clever experiment in which their participants studied unrelated list items (e.g., *slope*, *reindeer*) that were related to mediators (e.g., *ski*, *sleigh*) which all converged upon a single nonpresented theme word (e.g., *snow*). The researchers reasoned that if the theme words were falsely remembered, it must be due to activation-based false memory since the list words themselves did not share semantic characteristics. The researchers found that the theme words were indeed falsely remembered, supporting the activation-based theory.

Along a similar vein, Hutchison and Balota (2005) had participants study list words that all converged onto one meaning of the nonpresented theme word (e.g., *snooze*, *wake*, *rest*…., for *SLEEP*) or onto a homographic theme word (which had two different meanings; e.g., *autumn*, *season*, *trip*, *stumble*…, for *FALL*). The associative strength from the list items to the nonpresented theme word was equated across lists. The researchers argued that if meaningful semantic information was leading to false memories then the homograph themes would have lower false memory than the single meaning themes since the list words for the homograph themes contained two different meanings. However, an associative activation theory of false memory would predict no difference in false memory performance between the two conditions since associative strength was equated across the lists. The researchers found that there were no differences in false memory performance between the two conditions. These results were supportive of an association based theory rather than a semantic based theory (Hutchison & Balota, 2005).

One distinction between these two theories is that the activation/monitoring theory focuses on the associations between words (i.e., words that tend to co-occur which are often seen as related; Hutchison & Balota, 2005) whereas the fuzzy-trace theory emphasizes the semantic relatedness between words (i.e., words that share meaning or semantic features which are seen as related; Hutchison & Balota, 2005). While these two distinctions seem clear, it is often difficult to tease them apart. Hutchison (2003) found that in priming studies, there was a large degree of semantic overlap for ‘associated’ word pairs. In addition, he argued that it is hard to determine a universal method for determining 'similarity' in word pairs since results can differ across studies depending on how the researcher defines 'similar', what words are used in the experiment and the extent to which the word pairs are associated (Hutchison, 2003; see also Moss, Ostrin, Tyler & Marslen-Wilson, 1995 and McRae & Boisvert, 1998). Wu and Barsalou (2007; as cited in Howe, Wimmer, Gagnon & Plumpton, 2009) state that there are six types of semantic relations that can occur between words: antonymy (words of opposite value, e.g., *hot* and *cold*), entity (entity-property relationships between words, e.g., *chair* and *wood*), introspective (mental state-property between words, e.g., *happy* and *sun*), situational (words that are related to the same association, e.g., *medical treatment*), synonymy (words that have the same meaning, e.g., *bunny* and *rabbit*), and taxonomy (words that belong to the same taxonomic category, e.g., *cats* and *dogs* are both animals). Thus, words can be associated or semantically related in many different ways (see Table 1 of Hutchison, 2003). For instance, two words could be category members (*sheep-goat*), share property relations (*canary-yellow*) or be part of the same script (*orchard-apple*). Some of these relations have been said to be mostly semantic or mostly associative but for most of them there is much overlap. Therefore, the study of 'semantic relations' in any given experiment is unclear considering that some of the relation types can overlap (Brainerd et al., 2008) and are rarely specified in any given experiment. In addition, there is overlap in the definition of semantic and associative relationships as the antonymy semantic relation is very similar to how researchers think about association.

Words on the DRM lists are highly associated but are also semantically similar and semantically related words are not always associatively related (see Brainerd et al., 2008). Past research has found that the most semantically related word was not always the most associatively related word for DRM lists (Stevens, unpublished Master's Thesis, 2006). In her experiments, Stevens (unpublished Master's Thesis, 2006) obtained pairwise semantic relatedness ratings for DRM word lists. For this measure, each word on the list is paired with every other word on the list. Participants are instructed to give a semantic relatedness rating, on a scale of not at all related to highly related, for each pair of words. Pairwise relatedness ratings have been shown to be a valid method for obtaining measures of semantic relatedness between words (see Cooke, 1992; Cooke, Durso & Schvaneveldt, 1986). Thus, Stevens (unpublished Master's Thesis, 2006) was able to obtain a measure of semantic relatedness (from the ratings) and associative relatedness (from the established word norms) and found differences between the two. For instance, the highest associate to the nonpresented theme word *cold* is *hot* (per the group normed data). However, Stevens (unpublished Master's Thesis, 2006) found *hot* to have one of the lowest average semantic relatedness ratings (2.26 out of 5) and the word *winter* to be the highest semantically related word. This makes sense given the associative/semantic distinction; the words *cold* and *hot* often co-occur together while *cold* and *winter* share more semantic features and meaning. More research is needed to further explore the distinction between semantic and associative relatedness and how they play a role in the DRM paradigm; the current experiments were conducted with this in mind.

## Individual Differences in False Memory

Most of the previous literature investigating false memories has focused on the group level. Some research has been performed looking at differential false memory performance in special populations including older adults, Alzheimer's patients, those with frontal lobe dysfunction, children and those with domain expertise. Older adults (Butler, McDaniel, Dornburg, Price & Roediger, 2004; Fernandes, Ross, Wiegand & Schryer, 2008; Jacoby & Rhodes, 2006; Mather & Johnson, 2000; McCabe & Smith, 2002; Watson, McDermott & Balota, 2004) and those with Alzheimer's disease (Waldie & See, 2003) remember fewer studied items but falsely remember more nonpresented theme words than younger adults. This finding, that older adults are more likely to false alarm to nonpresented words, is consistent across numerous studies. Butler et al. (2004) found that the age difference was limited to those older adults that were low in frontal lobe function. Dewhurst and Robinson (2004) looked at the likelihood of false memories in children ages 5, 8 and 11. They found that the 5-year-olds were more likely to falsely recall words that rhymed with the studied items than to recall semantically similar or unrelated items while the 11-year-olds were more likely to falsely recall words that were semantically similar to the studied words than either rhyming words or unrelated words. The 8-year-olds were just as likely to falsely recall words that rhymed or were semantically similar to the studied items. Metzger, et al. (2008) asked 2$^{nd}$ graders, 8$^{th}$ graders and college student to study age-appropriate DRM lists. They found that the 2$^{nd}$ graders had the fewest false memories while the college students had the most false memories. The 8$^{th}$ graders were similar to the 2$^{nd}$ graders in low-demand conditions but more similar to the college students in high-demand conditions. The researchers argued that their results show that as children mature they make greater use of their semantic knowledge, which leads to an increase in semantic based errors (Metzger et al., 2008). Howe et al. (2009) studied the effects of associative strength (predicted by the activation/monitoring theory) and gist relations (predicted by the fuzzy-trace theory) on children's and adults' likelihood of false memory. Children (ages 5, 7, and 11) and adults (age 18) studied DRM and category lists and/or lists that differed in associative strength (particularly BAS) and semantic cohesion. They found that, for all ages, the likelihood of false memory was determined by BAS but were unaffected by gist manipulations supporting the activation/monitoring conceptualization of false memory (Howe et al., 2009). Finally, Castel, McCabe, Roediger and Heitman (2007) looked at cohort differences in false memories based on domain expertise. They had young adults study a list of football-related animal names (e.g., lions, broncos, bears) and a list of body parts. The high-football-knowledge group was more likely to intrude football-related animal names than the low-knowledge group, but there was no difference between the groups in false recall of body parts. Thus, expertise in a particular domain had an influence on the likelihood of false memory for that field.

This prior research involving differing populations has predicted average false memory performance. None of these studies have assessed the semantic networks of individuals and many of their conclusions are based on group differences in processing. A study performed by Ceci, Papierno and Kulkofsky (2007) is the closest known to date to attempt to investigate individual differences in false memory. In their study, the experimenters asked two groups of children (aged 6 and 9) to view different triads of pictures and to indicate which picture did not belong with the other two. One to three months later, these children, along with matched peers, were presented a story that contained critical and noncritical items from the earlier triad task. Two days later, all children were interviewed about the critical and noncritical items in the story. Half of the children were purposely misled about the critical items (for instance, if *lemon* had appeared in the study, then *orange* was suggested), while the other half (control group) received the same interview but it did not contain misinformation. Five to seven days later, the children were shown pictures and asked which ones they remembered seeing during the study. The researchers used nonmetric multidimensional scaling to obtain a measure of semantic proximity for the triads. They found that the participants' triad semantic proximity served as a basis for correctly predicting memory performance in the larger group of children; the closer a suggested distractor was to the original item's representation in proximity, the greater was the distractor's suggested influence (Ceci et al., 2007).

Ceci and colleagues' (2007) study differs from the proposed work in several important ways. First of all, Ceci et al. (2007) were interested in the misinformation effect which is different than the false memories studied in the current work. Misinformation involves the purposeful misleading or suggesting of false information; the current work does not involve misleading the participants. In addition, only twelve children performed the triad task. Ceci and colleagues (2007) analyzed the data for these twelve children and found that individual misinformation was related to semantic proximity. However, these ratings were then combined to create the full ratings set used in the final analysis for sixty-two children who had not completed the triad task. Thus, the results of these sixty-two children were based on other children's ratings. In addition, Ceci and colleagues (2007) did not use the ratings to manipulate test or study items; in the current experiments, each participant completed the ratings task and these ratings were used to manipulate later study and test items for each individual.

Thus, there is a dearth of research focusing on how differences in individual semantic networks may predict *individual* false memory performance. The typically used DRM lists are derived from group norms which are based on agreement across large groups of people about associations between words (see Nelson et al., 1998). False memory research using these lists cannot determine whether relative strength of associations that are specific to an individual will predict false memory performance. A stronger test of the theoretical accounts of the false memory phenomenon would examine the individual since this would allow for more thorough and fine grained predictions about false memory performance.

In her experiments, Stevens (unpublished Master's Thesis, 2006) investigated the relationship between the likelihood of false recognition and semantic relatedness ratings by individual participants. Participants were asked to study false memory lists and then their level of false recognition of nonpresented words was measured. After the false memory study-test procedure, participants rated the semantic relatedness of pairs of words (see Cooke, 1992; Cooke, Durso & Schvaneveldt, 1986) constructed from the studied list and nonpresented theme words on the recognition test. Stevens (unpublished Master's Thesis, 2006) found that nonpresented words that were falsely recognized had higher relatedness ratings to studied words than the nonpresented words that were correctly rejected. These results are consistent with the fuzzy-trace theory given that participants were asked to rate the semantic relationship between the words on the list and those nonpresented words rated as more semantically related to other words on the list were found to have a higher likelihood of false remembering. However, these results also support the activation/monitoring theory because those words that are more semantically related to the theme may also be stronger associates to the theme thus eliciting more activation of the theme during study and more false remembering.

A limitation to Stevens' study is that the order of the two tasks was always the same; the participants always received the false memory study and recognition task before the relatedness ratings. It is possible that the false memory task influenced the participant's ratings. The increased familiarity of the theme and list words (caused by completing the memory task) may have led the participants to rate the words differently than if they were not previously seen in the experimental context. This limitation will be addressed in the current studies.

## Purpose of the Present Studies

The purpose of the experiments proposed here is twofold. First, they extend and delve into the individual differences in false memory research by examining individual relatedness ratings and false recognition performance of DRM lists.

Additionally, experimental manipulation will provide evidence regarding which of the two false memory theories explain the phenomenon. The fuzzy-trace theory states that semantic relatedness is most relevant to false memories while the activation/monitoring theory proposes that the level of association influences false memory. Participants will be asked to give semantic relatedness ratings, which have been shown to be a valid measure of semantic knowledge, for the lists of words. The words that participants indicate are highly semantically related (semantically strong) or weakly semantically related (but associatively strong) will be manipulated at test (Experiments 2-4) and study (Experiment 4) at the individual participant level.

Stevens (unpublished Master's Thesis, 2006) suggested a relationship between the likelihood of false recognition and an individual's ratings of semantic relatedness between studied and nonpresented words. In this dissertation, a potential confound will be addressed in these studies and these findings will be extended by individually tailoring recognition tests to influence the likelihood of false memory. Experiment 1 addressed the potential confound of the false memory test always being presented first in the previous study by reversing the order of the tasks; the participants were asked to perform pairwise relatedness ratings on the list of words and then, two weeks later, were asked to perform a false memory task. In an experiment conducted before Experiment 1 (see Appendix A; hereafter called Appendix Experiment), it became clear that the instructions for the recognition test should be modified so that participants would base their memory judgments only on memories for the study period and not the prior relatedness ratings session. Thus, in Experiment 1, this was tackled by modifying the study-test instructions to warn participants to base memory judgments on the current session only. In Experiment 2, the prediction that recognition tests based on individual semantic relatedness ratings (rather than association-based group norms) will lead to more or less false memory performance was tested. To that end, a more precise recognition test was tailored to each individual based on his/her relatedness ratings. The methods were also modified to parallel similar group-based manipulations used previously. In Experiment 3, only the individually-derived semantic relatedness of recognition test items was manipulated while the group based association strength was held constant. Finally, in Experiment 4, individual semantic ratings were used to construct study lists that were highly or weakly semantically related to theme items based on individual semantic ratings. A table in Appendix B summarizes the methodological details, results and additional information for each experiment.

# CHAPTER 2: EXPERIMENT 1

The first study extended the work done by Stevens (unpublished Master's Thesis, 2006). In her previous work, Stevens (unpublished Master's Thesis, 2006) asked participants to perform a false memory task followed immediately by a relatedness ratings task. It is possible that exposure to the false memory task might have influenced the participants' ratings in the relatedness ratings task. As such, this study reversed the order of the two tasks and addressed this concern. Participants performed the relatedness ratings task first and then, two weeks later, performed the false memory task. A period of two weeks was chosen based on Blair, Lenton and Hastie's (2002) work demonstrating that an individual's DRM performance was stable across a 2-week period. Blair and colleagues (2002) asked participants to study the same DRM lists initially and then two weeks later. They found that each participant's likelihood of false alarms stayed relatively consistent across the two testing sessions. Participants who were likely to falsely remember nonpresented theme words at time 1 were likely to falsely remember the same nonpresented theme words at time 2 (Blair et al., 2002). Thus, a two week delay between the tasks in the current studies should be sufficient to eliminate any influence of the completion of the first task to the completion of the second task.

## Method

### *Participants*

Thirty-five undergraduate students at the University of New Mexico participated in this experiment for class credit. None of the participants were involved in any of the other studies. Participants' reliabilities on the relatedness ratings task were computed by correlating their first and second ratings for 15 repeated pairs for each list. A reliability cutoff of $r = 0.40$ was used to eliminate participants with low reliabilities. Seven participants with low reliabilities were not included in the analysis. In addition, six participants who falsely recognized all theme words and two participants who did not complete the second part of the study were excluded from the analysis. As a result, twenty participants remained. The mean reliability across the analyzed sample was $r = 0.68$ (SD = 0.14).

### *Materials*

The materials were selected from the Nelson et al. (1998) word norms (see Appendix C). The list (theme word *music*) contained 24 words: the theme word and the 23 additional words that the theme word elicited on a free association task (Nelson et al., 1998). The study list was created by removing the theme word and every other word on the list (i.e., the words in serial position 2, 4, 6, 8, etc.). The recognition test consisted of all 24 words (including the theme word) on the *music* list.

### *Procedure*

Participants were tested individually in a small room with a personal computer. Participants were asked to rate the semantic relatedness of all possible word pairs created from the studied lists and their semantically associated recognition themes. Thus, word pairs were created from a set of 24 words resulting in 276 possible word pairs (24 choose 2). The word pairs were presented randomly. In addition, 15 pairs were rated a second time to serve as a reliability check. These 15 pairs were presented after the initial 276 word pairs were displayed and were chosen randomly. Participants rated the semantic relatedness for each pair of words on a scale from 1 to 5, with “1” being not at all related and “5” being highly related. Participants took approximately 20 minutes to complete the relatedness ratings.

Two weeks later, participants returned to the lab for the memory portion of the experiment. They were instructed to focus only on the second session and disregard what they did for the first session. This instruction was based on results from the Appendix Experiment that was conducted before Experiment 1 (see Appendix A for a detailed description). In this Appendix Experiment, participants first performed relatedness ratings on DRM list and, two weeks later, studied the DRM lists and then completed a recognition test. Unfortunately, the results of this study were uninformative as the likelihood of false memory was at ceiling. One explanation might be that participants were remembering words from the semantic relatedness task when they completed the false memory portion of the experiment two weeks later. In other words, exposure to the lists during the relatedness ratings increased familiarity with the words across the two week period (Atkinson & Juola, 1974) and increased the likelihood of false remembering on the false memory test. Consequently, the instructions for the current studies were changed when participants returned for the second part of the experiment to explain that they should disregard what they did for the first session and only focus on the second session.

During the study phase, words were presented to participants in a female voice at a rate of one word every 1.5 seconds. Participants were told to attend closely to the words in preparation for a memory test immediately presented at the end of the list.

During the memory test, words were presented one-at-a-time on a computer screen. Participants were told to press a key labeled “old” if the word had been studied on the just-presented list and to press a key labeled “new” if the word had not been previously studied. A practice list, consisting of the names of 12 U.S. states, was presented and tested at the beginning of the experiment. The study and recognition tasks took approximately 10 minutes to complete.

At the end of the experiment, participants were debriefed and thanked for their participation.

## Results and Discussion

Hits for studied words and false alarms for nonstudied words were calculated. The average hit rate for the studied words was 0.80. The average likelihood of false recognition was 0.47.

The primary analysis compared participants’ average relatedness ratings for falsely recognized and correctly rejected words. There were twelve words that could be falsely recognized (i.e., the nonpresented theme word and the eleven nonstudied list words). For each word that could be falsely recognized, the average relatedness rating given by each participant to all pairs containing the word was calculated. Then, for each list, the mean relatedness ratings for words that were falsely recognized were compared to words that were not falsely recognized (i.e., correctly rejected). This comparison required that participants falsely recognize at least one, but not all, of the candidate words. As mentioned previously, six participants falsely recognized all of the candidate words and thus were not included in the analysis. As shown in Table 1, a paired samples t-test on the mean relatedness ratings revealed that falsely recognized words had significantly higher ratings than correctly rejected words for all lists; *t* (19) = 3.04, *p*<.006, converging with the findings from Stevens (unpublished Master's Thesis, 2006). Thus, the false memory task did not appear to influence the relatedness ratings task since the current experiment replicated the previous work with a 2-week delay between the tasks.

**Table 1. Mean relatedness ratings (and standard deviations) for the nonpresented theme word that was falsely recognized and correctly rejected for Experiment 1**

| Theme word | Falsely recognized | Correctly rejected |
|---|---|---|
| **Music** | 3.29 (0.55) | 2.94 (0.53)* |

Note: *p<.006

# CHAPTER 3: EXPERIMENT 2

Experiment 1 found that individual semantic relatedness ratings could be used to predict false memory performance. A follow-up to that experiment would be to use these relatedness ratings to manipulate items on the recognition test. Experiment 2 manipulated items on the recognition test and, by doing so, pitted the two false memory theories against each other. The activation/monitoring theory states that the level of association between the nonpresented theme word and the other words on both the study and test lists are what determines later false recognition. Thus, the activation/monitoring theory would posit that a participant would be likely to falsely recognize the nonpresented theme word if it were preceded by the highest associates on the recognition test. The level of associations between the nonpresented words and other words on the list was obtained from the group normed data (see Nelson et al., 1998).

Conversely, the fuzzy-trace theory states that semantic relatedness between the nonpresented theme word and the other words on the study list is what determines false recognition. Thus, this theory would hypothesize that the nonpresented theme word would most likely be falsely recognized if it were preceded by the words that were most semantically related to it because this would cause the individual to be more likely to utilize a gist trace when evaluating whether a nonpresented word was studied or not. A measure of semantic relatedness can be obtained by asking participants to perform pairwise relatedness ratings (Cooke, 1992; Cooke, Durso & Schvaneveldt, 1986).

Previous research has found that the nonpresented theme word was more likely to be falsely recognized when it is preceded by highly associated studied words on the recognition test (based on group normed data) than when not preceded by highly associated words (Coane & McBride, 2006; Kimbell, Muntean & Smith, 2010). However, this manipulation may have confounded semantic relatedness and association; it is important to tease these two apart. In addition, Experiment 1 included only list words (and no unrelated words) on the recognition test which is not common in the literature; Experiment 2 addresses this by incorporating unrelated words on the recognition test.

The crucial manipulation in Experiment 2 involved unstudied items that appeared on the recognition test, mirroring past research. For instance, Marsh, McDermott and Roediger's (2004) participants studied false memory lists and then performed a recognition test. The recognition test contained false memory list words and their respective nonpresented themes that had *not* been studied in the previous phase. The researchers manipulated how many unstudied (that is, words that were not presented during the study phase) but related list words were presented prior to their respective theme word on the recognition test and found that when three unstudied words preceded their respective nonpresented theme word on a recognition test, the likelihood of false memories increased compared to when zero unstudied words preceded the theme word. Thus, in Experiment 2, following the procedure of Marsh et al. (2004), the manipulation on the recognition test was of unstudied items rather than studied items.

Two important methodological changes were made in Experiment 2. Participants were asked to make judgments about the quality of their memories to encourage more attention to the source of the memory. Tulving (1985) has suggested that it is possible to distinguish between two states of awareness: remembering and knowing. Remembering is having a detailed memory of the occurrence of a word including perceptual details of its presentation. Knowing is having a sense of the occurrence of a word but no detailed memory of its presentation. By asking participants to make a remember/know judgment for each recognized word, one is able to ascertain what sort of memory they have for each word they indicate was studied. McCabe and Geraci (2009) found that terminology and instructions given to participants describing experiences of remembering and knowing affected remember/know judgments. Thus, the current experiment used their suggested test instructions to obtain the most accurate report of remembering and knowing. This will help distinguish whether the participants actually misremember the experience of studying the words that they indicated as studied.

Finally, it was essential that the participants understood the directions for the relatedness ratings and were making semantic (not associative) ratings. While past data showing that highly associated items may be rated as weakly semantically related (e.g., opposites) supports the assertion that participants are making semantic ratings (Stevens, unpublished Master's Thesis, 2006), two additional instructional measures (described in the Procedure section below) were added to check that participants understood the task.

Given the previous study's findings that semantic information predicted false memory performance, it is hypothesized that the presentation of words that are highly semantically related to a nonpresented theme word prior to that theme word on a recognition test evokes higher levels of false recognition than presentation of weakly semantically related, but highly associated words, supporting the fuzzy-trace theory.

# Method

## *Participants*

Forty-eight undergraduate students at the University of New Mexico participated in this experiment for class credit. None of the participants were involved in any of the other studies. Participants' reliabilities on the relatedness ratings task were computed by correlating their first and second ratings for 15 repeated pairs for each list. A reliability cut-off of $r = 0.40$ was established to eliminate those with low reliability. Three participants with low reliability were excluded from the analysis. In addition, one participant with exceptionally fast reaction times (RT<300ms) and three participants who had higher false recognition of unrelated words than correct recognition of studied items were excluded from the analysis. As a result, forty-one participants remained. The mean reliability for these remaining participants was $r = 0.71$ (SD = 0.10).

## *Materials*

The materials consisted of thirteen DRM lists: nine lists (theme words *fruit, girl, needle, thief, shirt, butterfly, cottage, doctor* and *slow)* were presented during the study phase of the false memory portion of the experiment (termed 'Filler memory lists') and four lists (theme words *music, cold, sleep* and *chair*) were presented during the relatedness ratings and false memory portion of the experiment (termed 'Relatedness ratings lists'). Each of the lists consisted of the theme word along with the 15 words that elicited the theme word on a free association task. The lists were carefully chosen such that there was no thematic or word overlap and were taken from Roediger, Watson et al. (2001).

The recognition test consisted of 109 items. The 54 studied items were taken from the Filler memory lists (6 items per list). The 55 nonstudied items included the 9 nonpresented theme words from the Filler memory lists, 30 unrelated words taken from other nonstudied DRM lists and 16 items (3 list items and 1 theme per list) from the Relatedness ratings lists. Following Marsh et al. (2004), the theme words were presented after their respective list items for all the Relatedness ratings lists and Filler memory lists on the recognition test. An unrelated word was always tested immediately before each of the nonstudied theme words.

There were two conditions (High Semantic and High Associate) for the words on the recognition test. For the High Semantic condition, the nonpresented theme word was preceded by three highly semantically related words (i.e., ratings of 4 and 5) that were low in their association to the nonpresented word. For the High Associate condition, the nonpresented theme word was preceded by three words that were rated as weakly semantically related (i.e., ratings of 3 or below) but highly associated to that theme word as determined by the normed group data (see Nelson et al., 1998). In order to satisfy these conditions, the word lists, ordered from highest to lowest association to the theme word, were split in half. The High Semantic words consisted of the words that were low associates (i.e., the lower half of the list) and rated as highly related. The words for the High Associate condition were high associates (i.e., the upper half of the list) and rated as weakly related. For this experiment, these conditions were only applicable to the Relatedness ratings lists. For each participant, two of the lists were randomly assigned to the High Semantic condition and the other two were randomly assigned to the High Associate condition. The words used for the High Semantic and High Associate conditions are presented in Appendix D.

The twelve words selected based on the Relatedness ratings for inclusion on the recognition test differed across participants. For each participant, half of the words (i.e., 6 words; 2 study lists) were chosen based on the High Semantic condition and the other half of the words were chosen based on the High Associate condition. The ordering of the two conditions and the lists chosen for the two conditions were counterbalanced across participants.

### *Procedure*

The procedure was the same as described in Study 1 with the following exceptions. Participants completed 135 (16 choose 2 plus 15) pairwise ratings for each Relatedness ratings list. As in Experiment 1, each list also included 15 repeated pairs to establish reliability. Two additional instructional measures were added in order to ensure that the participants understood the directions for the relatedness ratings task. During the practice trial of the relatedness ratings task, the participants were told "It is important for you to make *semantic* ratings for the pairs of words that you are going to see. We are interested in which pairs of words you think are highly related and which pairs of words you think are unrelated.  For instance, you might think these two words *sweet* and *candy* are highly related so you would give this word pair a high semantic rating, like a 4 or 5.  However, these next two words *bitter* and *soda* you might think are not as related so you would give them a lower rating, like a 1 or 2.  Does that make sense?  Do you have any questions?"  The participants then practiced rating five word pairs.  In addition, prior to starting the relatedness ratings task, participants were shown all of the words on the word list and encouraged to mentally think about which word pairs they would rate as more semantically related and which word pairs they would rate as less semantically related.  This procedure provided a context for making the pairwise ratings and might improve the participants' reliability.  Participants took about 35 minutes to complete the relatedness ratings task.

When participants returned for their second session, the study words for the false memory portion were displayed visually on the screen for 1000 ms (in 40-point font in black letters) following Marsh et al.'s (2004) protocol.  In addition, participants were given the following instructions: "You may notice that the words in this part (Part 2) of the experiment are similar to Part 1.  Please disregard what you did for Part 1 and focus only on this session (Part 2)".  Furthermore, the participants completed memory judgments on the recognition test.  Per McCabe and Geraci's (2009) instructions, participants were told:

"If you indicate that a word has been studied, we will ask you additional information about your memory for that word.  Specifically, we will ask if you had a Type A memory or a Type B memory.

If you had a Type A memory, then that word brought to mind the exact thought you had when you first studied the word at the beginning of the experiment. If you can recall the exact thought you had from when you studied the word earlier you should indicate that you had a Type A memory. Often when people give a Type A response it is because they can recall a personal association that came to mind when they first saw the word, or some other details about when they studied the word. There are other details you may recall about studying a word that would lead you to make a Type A response, such as a particular feeling you had when you saw the word, or a mental image that came to mind while you were studying the word. You may also recall that you associated the word with another word that you studied, or you may recall what the word looked like on the screen.

If you had a Type B memory, then you believe that word was presented but you cannot recall any specific association that you made when you studied it. In other words, a Type B response means you "just know" you studied the word, even though you cannot recall any details from when you studied it.

In sum, if you can recall specific details about when you studied the word, you had a Type A memory. If you just know the word was presented but you cannot recall any specific details, you had a Type B memory."

Participants were asked to press the 'A' key if they had a Type A memory and to press the 'B' key if they had a Type B memory. A paper-based detailed explanation of the two different types of memory was available to the participants while they made their judgments. In addition, the following text was displayed on the screen while the participants made their judgments: “Type A = you can recall specific details about when you studied the word.  Type B = you just know the word was presented but you cannot recall any specific details”.

These memory judgments map on to the standard ‘remember/know’ paradigm used in past experiments; Type A memories corresponds to the ‘remember’ judgments and Type B memories corresponds to the ‘know’ judgments.

Participants took about 20 minutes to complete the false memory task.  After the false memory task, they were debriefed and thanked for participating.

## Results and Discussion

The changes made to the semantic relatedness ratings instructions resulted in fewer unreliable participants (3/48 participants) compared to the Appendix Experiment (8/57 participants) and thus appeared to be effective.

Hits for studied words, false alarms to unrelated words and likelihood of false recognition were computed.  The hit rate of studied words was 0.59 and the false alarms to unrelated words were 0.08.  The likelihood of false recognition for the four Relatedness ratings lists was 0.29 for *chair*, 0.41 for *cold*, 0.20 for *music* and 0.34 for *sleep*.  These numbers are lower than what Marsh et al. (2004) found but still somewhat surprising given that these nonstudied lists were not presented in the study phase.  The average likelihood of false recognition across the nine Filler memory studied lists was 0.59, which is comparable to past research (e.g., Roediger, Watson, et al., 2001).

The mean semantic relatedness ratings for the High Semantic and High Associate conditions are presented in Table 2.

**Table 2. Mean relatedness ratings (and standard deviations) for words used in the High Semantic and High Associate conditions for Experiment 2.**

| Experiment | High Semantic | High Associate |
|---|---|---|
| 2 | 4.80 (0.026) | 2.33 (0.076) |

False recognition on the two lists that incorporated the highest semantically related words (High Semantic condition) as studied items were compared to false recognition on the two lists that incorporated the weakly semantically related words (High Associate condition). A t-test revealed a non-significant trend for the themes in the High Associate condition to be more likely to be falsely remembered than the themes in the High Semantic condition, $t$ (40) = 1.817, $p$=.077; see Figure 1. This is in direct contrast to what was predicted, supporting the activation/monitoring theory rather than the fuzzy-trace theory.

The likelihood of false recognition for the High Associate and High Semantic conditions was also compared to false recognition of the unrelated words. The two conditions had a higher likelihood of false memory compared to the unrelated words; $t$ (40) = 4.94, $p$<.0001, which replicates Marsh et al.'s (2004) previous work.

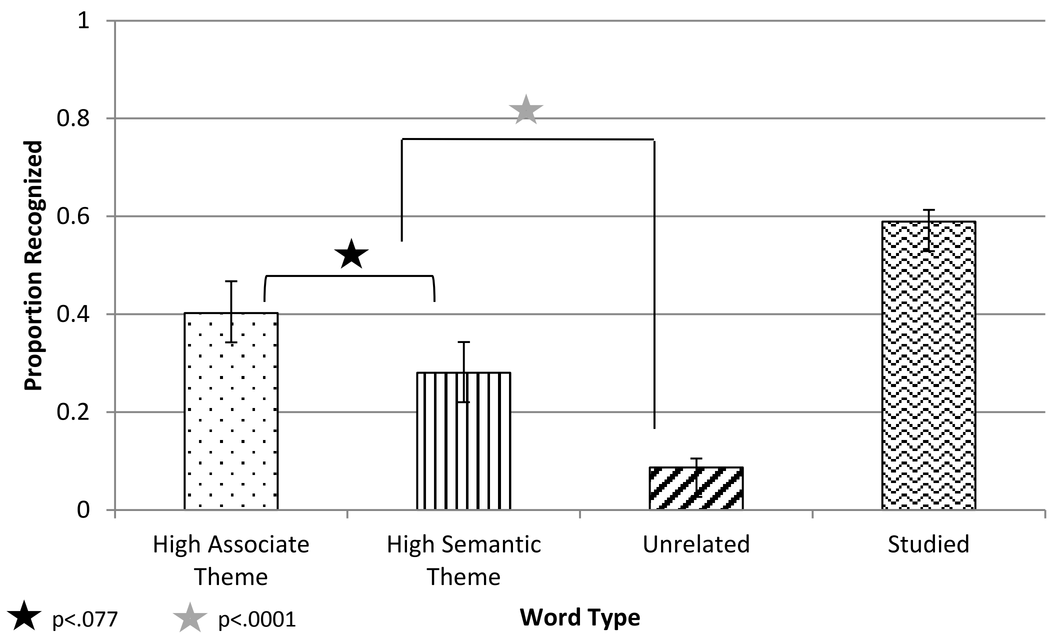



**Figure 1. Recognition of High Associate themes, High Semantic themes, Unrelated words and Studied Words (with standard error bars) for Experiment 2**

The memory judgments that participants attributed to their false memories were also analyzed. Remember that the participant chose between a Type A memory (having perceptual details about studying the word; corresponding to 'remember') and a Type B memory (just knowing that the word was studied but not having perceptual details about the experience; corresponding to 'know'). The total number of Type A versus Type B memory judgments for the High Semantic, High Associate and Studied Themes conditions is shown in Figure 2. The Studied Themes condition consisted of the nonpresented theme words from the nine Filler memory lists presented at the study phase of the experiment. Of particular interest was the comparison of memory attribution for the High Semantic and High Associate conditions. There was no significant difference between the two conditions for either Type A ($t$ (28) = .21, $p$=ns) or

Type B ($t$ (31) = .40, $p$=ns) memories. That is, on absolute levels, the High Semantic and High Associate conditions were equal on Type A and Type B memories.

Given that the participants saw only 3 words related to the nonpresented theme word on the recognition test (and that these 3 words were not presented during the study phase) for the High Semantic and High Associate conditions, it is not surprising that they did not have many false memories. However, it is interesting that they indicated that a portion of their false memories contained perceptual details considering that they should not have perceptual details for even the list words from the Relatedness ratings lists. It is unclear how these numbers correspond to other experiments using these instructions as no one has performed a false memory experiment using these instructions before (to the author's knowledge). However, these numbers are comparable to other studies that used the standard remember/know paradigm (e.g., Geraci & McCabe, 2006).

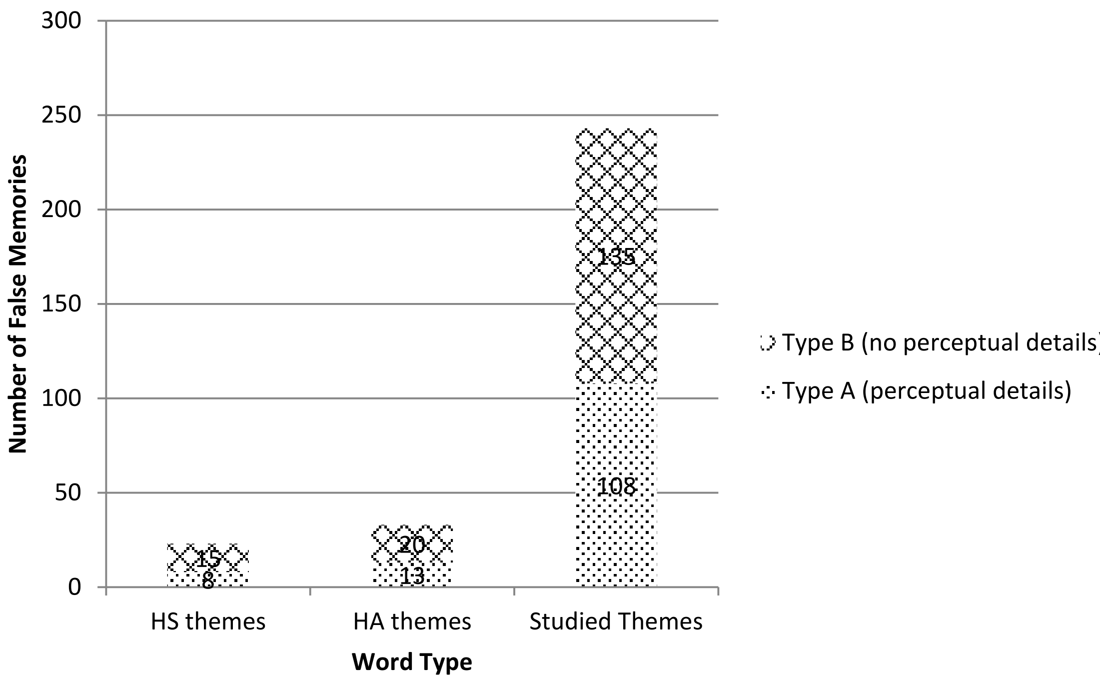



**Figure 2. Number of false memories attributed to a Type A and Type B memory for the High Semantic (HS), High Associate (HA) and Studied Lists (Studied Themes) conditions for Experiment 2**

Individual differences in accurate recognition of the studied items are often related to the likelihood of false memory. As such, a tertile split of the participants was done and the top third (N = 13) of participants with the best recognition performance of studied items were compared to those in the bottom third (N = 13). A 2 (Recognition Group: top third vs. bottom third) x 2 (Condition: HS vs. HA) ANOVA of false memory performance was executed (see Figure 3). There were no significant differences for either Condition or Recognition Group and there was no significant interaction.

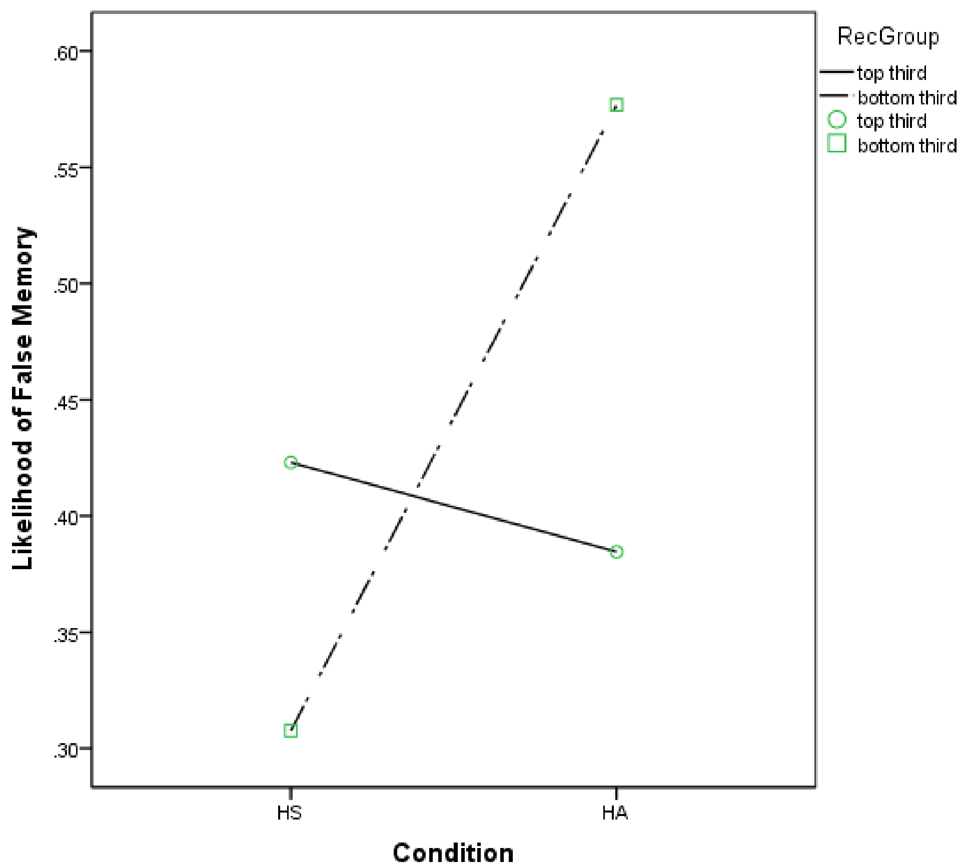



**Figure 3. Likelihood of false memory for participants with the best (top third) correct recognition of studied items compared to those with the worst (bottom third) correct recognition of studied items for High Semantic (HS) and High Associate (HA) conditions for Experiment 2**

The current experiment employed the testing methods used by Marsh et al. (2004) to assess false memory in individuals in the High Semantic versus the High Associate condition. The likelihoods of false memory for the two conditions were higher than the likelihood of false memory for unrelated words, replicating Marsh et al.'s (2004) work. The average likelihood of false memory for the Relatedness ratings lists for the current experiment (0.31) was slightly lower than what Marsh et al. (2004) found (0.49). It is unclear which false memory lists Marsh et al. (2004) used in their experiment so the differences in the lists used might explain the difference in the likelihood of false memory. In addition, the number of lists used in the experiments may explain the discrepancy; in the current experiment, the participants only studied nine lists while Marsh et al.'s (2004) participants studied eighteen or twenty-four lists. The fuzzy-trace theory would predict higher likelihood of false memory with an increase in the number of lists since the availability of verbatim traces would decline as the number of items to remember increases, leading the person to rely on gist-based traces. The likelihood of false recognition for the nine Filler memory lists was 0.29 for *butterfly*, 0.41 for *cottage*, 0.66 for *doctor*, 0.76 for *fruit*, 0.73 for *girl*, 0.85 for *needle*, 0.80 for *shirt*, 0.63 for *slow* and 0.78 for *thief*; average = 0.66. These percentages are generally in line with past research (e.g., Roediger, Watson et al., 2001). The hit rate of studied items (0.59) and unrelated words (0.08) in the current experiment was also lower than what Marsh et al. (2004) found (0.78 and 0.27 respectively). Again, this could be due to the differences in types of materials and number of lists used.

The critical comparison of the High Semantic and High Associate conditions resulted in a higher likelihood of false memory for the High Associate condition, which is in direct contrast to what was hypothesized. These results support the activation/monitoring theory. Thus, it seems that associative information, rather than semantic relatedness, is more influential in false memory performance. However, the current experiment explored only nonstudied words and did not keep the level of association constant across the two conditions. It might be that semantic relatedness is having an influence on false memory performance but the current design did not detect it. Thus, a different design was implemented in Experiment 3.

# CHAPTER 4: EXPERIMENT 3

The results from Experiment 2 did not support the hypothesis that semantic information would influence false memory performance; however, it is still possible that semantic relatedness is influencing the likelihood of false memory. While an effort was made to rule out the possibility that the level of association was affecting the results, it is important to note that the semantic relatedness ratings and the association strengths are measured with different levels of precision and are on two different scales. By manipulating both the individual semantic ratings and the association ratings of the previously tested words (i.e., by purposely picking low associates for the High Semantic condition and weak semantically related words for the High Associate condition) it is not clear which manipulation (or both?) is leading to the effect (or lack thereof). A manipulation of only the semantic relatedness (while keeping the level of association constant) might be a stronger test of whether semantic relatedness is predictive of false memory.

In addition, the design of Experiment 2 might not have been adequate to accurately test the hypotheses. The fuzzy-trace theory makes predictions about gist-based traces for studied items in false memory experiments, thus it is unclear what predictions the theory would make about the *unstudied* items used in Experiment 2 since it is uncertain how gist-based traces would be created or accessed for these items. Thus, the manipulation of semantic relatedness will focus on studied items in the current study. Coane and McBride (2006) manipulated the number of studied items (0, 6 or 12) that appeared before the nonpresented theme word on a recognition test.

They found an increase in false memory when six, but not zero, studied items preceded the nonpresented theme word (although see Kimball, Muntean & Smith, 2010 as these researchers did find an increase in false memories from zero to six studied items). Thus, the recognition test in Experiment 3 will consist of six studied items per list (rather than three unstudied items as used in Experiment 2).

In Experiment 3, the semantic relatedness of studied words on the test was manipulated and the average association of these manipulated test words was kept constant. For the High Semantic condition, words that participants rated as highly semantically related to the nonpresented theme word were chosen as studied items on the recognition test and these items were also highly associated to the nonpresented theme word per group norms. For the second condition, the Low Semantic condition, words that the participants rated as weakly semantically related to the nonpresented theme word were chosen as studied items on the recognition test and these items were also highly associated to the nonpresented theme word per group norms. Keeping the level of association constant (that is, by including highly associated words for both conditions) provided more evidence for whether an individual's rating of semantic relatedness was playing a role in false memory performance. This change in procedure will help reveal an effect if it is present. It is predicted that the High Semantic condition will lead to a greater likelihood of false remembering than the Low Semantic condition.

## Method

### *Participants*

Sixty-six undergraduate students at the University of New Mexico participated in this experiment for class credit. None of the participants were involved in any of the other studies. Participants' reliabilities on the relatedness ratings task were computed by correlating their first and second ratings for 15 repeated pairs for each list. A reliability cut-off of $r = 0.40$ was established to eliminate those with low reliability. Six participants with low reliabilities, seven participants who did not have enough variability in their relatedness ratings to meet the criteria for the High Semantic and Low Semantic conditions, one participant who had high false recognition for the unrelated items and five participants who did not complete the second part of the study were excluded from the analysis. In addition, eleven participants had to be excluded from the analysis due to a programming error. As a result, 36 participants were included in the analysis. The mean reliability for these remaining participants was $r = 0.67$ (SD = 0.12).

### *Materials*

The materials consisted of six lists (theme words *breakfast*, *cotton*, *sticky*, *summer*, *theater* and *tiger*). The lists consisted of a theme word and the 20 associates that the theme word elicited on a free association task (Nelson et al., 1998; see Appendix C). Per Coane and McBride (2006), the lists were studied in two blocks of three lists; each study list consisted of the 20 associates to the theme word. Each study-test block contained two Experimental lists (one from each condition as explained below) and one Control list and was followed by a recognition test. On the recognition tests, six study items from each Experimental list preceded the theme word (corresponding to Coane & McBride's six studied item condition); for Control lists, study items were presented following the theme word (corresponding to Coane & McBride's zero studied item condition). The ordering of these lists was counterbalanced across the participants.

There were two conditions for the Experimental lists on the recognition test. In one condition (termed the High Semantic condition), the studied words on the recognition test were selected to be words that the individual rated as highly semantically related to the theme word (i.e., ratings of 4 and 5 out of 5) and also high associates to the theme word according to group norms. In the second condition (termed Low Semantic condition), the studied words on the recognition test were selected to be words that the individual rated as weakly semantically related to the theme word (i.e., ratings of 3 and below out of 5) and also highly associated to the theme word according to group norms. For each participant, two of the Experimental lists were assigned to the High Semantic condition and the other two Experimental lists were assigned to the Low Semantic condition. For the Control condition, the study items were the six highest associates per the association norms. The lists chosen for the High Semantic, Low Semantic and Control conditions were counterbalanced across participants. The words used for the High Semantic, Low Semantic and Control conditions are displayed in Appendix E.

Each of the two recognition tests consisted of 38 items; 18 items from studied lists (6 from each Experimental list, 6 from the Control list) and 20 nonstudied items consisting of the three theme words from the studied lists and 17 unrelated items taken from other nonstudied DRM lists.

### *Procedure*

The procedure was the same as Experiment 2 with the following exceptions. During the first session, participants performed 225 relatedness ratings for each list (21 choose 2 plus 15 repeated pairs to establish reliability). The participants performed relatedness ratings for all six lists; they took approximately 50 minutes to complete the ratings.

Two weeks later during the study session, Coane and McBride's (2006) procedure was followed. The study lists were displayed visually (in 40 point black font) on the screen for 3000ms. Each of the two study blocks (3 lists per block) was followed by a 30 second distractor task consisting of mathematical problems. After the distractor task was completed, the recognition test was presented. This portion of the experiment took 20 minutes to complete. After the end of the experiment, participants were debriefed and thanked.

## Results and Discussion

The hit rate for the studied words, the false alarms to unrelated words and likelihood of false recognition of themes were computed. The hit rate of studied words was 0.92 and false alarms to unrelated words were 0.03.

The mean semantic relatedness ratings for words used in the High Semantic and Low Semantic conditions are presented in Table 3.

**Table 3. Mean relatedness ratings (and standard deviations) for words used in High Semantic and Low Semantic conditions for Experiments 3 and 4**

| Experiment | High Semantic | Low Semantic |
|---|---|---|
| 3 | 4.76 (0.026) | 2.60 (0.057) |
| 4 | 4.84 (0.018) | 2.41 (0.059) |

In manipulating the studied items for the High and Low Semantic conditions, it is important to double check that the BAS of the words in each condition are equivalent (otherwise, any results may be due to differences in BAS, which has already been shown to influence false memory performance). Thus, the average BAS values for words used for the High Semantic, Low Semantic and Control conditions for each list were calculated (see Table 4). The analysis shows that there was no difference in BAS between the High and Low Semantic conditions; $t\ (35) = 0.807$, $p$=ns. There was a significant difference in BAS values between the Control and HS condition ($t\ (35) = 3.31$, $p = .002$) and between the Control and LS condition ($t\ (35) = 5.80$, $p < .0001$) in which the Control condition had significantly higher BAS for both cases. This is not surprising given that the Control condition consisted of the highest associates per the association norms.

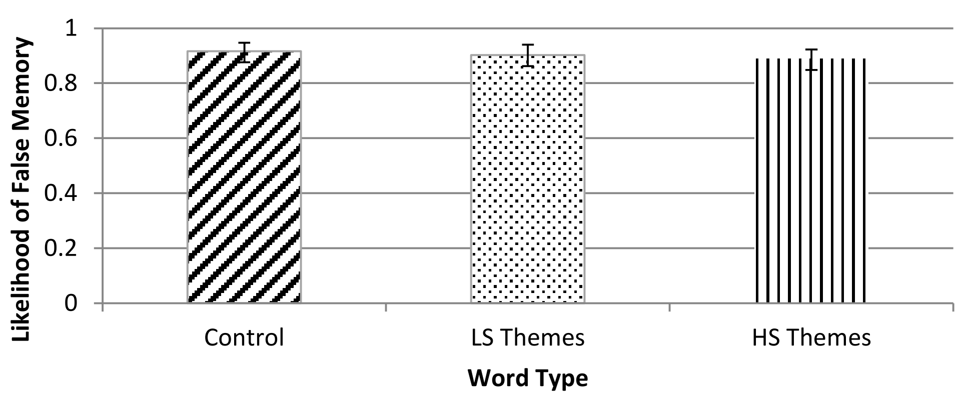
**Table 4. Average BAS values (and standard deviations) for words used in High Semantic, Low Semantic and Control conditions for Experiments 3 and 4**

| Experiment | High Semantic | Low Semantic | Control |
|---|---|---|---|
| **3** | 0.028 (.018) | 0.024 (.017) | 0.049 (.027) |
| **4** | 0.025 (.034) | 0.021 (.020) | 0.050 (.021) |

False recognition for the two lists that incorporated the highest semantically related words as studied items (High Semantic condition; 0.88) was compared to false recognition on the two lists that incorporated the weakly semantically related words (Low Semantic condition; 0.90). A t-test revealed no difference between the two lists ($t$ (35) = 0.37, $p$=ns). This is in contrast to what was predicted. It is important to note that this is independent of the degree of association between the studied words and theme words since the words chosen for the High and Low Semantic conditions were also high associates according to the group norms. There was a very slight trend for the Low Semantic condition to have higher likelihood of false recognition compared to the High Semantic condition but further analysis indicates that it would take 648 participants for this difference to be significant. These results support the activation/monitoring theory since this theory would predict no difference between the two conditions since the BAS values across conditions were equated.

The likelihood of false recognition for the High Semantic and Low Semantic condition was also compared to the likelihood of false recognition for the Control (0.91) condition. Once again, a t-test revealed no statistical differences between the Control condition and the High Semantic ($t$ (35) = 0.62, $p$=ns) or between the Control and the Low Semantic ($t$ (35) = 0.27, $p$=ns) conditions (see Figure 4).

The likelihood of false recognition for the two Experimental conditions increased as compared to Experiment 2, which is in line with past work. The likelihood of false memory for these two conditions was higher in this experiment (0.89) as compared to Coane and McBride (0.70; Experiment 1). The current experiment used different lists than Coane and McBride (2006) used which may explain the differences in the overall likelihood of false memory.  In addition, there was also a high likelihood of false recognition for the Control condition.  Thus, the contribution of semantic information from the studied items to false memory may have been so strong from the study phase that the additional manipulation on the test was not able to discriminate the Semantic conditions from the Control condition.  The finding that false memory for the Control condition did not differ from the other two conditions is different than what Coane and McBride (2006) found (but see their Experiment 1).  The high likelihood for false memory for the Control condition, in which the theme is presented before any of the list items on the recognition test, supports the idea that activation and/or gist representation of the theme word is occurring at study (see McDermott & Watson, 2001; Brainerd & Reyna, 2002).




Note: There are no significant differences

**Figure 4.  Likelihood of false memory of nonpresented theme words for Control, Low Semantic (LS) and High Semantic (HS) conditions (with standard error bars) for Experiment 3**

The memory judgments that participants attributed to their false memories was also assessed.  Recall that the participant chose between a Type A memory (having perceptual details about studying the word) and a Type B memory (just knowing that the word was studied but not having perceptual details about the experience). Once again the absolute number of Type A versus Type B memories was assessed. The attribution of Type A and Type B memories to the High Semantic, Low Semantic and Control conditions can be seen in Figure 5. There were no significant differences for memory attribution between any of the three conditions. However, for all conditions, there was more Type A false memories than Type B false memories. This is in contrast to what was found in Experiment 2 which is not surprising since the participants studied the pertinent word lists in the current experiment whereas they did not in Experiment 2.

When comparing the memory judgments with those studies incorporating the standard remember/know instructions (Geraci & McCabe, 2006), the memory judgments for Type A memories in the current work are higher than what has been found in the past. Given that McCabe and Geraci (2009) were able to decrease the proportion of Type A memories with their modified instructions in their paradigm, these numbers are rather surprising. However, there is one major difference between the two studies; in this experiment, the participants studied 3 lists of words before the presentation of the recognition test whereas in the Geraci and McCabe (2006) work, the participants studied 12 lists of words before being given the recognition test. The increase in list size might have decreased the participants' confidence in remembering perceptual details for the nonpresented theme word for Geraci and McCabe's (2006) study. Again, with an increased number of lists, the fuzzy-trace theory would predict a decrease in the availability of verbatim traces leading participants to rely on gist-based traces. Without the verbatim traces, participants might be more likely to indicate that their memory does not include perceptual details  Also, the current experiment did not offer a ‘guess’ or ‘not sure’ option for the memory judgments as did Geraci and McCabe (2006).  The lack of this option likely overinflated the Type A and Type B memories in the current experiment as participants would have been forced to choose between two options that may have been deemed invalid.

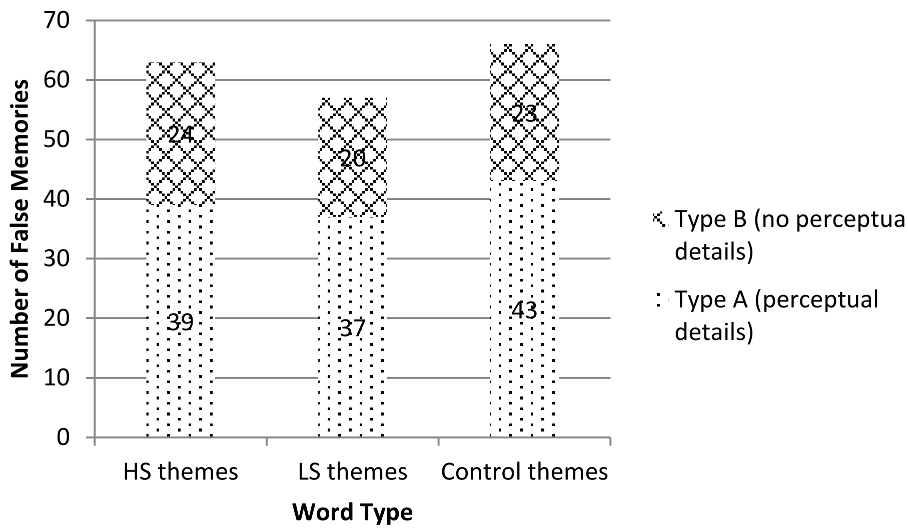



**Figure 5.  Number of false memories attributed to a Type A and Type B memory for nonpresented theme words for High Semantic (HS), Low Semantic (LS) and Control conditions for Experiment 3**

In addition, the likelihood of false memories for each of the word lists was evaluated.  Overall, the lists had fairly high likelihoods of false memory; the theme word *theater* had the highest likelihood of false memory (0.97), and the theme words *sticky* and *breakfast* had the lowest likelihood of false memory (0.83 and 0.84, respectively) and the other themes fell in between (*tiger* = 0.88, *summer* = 0.90, *cotton* = 0.94).  To the author’s knowledge, there are no known established norms on these particular lists so it is unclear if these percentages are reasonable or not.  In comparison to the established norms, the likelihood of false memory for the *breakfast and sticky* lists are comparable (see Roediger, Watson et al., 2001) but the other lists are well above what has already been established. This increase in false memory can be explained by the list lengths; the normed lists employed 15 study words while the current lists consisted of 20 study words. An increase in study words has been shown to result in an increase in false memories (e.g., Robinson & Roediger, 1997) which could explain the slightly higher likelihood of false memory in the current studies compared to past research. The fuzzy-trace theory would assert that studying 20 words might lead to an increased gist trace along with a decreased accessibility to verbatim traces for the entire set of items. The activation/monitoring theory would assert that studying 20 words would increase the activation of the theme word, leading to an increase in false memory performance.

Once again, a tertile split of the participants on accurate recognition of studied items was performed and the top third of participants (N=12) with the best correct recognition of studied items was compared to the bottom third of participants (N=12). A 2 (Recognition Group: top third vs. bottom third) x 2 (Condition: HS vs. LS) ANOVA of false memory performance was executed (see Figure 6). A significant effect of Recognition Group was found such that those in the top third had higher false memory rates than those in the bottom third for both conditions ($F$ = 9.32, $p$<.004). This is consistent with current understanding of the false memory phenomenon since participants who studied and remembered the studied items might be more prone to spreading activation of those words to the theme word (and lead to false remembering of the theme word), supporting the activation/monitoring theory.  There was no effect of Condition and no significant interaction.

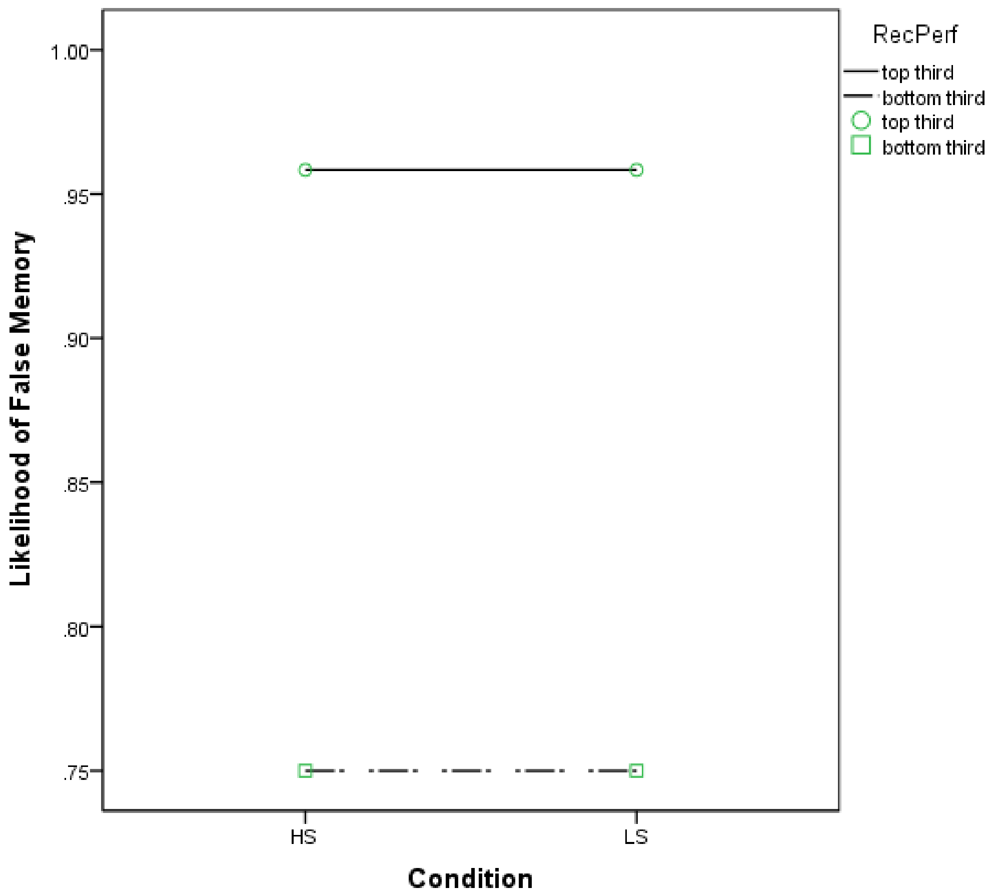



**Figure 6.  Likelihood of false memory for participants with the best (top third) correct recognition of studied items compared to those with the worst (bottom third) correct recognition of studied items for High Semantic (HS) and Low Semantic (LS) conditions for Experiment 3**

In summary, the current experiment did not replicate Coane and McBride's (2006) findings that presenting six words before the theme word on the recognition test resulted in a higher likelihood of false memory than presenting zero words (but see their Experiment 1).  There was no difference in false memory performance between the High Semantic and Low Semantic conditions, supporting the activation/monitoring theory.

The main difference between the studies was that the current study presented twenty words at study whereas Coane and McBride (2006) presented twelve. Perhaps presenting twenty list words is enough to dissolve the false memory differences between the Control condition and the Semantic conditions. This is not surprising since past research has found that, as the study list increases, so does the likelihood of false memory (Robinson & Roediger, 1997). This is also supported by the fact that the average likelihood of false memory (0.86) of the current study was higher than what Coane and McBride (2006) found (0.70) and the hit rate of studied items of the current study was greater for the current experiment (0.92) compared to what Coane and McBride (2006) found (0.80). Finally, memory judgments were found to be higher than in standard remember/know paradigms but it is unclear how these two different types of memory judgments can be compared since there are no established norms for the McCabe and Geraci (2009) modified instructions.

# CHAPTER 5: EXPERIMENT 4

The previous two experiments have manipulated the test phase of the experiment by varying the individual semantic relatedness of test items. However, past research has suggested that memory processes at both study **and** test influence false memories in the DRM paradigm (see Coane & McBride, 2006). For instance, some theories propose that activation processes that occur primarily at study combine with source monitoring errors to elicit high rates of false memory (McDermott & Watson, 2001). In addition, deeper processing at study can enhance rates of false recall perhaps because more meaningful encoding strengthens the association between items (Toglia, Neuschatz & Goodwin, 1999). During encoding, gist representations are processed and stored (Brainerd & Reyna, 2002). Finally, there is evidence that thematic priming can occur for studied and nonpresented theme words through the spreading of semantic activation during testing as well as during study (Kimball, Muntean & Smith, 2010). Evidence also suggests that retrieval processes may play a role in creating false memories. In their seminal paper, Roediger and McDermott (1995) analyzed the serial position of the free recall data and found that participants tended to falsely recall the nonpresented theme word toward the end of the recall session. This may indicate that prior recall of list words served as a cue for the theme word or that recalling the list items increased the probability that participants would recall the theme word as an item highly associated to studied items (Coane & McBride, 2006). The previous two experiments only incorporated individual manipulations based on semantic ratings at test. Since past research has suggested that processes at both study and test influence false memory, a manipulation at study should be informative. As such, Experiment 4 will further explore the effect of individual ratings on false memory performance by incorporating manipulations based on individual semantic ratings at the study phase of the experiment; the words presented at study will be based on participants' relatedness ratings. Given the high likelihood of false memory for the Control condition in the previous experiment, the study lists for Experiment 4 were shortened in hopes that the manipulation would be more effective.

In addition, it is important to determine whether the semantic relatedness ratings are capturing semantic relations between word pairs, especially given that the manipulation depends on semantic information. Given the earlier discussion on the overlap between semantic and associative features, this may not be the case. Great concern was taken to ensure that the participants understood the semantic nature of the relatedness ratings (the instructions explained semantic relatedness, gave an example, asked the participants to practice giving semantic judgments and gave the participants the opportunity to ask questions) but this does not guarantee that semantic information was obtained. As such, participants were asked to perform an associative ratings task at the end of the experiment in which they made associative judgments on the same word pairs that they rated on semantic relatedness. These two measures can then be analyzed to determine whether there are, in fact, differences in how participants rate semantic relatedness versus level of association for the false memory word pairs.

Once again, false memory performance was expected to be higher for the High Semantic condition compared to the Low Semantic condition, supporting the fuzzy-trace theory.

## Method

### *Participants*

Sixty-nine undergraduate students at the University of New Mexico participated in this experiment for class credit. None of the participants were involved in any of the other studies. Participants' reliabilities on the relatedness ratings task were computed by correlating their first and second ratings for 15 repeated pairs for each list. A reliability cut-off of $r = 0.40$ was established to eliminate those with low reliability. Nine participants who had low reliability, twelve participants who did not complete the second session, six participants who did not have enough variability in their relatedness ratings to meet the criteria for the High and Low Semantic conditions and five participants who had poor recognition performance were excluded from the analysis. As a result, 37 participants were included in the analysis. The mean reliability for these remaining participants was $r = 0.71$ (SD = 0.11).

### *Materials*

The materials used for the semantic relatedness ratings task, false memory tasks and associative ratings task included the same six lists from Experiment 3 (theme words *breakfast*, *cotton*, *sticky*, *summer*, *theater* and *tiger*; see Appendix C). For the false memory portion, the study items were determined by the participant's relatedness ratings. In the High Semantic condition, the six studied words were chosen because they were rated as highly semantically related to the theme word by that participant (i.e., ratings of 4 and 5 out of 5) and because they also were high associates to the theme word according to group norms. In the Low Semantic condition, the six studied words had been rated as weakly semantically related to the theme word (i.e., ratings of 3 and below out of 5) and also were highly associated to the theme word according to group norms. For each participant, two of the lists were assigned to the High Semantic condition, two of the lists were assigned to the Low Semantic condition and two of the lists were assigned to the Control condition. Each of the two blocks consisted of a list from each of the three conditions. For the Control condition, the study and test words consisted of the six highest associates (per the association norms). The assignment of lists to each condition was counterbalanced across participants.

The recognition tests consisted of 72 items: 36 that had been studied (6 from each list) and 36 nonstudied items that consisted of the six nonpresented theme words from the studied lists and thirty unrelated items from other unstudied DRM lists. The six studied words presented at test were the same six words that were presented at study. These six list words preceded the theme word on the recognition test for the High Semantic and Low Semantic conditions; the theme word preceded the six list words for the Control condition.

### *Procedure*

The procedure was similar to Experiment 3; participants first performed semantic pairwise ratings on all possible pairs for each of the six lists (225 total ratings for each list). The participants took about 50 minutes to perform the ratings task. Two weeks later, participants came back to the lab and performed the false memory task. The study lists were displayed visually (in 40 point black font) on the screen for 3000ms. Each of the two study blocks (3 lists each) was followed by a 30 second distractor task consisting of mathematical problems. After the distractor task was completed, the recognition test was presented.

Following the recognition test, participants were asked to perform an associative ratings task on four of the six word lists (the four lists were randomly determined). Only four lists were chosen in the interest of keeping the session under one hour so that further IRB approval was not needed. The associative ratings task was the same as the relatedness ratings task except that participants were asked to judge the level of association between the word pairs (with '1' being not at all associated and '5' being highly associated). As with the semantic relatedness ratings task, participants were given detailed instructions which explained association, provided an example of word pairs that are associated and provided an opportunity to ask questions or get clarification. The participants took about 50 minutes to perform the second session. At the end of the experiment, the participants were debriefed and thanked.

## Results and Discussion

Hits for studied words, false alarms to unrelated words and likelihood of false recognition to themes were computed. Hits to studied words were high (0.93) and false alarms to unrelated words were infrequent (0.02).

The mean semantic relatedness ratings used in the High Semantic and Low Semantic conditions are presented in Table 3.

In manipulating the studied items for the High and Low Semantic conditions, it is important to double check that the BAS of the words in each condition are equivalent (otherwise, any results may be due to differences in BAS, which has already been shown to influence false memory performance). Thus, the average BAS values for words used for the High Semantic, Low Semantic and Control conditions were calculated (see Table 4). The analysis shows that there was no difference in BAS between the High and Low Semantic conditions ($t$ (36) = 0.490, $p$=ns), suggesting that the BAS of the two conditions were in fact equal. There was a significant difference for BAS values between the Control and HS conditions ($t$ (36) = 3.32, $p$ = .002) and Control and LS conditions ($t$ (36) = 5.56, $p$ <.001) in which the Control condition had significantly higher BAS values in both cases. Again, this is not surprising since the Control condition consisted of the highest associates per the association norms.

The likelihood of false recognition for the three lists that incorporated the highest semantically related words (High Semantic condition) as studied and tested items were compared to the three lists that incorporated the weakest semantically related words (Low Semantic condition). A t-test revealed no significant difference between the two conditions ($t$ (36) = 0.59, $p$=ns). There was a slight increase in the likelihood of false memory for the High Semantic (0.56) versus the Low Semantic (0.52) condition but further analysis indicated that it would take 259 participants for this difference to be significant. Once again, the results are in support of the activation/monitoring theory since the theory would predict no difference in false memory performance between the two conditions since BAS between the two conditions was equated. Again, this is in contrast to what was predicted.

The likelihood of false memory for the High Semantic and Low Semantic conditions were also compared to the false memory performance for the Control (0.67) lists. T-tests revealed a significant difference between the Control and Low Semantic conditions ($t$ (36) = 2.22, $p$<.03) in which the Control condition had the greater likelihood of false memory (see Figure 7). Again, the current experiment did not replicate Coane and McBride's (2006) findings that presenting six items before the theme word on the recognition test results in higher likelihood of false memory compared to presenting zero items. However, the current study's result can be explained by the way in which the study lists were presented; for the Control list, the study list was presented in the order allocated by the association norms (arranged from highest associate to lowest associate) as has been done in previous research. However, it was impossible to order the High Semantic and Low Semantic lists according to semantic ratings because participants often listed all words as 5's. Thus, the study lists for the High and Low Semantic conditions were presented alphabetically (to ensure consistency across participants) instead of in order of semantic relation. The Control list was *not* presented alphabetically. In hindsight, the alphabetized order for the High and Low Semantic conditions likely allowed the participants to use alphabetization information to their advantage when making decisions on the recognition test. For instance, they could have used a 'recall-to-reject' strategy (Gallo 2004) because they were better able to recall all of the study words from certain parts of the alphabet and thus reject the nonpresented theme word if it fell in that part of the alphabet (i.e., the participant did not remember studying the theme word) or if it did not fall in that part of the alphabet (i.e., the participant did not remember studying a word from that particular part of the alphabet).  In order to assess whether this type of recall-to-reject strategy could have been used, for all participants, the study and theme words used for the High Semantic and Low Semantic conditions were evaluated.  Of particular interest was for how many participants the theme word fell outside of the parts of the alphabet that the studied words represented.  For 32 of the 37 participants, the theme word was the only word that started with its respective letter for at least one list in the High Semantic or Low Semantic condition.  Thus, 32 of the 37 participants could have used this recall-to-reject strategy.

This makes it difficult to compare the likelihood of false memories for the two Semantic conditions to the Control condition since two different strategies might have been implemented.  It is interesting; however, that the likelihood of false memory for both the High Semantic and Low Semantic lists was still relatively high.  Perhaps even with the distinctive information and the ability to 'recall-to-reject', six related words are enough to demonstrate the false memory phenomenon at a reasonable rate.  The false memory likelihoods of the three conditions are comparable to previous research (e.g., Roediger et al., 2005).

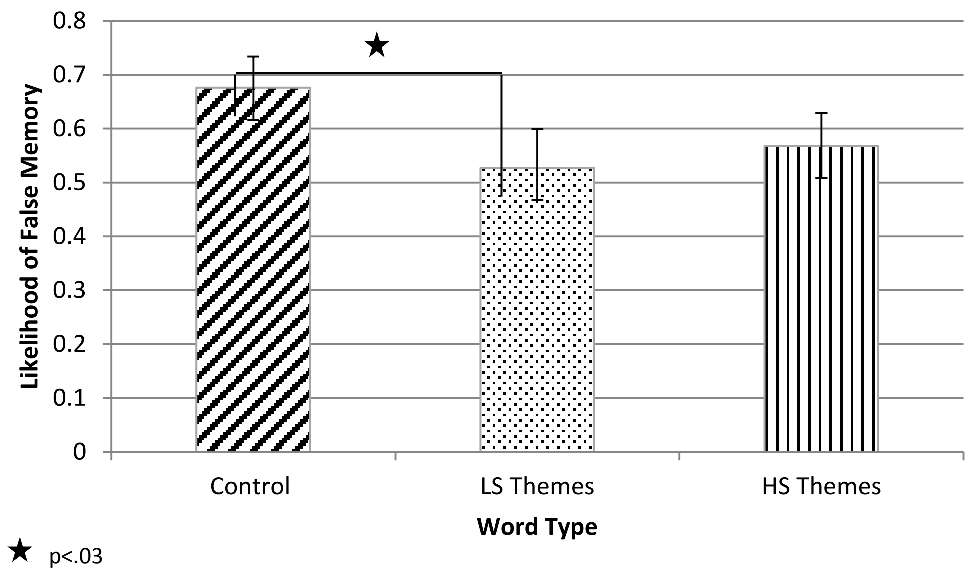



**Figure 7. Likelihood of false memory of nonpresented theme words for Control, Low Semantic (LS) and High Semantic (HS) conditions (with standard error bars) for Experiment 4**

The number of Type A and Type B memory judgments for falsely remembered words was also analyzed using the same procedure as in Experiment 3 (see Figure 8). The number of false memories attributed to a Type A ($t$ (48) = 0.922, $p$=ns) memory did not differ between High Semantic and Low Semantic conditions. However, there was a trend for the High Semantic condition to have more Type B memories ($t$ (51) = 1.43, $p$=.10) than the Low Semantic condition. That is, for the Low Semantic condition, fewer participants attributed their false memories to a Type B memory than in the High Semantic condition. This pattern of data suggests that the semantic manipulation might have influenced the memory judgments since those in the Low Semantic condition had a higher number of Type A compared to Type B memories, thus attributing a higher proportion of their memories to having perceptual details. Perhaps words that are studied and then tested prior to the theme word that also have low semantic relatedness have a different quality of false memory than those words that are highly semantically related. There was also a higher attribution of Type B memories for the Control condition compared to the Low Semantic condition; $t\ (54) = 1.92$, $p<.05$.

Overall, the number of Type A memories is lower than seen in Experiment 3. The differing results between the two experiments may be due to the variations in the study list lengths; Experiment 3 used 20 study words while Experiment 4 used only 6 study words. According to the false memory theories, studying 20 words would lead to greater overall activation and/or richer gist which may result in the participant attributing false memories of the nonpresented theme word to a memory with more specific/detailed information. The proportion of Type A memories was once again higher than in past research using the standard remember/know paradigm (Geraci & McCabe, 2006). Again, this could be due to the differences in the experimental design and number of lists studied along with not having a ‘guess’ option for the memory judgments.

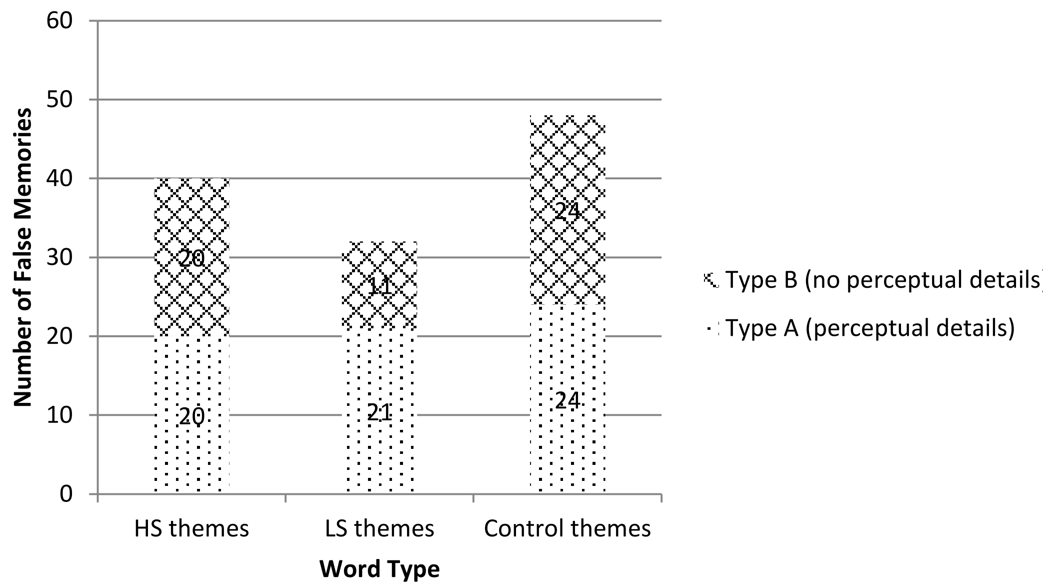



**Figure 8. Number of false memories attributed to a Type A and Type B memory for nonpresented theme words for High Semantic (HS), Low Semantic (LS) and Control conditions for Experiment 4**

The likelihood of false recognition (collapsed across conditions) did differ between lists with two lists having a fairly high likelihood of false memory (*theater* = 0.81; *cotton* = 0.68), one having a fairly low likelihood of false memory (*breakfast* = 0.31) and the remaining two being in between (*tiger* = 0.50; *sticky* = 0.51; *summer* = 0.63). Again, there are no established norms for these lists of words so it is uncertain as to whether these numbers are reasonable or not; however, they fall within the range of other normed false memory lists (see Roediger, Watson, et al., 2001). These rates are lower than seen in Experiment 3 but this is not surprising given that only six items were presented during the study phase. The fuzzy-trace theory would predict lower false memory rates for this experiment compared to Experiment 3 since the gist-based trace would not be as robust with six words compared to twenty leading to decreased false memory performance. The activation/monitoring theory would predict decreased overall activation with six words compared to twenty words leading to a decrease in false memory performance.

A few other analyses were performed to see if anything could be discovered from the data. In calculating the reliabilities, the average reliability (across the four lists) was used in determining whether or not a participant met the criteria for inclusion in the analysis. Another way of doing it, however, is to look at the reliability for the individual lists for each participant. In doing this, lists from four additional participants were included in the analysis. However, even when including these additional participants, there was still no significant difference between the High Semantic condition and the Low Semantic condition.

As before, a tertile split of the participants on accurate recognition of studied items was performed and the top third of participants (N = 12) with the best correct recognition of studied items was compared to the bottom third of participants (N = 12). A 2 (Recognition Group: top third vs. bottom third) x 2 (Condition: HS vs. LS) ANOVA of false memory performance was executed. There were no significant effects of Condition or Recognition Group. There was also not a significant interaction (see Figure 9).

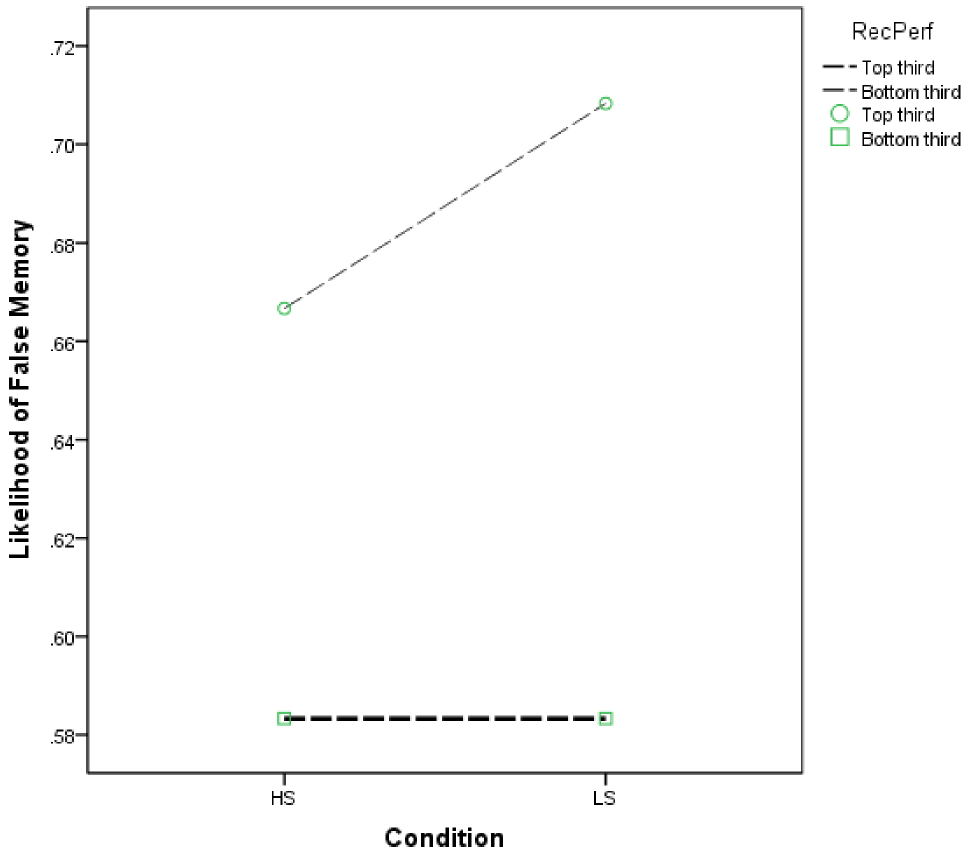



**Figure 9. Likelihood of false memory for participants with the best (top third) correct recognition of studied items compared to those with the worst (bottom third) correct recognition of studied items for High Semantic (HS) and Low Semantic (LS) conditions for Experiment 4**

Appendix F shows the frequency with which the words were used for the High Semantic, Low Semantic and Control conditions. For the Control condition, the top 6 associates (per the association norms) to the nonpresented theme word were used for both the study and test phase. It is interesting to note that a few of the list words were used for only the High Semantic or Low Semantic conditions (this was due to the participants rating them as such). Perhaps there are some words that participants agree are either highly semantically related or weakly semantically related to the nonpresented theme word. In terms of the High Semantic condition, for one of the lists (*sticky*), the word that was exclusively used for the High Semantic condition (*glue*) is the highest associate for that list. However, for two of the other lists, the words solely used for the High Semantic condition were not the highest associates (for the *breakfast* list, *omelet* is the 13$^{th}$ highest associate; for the *summer* list, *vacation* is the 6$^{th}$ highest associate, *season* is the 3$^{rd}$ highest associate and *heat* is the 7$^{th}$ highest associate; for the *theater* list, *popcorn* is the 9$^{th}$ highest associate).

For the Low Semantic condition, for three of the words lists, the words used exclusively were low associates (thus in agreement with the association norms). However, for the *tiger* list, several of the words rated as weakly semantically related were high associates per the association norms (*cougar* is the 5$^{th}$ associate, *panther* is the 3$^{rd}$ associate and *tame* is the 9$^{th}$ associate).

For the majority of the words, however, there was a fairly even split of whether they were used for the High Semantic or Low Semantic conditions. The fact that most of the words (even the low associates) were rated as highly semantically related indicates that the semantic relatedness ratings were capturing different information than the normative associations. It also strengthens the assertion that individual differences in semantic knowledge and semantic relatedness of these lists exists.

One of the crucial ideas of these experiments was that an individual's semantic knowledge could be derived from semantic relatedness ratings and that this measure would capture different information than the association based norms. To address this notion, both semantic and associative ratings were obtained from the participants. That is, the participants gave semantic ratings of the memory lists at the beginning of the experiment (which were used for the study and test manipulations) and then gave associative ratings to the same lists two weeks later. A measure of correlation might indicate the similarity of the two sets of ratings; the semantic and associative ratings were correlated for each participant and each list; see Appendix G. With the exception of two lists from one participant, all of the correlations were significant (at the 0.01 level). This suggests that the associative and semantic ratings are highly similar and could be measuring the same thing. However, earlier work (e.g., Stevens, unpublished Master's Thesis, 2006) suggested that the semantic/associative difference could be seen with opposites (such as *hot* and *cold*). Stevens (unpublished Master's Thesis, 2006) found that opposites tended to have lower semantic ratings than the association norms would suggest. Thus, in comparing the associative and semantic ratings, it is expected that the participants' associative ratings would be higher than their semantic ratings for the opposite pairs. Two of the lists in the current study contained opposites: *summer* & *winter* (from the *summer* list) and *breakfast* & *supper* (from the *breakfast* list). For the *summer* & *winter* pair (N = 18), the average associative rating was 3.55 and the average semantic rating was 3.22. For the *breakfast* and *supper* pair (N = 20), the average associative rating was 3.2 and the average semantic rating was 2.95. Thus, the same pattern that Stevens (unpublished Master's Thesis, 2006) discovered was also found in the current studies. This supports the idea that the semantic and associative ratings were measuring different things, at least in terms of opposites.

In summary, the current experiment again used Coane and McBride's (2006) false memory procedure of displaying six studied words before the nonpresented theme word on the recognition test. However, this study only presented six studied items in the study phase as well. There was no difference in false memory performance between the High Semantic and Low Semantic conditions, which is in contrast to what was predicted. These results are in support of the activation/monitoring theory, which would predict no differences in false memory performance between the two conditions since the conditions were equated in BAS. The current experiment did not replicate Coane and McBride's (2006) finding that presenting six studied words before the theme word on the recognition test resulted in a higher likelihood of false memory compared to when zero studied items are presented. The lack of this finding in the current study is likely due to the way in which the study lists were presented and that the participants could have used a 'recall-to-reject' strategy in the Control condition.

# CHAPTER 6: GENERAL DISCUSSION

Four experiments investigated false memory performance at an individual level. For all experiments, participants made semantic relatedness ratings to pairs of words from false memory lists and, two weeks later, performed a false memory task. In Experiment 1, the relatedness ratings were found to be predictive of later false memory performance; the theme words that were falsely recognized had significantly higher semantic relatedness ratings to other list words compared to the theme words that were correctly rejected. For Experiments 2-4, this information was used to manipulate test items on the recognition test and, for Experiment 4, study items were also manipulated. The individual semantic relatedness ratings did not predict false memory for Experiments 2-4; manipulating semantic relatedness at study or on the recognition test did not influence false memory performance compared to using associative information. While Experiment 1 found a significant difference in the likelihood of false memory when using semantic relatedness ratings, this difference is likely due to associative information that was captured by the ratings.

Two competing hypothesis of false memory were pitted against each other in that both associative information (supporting the activation/monitoring theory) and semantic information (supporting the fuzzy-trace theory) were included. Most of the results supported the activation/monitoring theory.

## Fuzzy-Trace Theory

The results were expected to be supportive of the fuzzy-trace theory of false memory since the participants' individual ratings of *semantic* information were used in the manipulation.  Contrary to the hypotheses, the results did not show much support for this theory.

## Activation/Monitoring Theory

Most of the results from the current experiments support the activation/monitoring theory.  In Experiment 2, the High Associate condition had a higher likelihood of false memory than the High Semantic condition, supporting the idea that level of association has a greater influence on false memory than semantic information.  There was no difference in false memory performance between the High Semantic and Low Semantic conditions in Experiments 3 and 4, supporting the activation/monitoring theory since the theory would predict no difference since the two conditions were equated on BAS.

## Comparison to Previous Research

The current work included paradigms used by Stevens (unpublished Master's Thesis, 2006), Marsh et al. (2004) and Coane and McBride (2006).  The first experiment replicated the results found in Stevens (unpublished Master's Thesis, 2006) with a procedure gathering the relatedness ratings before the memory test; the theme words that were falsely recognized had significantly higher semantic relatedness ratings than those theme words that were correctly rejected.  Experiment 2 followed Marsh et al.'s (2004) work in which the likelihood of false memory for theme words from *unstudied* lists was of interest.  Experiment 2 replicated Marsh et al.'s (2004) finding that presenting three unstudied words before the theme word on a recognition test resulted in a higher likelihood of false memory compared to presenting zero unstudied words before the theme word. Experiments 3 and 4 employed the procedure used by Coane and McBride (2006) in which six studied items were presented before their respective theme word on the recognition test. The experiments did not replicate Coane and McBride's (2006) main finding that presenting six study items before the nonpresented theme word resulted in an increase of false memory as compared to when zero study words were presented before the theme word (although Experiment 3 did replicate their Experiment 1). While this is surprising, the lack of replication might be due to the differences in the materials used. The study list for Experiments 3 employed 20 words while the study lists for Coane and McBride's (2006) work used 12 words. It is possible that the increase in study words resulted in stronger gist-traces during encoding and that the contribution of semantic information from the studied items to false memory may have been so strong from the study phase that the additional manipulation on the test was not able to discriminate the Semantic conditions from the Control condition. In addition, the DRM lists used for Experiments 3 and 4 differed from those used by Coane and McBride and it is possible that this also had an influence. For Experiment 4, the lack of replication for Coane and McBride's (2006) study might be due to the alphabetized presentation of list items in the Low and High Semantic conditions and the possibility that participants used this alphabetization to make decisions on the recognition test. The list items in the Control condition were not alphabetized, thus, it is difficult to compare the Control condition with the High and Low Semantic conditions since participants might have used different strategies in determining whether a word had been studied on the recognition test.

## Source Monitoring

A body of false memory literature has looked at the role of source monitoring in false memories which might have played a role in the current experiments. This literature suggests that when participants are unable to remember source information (e.g., whether memories were internally generated or externally presented) of the items, they falsely remember the critical theme words (see Johnson, Hashtourdi & Lindsay, 1993 for an overview of the framework). Previous work has shown that asking participants to engage in source monitoring (e.g., paying close attention to source information of externally presented items) decreases the likelihood of false memory. Dewhurst, Knott and Howe (2011) asked participants to engage in source monitoring in three different experiments; participants were asked to give remember/know judgments, were asked to remember whether a male or female presented the list words or were presented both studied and unstudied list items on the recognition test. In all of these conditions, participants were required to engage in source monitoring to complete the recognition test. These researchers were interested in an effect termed 'test-induced priming' which involves inducing false memories at retrieval by manipulating the number of studied items that precede the critical theme words on the recognition test (Dewhurst et al., 2011). The researchers found decreased false memory performance in all three experimental conditions compared to participants who were simply asked to make old/new judgments. Interestingly, the researchers found this decrease in false memory performance for the source monitoring participants only in lists that had been previously studied but not for unstudied lists.

The present studies replicate the results from Dewhurst et al. (2011). Experiment 2 replicated their findings that, *for unstudied* lists, the test-induced priming effect was present. That is, even though the participants were instructed to engage in source monitoring by making remember/know judgments, there was a significant increase in false memory performance for themes that were preceded by three unstudied list words on the recognition test compared to themes that were preceded by no unstudied list words on the recognition test. This result was also what Marsh and colleagues (2004) found.

Experiment 3 and 4 replicated Dewhurst and colleagues (2011) finding that, *for studied* lists, source monitoring reduced the effects of test-induced priming. That is, when participants were instructed to engage in source monitoring by making remember/know judgments, there was no difference in false memory performance for those themes that were preceded by studied list items on the recognition test compared to those themes that were preceded by zero list items on the recognition test. Coane and McBride (2006) did not ask their participants to perform remember/know judgments which may explain why they found a test-induced priming effect.

## Individual Differences

Previous studies looking at false memory phenomenon took a nomothetic approach by manipulating variables at the group level. However, Underwood (1975) proposed that a critical test of any theory is to determine whether individual variation in the proposed construct correlates with performance. Before this series of studies (and

Stevens, unpublished Master's Thesis, 2006), predicting false memory performance at the *individual* level had yet to be done. These studies were the first to evaluate false memory performance at the individual level using individual semantic relatedness information. While there was variability in the semantic relatedness ratings given by individual participants and variability in individual false memory performance, suggesting that individual differences are present, the current results do not provide much support at the individual differences level. Given the substantial evidence in support for the false memory theories, the results suggest methodological considerations and limitations in the current studies, which are discussed next.

## Evaluation of Methods

The current studies were the first known experiments to use individual ratings to predict or manipulate false memory performance. As such, it is important to discuss the successes and limitations of this unique approach.

First, it is important to assess the validity of the ratings. Pairwise ratings have been shown to be a valid measure of semantic information (Cooke, 1992) for at least one of the six types of semantic relations as described earlier (Wu & Barsalou, 2007; as cited in Howe, et al., 2009). Past research using pairwise ratings has focused mostly on the taxonomy relation including types of animals (Cooke, 1992), statistical design concepts (Goldsmith, Johnson & Acton, 1991) and mathematical concepts (Johnson, Goldsmith & Teague, 1994). The false memory lists used in the current studies incorporated not only the taxonomy relation (e.g., *lion* and *bear*) but all of the other five relations; antonymy (e.g., *breakfast* and *dinner*), entity (e.g., *tiger* and *stripe*), introspective (e.g., *sleep* and *peace*), situational words (e.g., *summer olympics*) and synonymy (e.g., *theater* and *cinema*). In addition, all of the lists used in the current experiments consisted of at least two relations, supporting the idea that semantic relations include multiple, overlapping categories. Thus, while pairwise semantic relatedness ratings have been shown valid for studies using the taxonomy relation, it is conceivable that they are not valid for the other types of semantic relations.

In performing pairwise ratings in the current experiments, participants were given detailed instructions explaining what the ratings entailed, were given the ability to ask questions and indicated that they understood the instructions before continuing with the experiment suggesting that participants understood the directions. However, 'semantic relatedness' (or which semantic relation was of most importance) was not clearly defined and might have been interpreted differently across participants. Future studies may involve describing the pertinent semantic relation(s) in the instructions so that participants know what exactly is meant by ratings of semantic relatedness. The ratings seemed to capture different information than what the associative norms would suggest. For instance, some very highly associated words (e.g., *hot*) that were opposites of their respective theme word (e.g., *cold*) were given relatively low semantic ratings (e.g., a 2.5 out of 5). However, the semantic and associative ratings obtained in Experiment 5 were found to be highly correlated which is a concern for the manipulations used in the present experiments. The ratings in the current studies were based on a 1 to 5 scale which, while common, did not lead much room for variability which was especially crucial in the current experiments. Future studies could involve a more variable scale (such as 1 to 7 or 1 to 10 scale).

The presentation of the rating pairs might have influenced the ratings. The pairs were presented such that the words were side by side (e.g., *Cold – Snow*) and, while not intentional, it is possible that the participant might have seen the left word (*Cold*) as a cue for the right word (*Snow*) when considering the relatedness of the words. This presentation might have directed participants to interpret one word as a cue for the other leading to a different type of cognitive processing of the word pairs than what was intended. Since associations are often obtained by giving a cue word and asking for a target word, the ratings may have also captured associative relatedness rather than purely semantic information. Further, the first word in the pair was sometimes the theme word and other times a studied word and thus the type of association information influencing the ratings may have varied. If this was indeed happening, then the manipulation of the semantic relatedness based on the ratings was not valid. Future studies could include a different kind of rating, such as a controlled association task, to obtain semantic relatedness. In a controlled association task, participants are given a word and asked which other words on the list are highly related to it. This sort of relatedness procedure might have alleviated some of the problems incurred in the current studies since it is generally quicker to perform and participants are shown many words on which to base their relatedness judgments which may eliminate the propensity to rate cue-target associative relation between only two words.

Additionally, the presentation of the ratings task before the false memory task could have had an influence on false memory performance.  While an individual's DRM performance has been found to be stable across a 2-week period (Blair & colleagues, 2002), there was no check to make sure that this was true for the current experiments. A follow-up study could include false memory lists that are presented during the false memory portion but were **not** included in the semantic relatedness ratings portion two weeks before.  The likelihood of false memory for these unrated lists could be compared to the likelihood of false memory for lists that were rated two weeks before.

Second, it is important to assess the overall methodology of the approach.  The design and methodology used in Experiment 1 was easy to execute and would be easy to replicate.  It took relatively little time to run the experiment and there was minimal work on the experimenter's part.  The other experiments fine-tuned the idea of assessing individual false memory performance by manipulating items at the individual level.

The two week delay, while necessary for the experiments, increased the difficulty of the studies in several ways.  First of all, the time necessary to complete one experiment was twice that of a study with only one study session.  This is certainly an issue for researchers wanting to conduct research and subsequently publish in a timely fashion.  In addition, the participant drop-out rate for the second session proved problematic; the current experiments had to run double the number of desirable participants to ensure a sufficient number for the data analysis.

The manipulation using the participants' relatedness ratings proved to be time consuming on the experimenter's part. Since the manipulation was based on individual ratings, the experimenter was required to sift through each person's ratings and determine which words would be used for their portion of the experiment. In addition, since each participant was given a unique test, the experimenter was required to create a different program for each participant resulting in possible confusion for the numerous researchers running the study.

Given the drawbacks found in the methods of the current experiments, it is recommended that future studies make changes to this approach (or perhaps change it entirely). The study of individual differences is still an important topic to consider. Future studies could involve further exploring and fine-tuning the ideas laid out in these current experiments.

## Future Studies

One potential next step in this line of work would be to try to replicate Ceci and colleague's (2007) study using adults and the standard DRM paradigm (not the misinformation effect). The authors found that the misinformation effect was influenced by semantic proximity in children. It would be informative to address whether these results are constrained to children and the misinformation effect or whether they can be extended to adults in the standard DRM paradigm since it is possible that semantic information is particularly useful for children but that adults rely more on associative information in false memory studies. The follow-up study could include Ceci et al.'s (2007) triad items which seemed to include taxonomic semantic relations (such as types of animals, types of foods, etc.) or could include other false memory lists that focus on taxonomic information (such as the list with the theme word *fruit*) keeping in line with past DRM false memory research. These ratings could be used to construct and manipulate recognition tests. As in the present experiments, associative information could also be included to further test the influence of semantic information.

Another future study could also include using a different method to collect semantic information other than the pairwise relatedness ratings used in the present experiments. As mentioned previously, a controlled association task collects semantic information by asking participants to indicate which words are most related. For this task, participants are able to compare the word of interest with every other word on the list at one time and may eliminate the drawbacks from the pairwise ratings in the current experiments. This method may better tap into whether semantic information influences false memory performance.

Another potentially fruitful approach would involve obtaining both semantic and associative information from the participants. One of the disadvantages of the current studies is that the semantic information was obtained for each individual participant but the associative information was obtained from the group norms. It is hard to compare these two sets of information since they are on different scales and measured with difference levels of precision. Obtaining both sets of information from the participants would allow the researcher to directly use and compare the semantic and associative information in the false memory portion of the experiment. However, given the overlap of semantic and associative information found in past studies and in the current experiments, the study would have to take tremendous efforts to ensure that the semantic and associative information gathering did not influence each other.

Finally, the distinction between semantic and associative information should be further assessed and tested. Given the various types of semantic information and the problems with teasing apart semantic and associative information (Hutchison, 2003), more research is needed to understand this distinction. The current experiments also found a large amount of overlap between participants' semantic and associative ratings. The one semantic relation that seems to differ from associative information is the antonymy relation (i.e., opposites). Both Stevens (unpublished Master's Thesis, 2006) and the current studies found the trend for opposite word pairs to have high associative ratings but low semantic ratings. Since the current experiments included lists that contained all six types of semantic relations, this may have muddied the waters and made it difficult to find a significant effect if one existed. As it stands, it is unclear exactly what type(s) of semantic relations the fuzzy-trace theory would predict as having an influence on false memory. Follow-up studies focusing on only one type of semantic relation might shed some light on the influence of semantic information on false memory performance.

# REFERENCES


Atkinson, R. C. & Juola, J. F. (1974). Search and decision processes in recognition memory. In D. H. Kratz, R. C. Atkinson, R. D. Luce & P. Suppes (Eds.) *Contemporary Developments in Mathematical Psychology: I. Learning, Memory and Thinking.* W. H. Freeman: Oxford, England.

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

Blair, I. V., Lenton, A. P. & Hastie, R. (2002). The reliability of the DRM paradigm as a measure of individual differences in false memories. *Psychonomic Bulletin & Review, 9(3),* 590-596.

Brainerd, C. J. & Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science, 11(5),* 164-169.

Brainerd, C.J., Reyna, V. F. & Kneer, R. (1995). False-recognition reversal: When similarity is distinctive. *Journal of Memory and Language, 34(2)*, 157-185.

Brainerd, C. J., Yang, Y., Reyna, V. F., Howe, M. L. & Mills, B. A. (2008). Semantic processing in “associative” false memory. *Psychonomic Bulletin & Review, 15(6),* 1035-1053.

Butler, K. M., McDaniel, M. A., Dornburg, C. C., Price, A. L. & Roediger, H. L. (2004). Age differences in veridical and false recall are not inevitable: The role of frontal lobe function. *Psychonomic Bulletin & Review, 11(5),* 921-925.

Castel, A. D., McCabe, D. P., Roediger, H. L., & Heitman, J. L. (2007). The dark side of expertise: Domain-specific memory errors. *Psychological Science, 18(1)*, 3-5.

Ceci, S. J., Papierno, P. B. & Kulkofsky, S. (2007). Representational constraints on children’s suggestibility. *Psychological Science, 18(6),* 503-509.

Chi, M. T. H. & Koeske, R. D. (1983). Network representation of a child's dinosaur knowledge. *Developmental Psychology, 19(1)*, 29-39.

Coane, J. H. & McBride, D. M. (2006). The role of test structure in creating false memories. *Memory and Cognition, 34(5),* 1026-1036.

Collins, A. M. & Loftus, E. F. (1975). A spreading activation theory of semantic processing. *Psychological Review, 82(6)*, 407-428.

Cooke, N. J. (1992). Predicting judgment time from measures of psychological proximity. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18(3)*, 640-653.

Cooke, N. J., Durso, F. T & Schvaneveldt, R. W. (1986). Recall and measures of memory organization. *Journal of Experimental Psychology: Learning, Memory and Cognition, 12(4)*, 538-549.

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

Dewhurst, S. A., Knott, L. M. & Howe, M. L. (2011). Test-induced priming impairs source monitoring accuracy in the DRM procedure. *Journal of Experimental Psychology: Learning, Memory and Cognition, 37(4),* 1001-1007.

Dewhurst, S. A. & Robinson, C. A. (2004). False memories in children. *Psychological Science, 15(11),* 782-786.

Fernandes, M., Ross, M., Wiegand, M. & Schryer, E. (2008). Are the memories of older adults positively biased? *Psychology and Aging, 23(2)*, 297-306.

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

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

Geraci, L. & McCabe, D. P. (2006). Examining the basis for illusory recollection: The role of remember/know instructions. *Psychonomic Bulletin & Review, 13(3)*, 466-473.

Goldsmith, T. E., Johnson, P. J. & Acton, W. H. (1991). Assessing structural knowledge. *Journal of Educational Psychology, 83(1),* 88-96.

Hicks, J. L. & Hancock, T. W. (2002). Backward associative strength determines source attributions given to false memories. *Psychonomic Bulletin & Review, 9(4)*, 807-815.

Howe, M. L., Wimmer, M. C., Gagnon, N. & Plumpton, S. (2009). An associative-activation theory of children's and adults' memory illusions. *Journal of Memory and Language, 60,* 229-251.

Huff, M. J. & Hutchison, K. A. (2011). The effect of mediated word lists on false recall and recognition. *Memory and Cognition, 39(6),* 941-953.

Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A *micro*analytic review. *Psychonomic Bulletin & Review, 10(4)*, 785-813.

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

Jacoby, L. L. & Rhodes, M. G. (2006). False remembering in the aged. *Current Directions in Psychological Science, 15(2),* 49-53.

Johnson, M. K., Hashtroudi, S. & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114(1),* 3-28.

Johnson, P. J., Goldsmith, T. E. & Teague, K. W. (1994). Locus of the predictive advantage in Pathfinder-based representations of classroom knowledge. *Journal of Educational Psychology, 86(4)*, 617-626.

Kimball, D. R., Muntean, W. J. & Smith, T. A. (2010). Dynamics of thematic activation in recognition testing. *Psychonomic Bulletin & Review, 17(3)*, 355-361.

Lampinen, J. M., Neuschatz, J. S., & Payne, D. G. (1999). Source attributions and false memories: A test of the demand characteristics account. *Psychonomic Bulletin & Review, 6*, 130-135.

Markman, A. B. (2002). Knowledge representation. In D. Medin & H. Pashler (Eds.) *Stevens' Handbook of Experimental Psychology (3rd ed.) Volume 2: Memory & Cognitive Processes (p. 165-208).* John Wiley & Sons, Inc.: New York.

Marsh, E. J., McDermott, K. B. & Roediger, H. L. (2004). Does test-induced priming play a role in the creation of false memories? *Memory, 12(1)*, 44-55.

Mather, M. & Johnson, M. K. (2000). Choice-supportive source monitoring: Do our decisions seem better to us as we age? *Psychology and Aging, 15(4)*, 596-606.

McCabe, D. P. & Geraci, L. D. (2009). The influence of instructions and terminology on the accuracy of remember-know judgments. *Consciousness and Cognition, 18*, 401-413.

McCabe, D. P. & Smith, A. D. (2002). The effects of warnings on false memories in young and older adults. *Memory and Cognition, 30(7)*, 1065-1077.

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

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

McEvoy, C. L., Nelson, D. L. & Komatsu, T. (1999). What is the connection between true and false memories? The differential roles of interitem associations in recall and recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition, 25(5)*, 1177-1194.

McKelvie, S. J. (1999). Effect of retrieval instructions on false recall. *Perceptual and Motor Skills, 88*, 876-878.

McRae, K. & Boisvert, S. (1998). Automatic semantic similarity priming. *Journal of Experimental Psychology: Learning, Memory and Cognition, 24*, 558-572.

Metzger, R. L., Warren, A. R., Shelton, J. T., Price, J., Reed, A. W. & Williams, D. (2008). Do children "DRM" like adults? False memory production in children. *Developmental Psychology, 44(1)*, 169-181.

Meyer, D. E. & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of dependence between retrieval operations. *Journal of Experimental Psychology, 90(2)*, 227-234.

Moss, H. E., Ostrin, R. K., Tyler, L. K. & Marslen-Wilson, W. D. (1995). Accessing different types of lexical semantic information: Evidence from priming. *Journal of Experimental Psychology: Learning, Memory & Cognition, 21*, 863-883.

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

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

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

Roediger, H. L., Balota, D. A. & Watson, J. M. (2001). Spreading activation and arousal of false memories. In H. L. Roediger, J. S. Nairne, I. Neath & A. M. Surprenant (Eds.), *The Nature of Remembering: Essays in honor of Robert G. Crowder (pp. 95-115)*. American Psychological Association: Washington, DC.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21*, 803-814.

Roediger, H. L., Watson, J. M., McDermott, K. M. & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review, 8(3)*, 385-407.

Seamon, J. G., Luo, C. R., & Gallo, D. A. (1998). Creating false memories of words with or without recognition of list items: Evidence for nonconscious processes. *Psychological Science, 9(1),* 20-26.

Stevens, S. M. (2006). Predicting false memories from semantic networks. Unpublished Master's Thesis.

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

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

Underwood, B. J. (1975). Individual differences as a crucible in theory construction. *American Psychologist, 30(2)*, 128-134.

Waldie, B. D. & See, S. T. K. (2003). Remembering words never presented: False memory effects in dementia of the Alzheimer type. *Aging Neuropsychology and Cognition, 10(4)*, 281-297.

Watson, J. M., McDermott, K. B. & Balota, D. A. (2004). Attempting to avoid false memories in the Deese/Roediger-McDermott paradigm: Assessing the combined influence of practice and warnings in young and older adults. *Memory and Cognition, 32(1),* 135-141.

# APPENDIX A. DETAILED DESCRIPTION OF THE METHODS AND RESULTS USED IN EXPERIMENT CONDUCTED BEFORE EXPERIMENT 1

## Method

### *Participants*

Fifty-seven undergraduate students at the University of New Mexico participated in this experiment for class credit. None of the participants were involved in any of the other studies. Participants' reliabilities on the relatedness ratings task were computed by correlating their first and second ratings for 15 pairs of words from each list that were repeated. A reliability cut-off of $r = 0.40$ was established to eliminate those participants with low reliability. As a result, 49 participants were included in the analysis. The mean reliability correlation of the repeated ratings across the analyzed samples was $r = 0.70$ (SD = 0.11). In analyses of average relatedness ratings for repeated pairs, the average of the two ratings given was used.

### *Materials*

Four lists of words (theme words *car, chair, cold* and *sleep*) were selected (see Roediger, et al., 2001). Each of the lists consisted of the theme word along with 15 words that elicit the theme word on the free association task (Nelson, et al., 1998). The ordering of the lists was counterbalanced across participants.

There were two conditions for the studied words on the recognition test. In one condition (termed the High Semantic condition), studied words on the recognition test were selected to be words that the individual rated as highly semantically related to the theme word (i.e., ratings of 4 and 5 out of 5) and also low associates to the theme word according to the association strength group norms. In the second condition (termed High Associate condition), the studied words on the recognition test were selected to be words that the individual rated as more weakly semantically related to the theme word (i.e., ratings of 3 and below out of 5) but highly associated to the theme word according to association strength group norms.

The recognition test consisted of 12 studied items (3 per list) and 24 nonstudied items. There were two types of nonstudied items: the 4 nonpresented theme words (*car, chair, cold* and *sleep*) and 20 words generally unrelated to any items on the three lists. The studied words on the recognition test differed across participants and were based on the relatedness ratings obtained in the first stage of the experiment. For each participant, half of the studied items (i.e., 6 items; 2 study lists) were chosen based on the High Semantic condition and the other half of the studied items were chosen based on the High Associate condition. The ordering of the two conditions and the lists chosen for the two conditions were counterbalanced across participants.

The words were displayed in a fixed order on the recognition test. All three of the studied items for a certain list of words were presented at intervals before the respective theme word. For the *car* list, the studied items were presented in serial positions 19, 22 and 28 and *car* was presented in serial position 30. For the *cold* list, the studied items were presented in serial positions 2, 7, and 13 and *cold* was presented in serial position 24. For the *chair* list, the studied items were presented in serial positions 5, 10 and 14 and *chair* was presented in serial position 27. Finally, for the *sleep* list, the studied items were presented in serial positions 4, 8 and 16 and *sleep* was presented in serial position 23. The participants received the same test with the exception of the studied items.

### *Procedure*

The participants were tested individually in a computer room. In the first session, participants were asked to rate the semantic relatedness of all possible word pairs created from the studied lists and their semantically associated recognition themes. The participants were asked to rate the semantic relatedness between each pair of words on a scale of 1 ‘not at all related’ to 5 ‘highly related’. The participants completed 120 (16 choose 2) pairwise relatedness ratings for each list. Each list also included 15 repeated pairs to establish reliability (thus, the total number of ratings for each list was 135). Participants took approximately 35 minutes to complete the relatedness ratings task.

Following a three week delay, the participants returned to the lab and completed the false memory task. During the study phase, words were presented to participants in a female voice at a rate of one word every 1.5 seconds. Participants were told to attend closely to the words in preparation for a memory test immediately presented at the end of the list. After the presentation of the words, the participants engaged in a distractor task (i.e., solving math problems) for 30 seconds. Finally, they completed the memory test in which words were presented one-at-a-time on a computer screen. Participants were told to press a key labeled "old" if the word had been studied on the just-presented list and to press a key labeled "new" if the word had not been previously studied. A practice list, consisting of the names of 12 U.S. states, was presented and tested at the beginning of the experiment. The participants took about 20 minutes to complete the false memory portion of the experiment.

**Results and Discussion**

Hits and false alarms for all four lists were computed. Specifically, for each participant, the two lists that incorporated the highest semantically related words (High Semantic condition) as studied items were compared to the two lists that incorporated the highest associates (High Associate condition). The average hit rate of studied words across the two conditions (hit rate = .88) was very comparable to the established literature (e.g., Roediger & McDermott, 1995). The hit rate for the High Semantic condition was .86 and the hit rate for the High Associate condition was .91. There is no difference between these two hit rates ($p$ = 0.12). The likelihood of false memory, however, was significantly higher than previously established. In fact, it was near ceiling. The average false memory for both the High Semantic and High Associate conditions was .99.  In addition, 41 of the 49 participants falsely recognized all of the theme words. Consequently, any comparison of the two conditions was pointless.

One potential explanation for the high levels of false memory is that the participants were incorporating external cues in their task performance for the second part of the study. Despite the 3 week delay, the cues in the environment (sitting in the same room at the same computer as the first session) might have had an influence on the participant's performance. The participants might have encoded the cues in the environment during the first part of the study and then used these retrieval cues during the second part of the study to judge that themes were studied.

Another condition of the procedure that may have inflated false alarms was that the number of unstudied items on the recognition test significantly outnumbered the number of studied items.  Participants often expect about half of the items on the recognition test to be studied items.  Thus, the participants, expecting half of the items to be studied items, might have indicated all words that would reasonably be a studied item as studied.  Hence, the high level of false remembering may simply be due to the participants' expectation of the number of studied items on the recognition test.  To address this concern in the next experiment, the number of studied and unstudied items should be comparable on the recognition test.

Taking these ideas into consideration, Experiments 1-4 asked participants to only focus on the task at hand for the false memory procedure (and disregard what they did for the semantic relatedness ratings).  The proportion of studied/unstudied items on the recognition test was changed and additional measures were taken to ensure that the participants understood the relatedness ratings task.

## APPENDIX B. TABLE SUMMARIZING THE METHODS, RESULTS AND ADDITIONAL INFORMATION FOR ALL EXPERIMENTS

| Experiment | Methods | Results | Additional information |
|---|---|---|---|
| **1** | Reversed order of tasks (relatedness ratings, 2 week delay, false memory portion) | N = 22<br>Falsely recognized words had higher relatedness ratings; $t(21) = 3.17$, $p<.006$ (replicating previous work) | Actually ran 35 participants; those excluded had low reliability (below 0.40) or falsely recognized all theme words |
| **2** | Followed Marsh et al.'s procedure<br>Relatedness ratings on unstudied lists<br>Manipulate 3 **unstudied** words on **test**<br>**2 conditions:** High Semantic (Highly rated words AND low associates); High Associate condition (High associates AND weakly rated) | N= 41<br>Found non-significant trend for High Associate themes (0.40) to have higher false memory than High Semantic themes (0.28); $t(40) = 1.817$, $p = .077$<br>High Associate and High Semantic conditions both had higher likelihood of false memory compared to unrelated words (0.08) (replicating Marsh's work) | Actually ran 48 participants; those excluded had low reliability (below 0.40), had exceptionally fast reaction times on the recognition test (RT<300ms) or had a higher false memory of the unrelated items than correct recognition<br>Likelihood of false memory was lower than what Marsh et al found (0.49). They used 18 lists and 297 items on the recognition test which might have increased overall false memory. |
| **3** | Followed Coane & McBride's procedure (2 blocks of 3 lists)<br>Manipulated 6 **studied** words on **test**<br>**2 conditions:** High Semantic (Highly rated words AND high associates); Low Semantic (Weakly rated words AND high associates) | N= 36<br>Found no difference between false memory in High Semantic condition (0.88) and Low Semantic condition (0.90)<br>No difference between false memory in both High Semantic and Low Semantic conditions and Control condition (0.91) which does not replicate Coane & McBride's work | Actually ran 66 participants; those excluded had low reliability (below 0.40), did not complete the second session, had an error in their program or did not have enough variability in their relatedness ratings<br>Likelihood of false memory was higher than what Coane & McBride found (0.70-0.74). Might be due to the fact that current study presented 20 studied items while Coane & McBride only presented 12. The increase in study items is likely to increase the likelihood of false memory. |
| **4** | Manipulated 6 **study** items<br>**2 conditions**: High Semantic (words rated as highly related AND high associates); Low Semantic (words rated as weakly related AND high associates) | N= 37<br>No significant difference between false memory in High Semantic condition (0.56) and Low Semantic condition (0.52)<br>Control condition (0.67) had significantly higher likelihood of false memory than Low Semantic condition; t(36) | Actually ran 69 participants; those excluded had low reliability (below 0.40), did not complete the second session, did not have enough variability in their relatedness ratings or had poor recognition performance for studied items<br>Higher likelihood of false memory in Control condition likely due to the fact that, for some participants, the study words in the High Semantic and Low |

| | | | |
|---|---|---|---|
| | | = 2.22, p = .03. This does not replicate Coane & McBride's work<br>No two-tailed difference between false memory in Control condition and High Semantic condition, significant one-tailed difference; t(36) = 1.60, p = .05 | Semantic conditions were presented alphabetically (study words for control condition were never presented alphabetically). Participants could have used this distinctive alphabetical information to their advantage on the recognition test<br>Trend for the higher false memory in the High Semantic condition compared to the Low Semantic condition (4% increase); however, it would take 259 participants for this difference to be significant |
| **Appendix** | Manipulated recognition **test**<br>2 conditions: High Semantic (Highly rated words AND low associates); High Associate (High associates AND weakly rated) | N = 49<br>Found near-ceiling likelihood of false memory for both High Semantic and High Associate conditions (both 0.99) | Actually ran 57 participants; those excluded had low reliability (below 0.40)<br>41 of the 49 participants falsely recognized all theme words<br>Possible reasons for high false memory: participants were incorporating external cues from the environment in their performance on the second portion; the number of unstudied items outweighed the number of studied items on the recognition test.<br>Addressed these possibilities in Experiment's 1-4 |

## APPENDIX C. UNIQUE LISTS USED FOR EXPERIMENTS 1, 3 AND 4 ALONG WITH THE BACKWARD ASSOCIATION STRENGTH (BAS) VALUES PER NELSON ET AL. (1998)

### Experiment 1

| Theme | List Words | BAS | List Words | BAS |
|---|---|---|---|---|
| **Music** | Rock | .045 | Piano | .230 |
| | Song | .209 | Record | .116 |
| | Note | .132 | Sound | .205 |
| | Radio | .270 | Flute | .289 |
| | Stereo | .333 | Hall | .000 |
| | Tune | .311 | Hear | .022 |
| | Dance | .101 | Listen | .011 |
| | Art | .020 | Loud | .018 |
| | Band | .432 | Mellow | .000 |
| | Guitar | .203 | Pretty | .000 |
| | Melody | .243 | Sing | .033 |
| | Noise | .038 | | |

### Experiments 3 and 4

| Theme | List Words | BAS | Theme | List Words | BAS | Theme | List Words | BAS |
|---|---|---|---|---|---|---|---|---|
| **Breakfast** | Cereal | .054 | **Cotton** | Swabs | .053 | **Sticky** | Glue | .185 |
| | Oatmeal | .000 | | Polyester | .020 | | Hairspray | .000 |
| | Pancakes | .014 | | Qtips | .000 | | Paste | .000 |
| | Waffles | .000 | | Fabric | .000 | | Sticker | .000 |
| | Orange juice | .000 | | Puff | .000 | | Tape | .000 |
| | Sausage | .000 | | Wool | .040 | | Syrup | .017 |
| | Doughnut | .000 | | Fluffy | .020 | | Tar | .000 |
| | Grits | .000 | | Gauze | .000 | | Gum | .076 |
| | Toast | .000 | | Fuzz | .000 | | Goo | .084 |
| | Bagel | .000 | | Silk | .000 | | Prickly | .000 |
| | Bacon | .000 | | Cloth | .000 | | Stuck | .017 |
| | Eggs | .068 | | Denim | .000 | | Slimy | .000 |
| | Omelet | .000 | | Nylon | .013 | | Humid | .000 |
| | Oats | .000 | | Soft | .166 | | Sap | .000 |
| | Biscuit | .000 | | Cloud | .000 | | Vaseline | .000 |
| | Instant | .000 | | Dress | .000 | | Peanut Butter | .000 |
| | Supper | .000 | | Yarn | .000 | | Porcupine | .000 |
| | Tart | .000 | | Mill | .000 | | Molasses | .000 |
| | Cornbeef | .000 | | Sheep | .000 | | Cobweb | .000 |
| | Lunch | .473 | | Clothes | .053 | | Honey | .059 |

**Experiments 3 and 4 (continued)**

| Theme | List Words | BAS | Theme | List Words | BAS | Theme | List Words | BAS |
|---|---|---|---|---|---|---|---|---|
| **Summer** | Winter | .396 | **Theater** | Movie | .439 | **Tiger** | Lion | .308 |
| | Spring | .051 | | Act | .014 | | Leopard | .000 |
| | Season | .000 | | Drama | .041 | | Panther | .000 |
| | Camp | .000 | | Cinema | .020 | | Saber | .000 |
| | Shorts | .000 | | Audience | .000 | | Cougar | .000 |
| | Vacation | .000 | | Production | .000 | | Fierce | .000 |
| | Heat | .000 | | Stage | .014 | | Roar | .021 |
| | Sandals | .000 | | Film | .014 | | Stripe | .077 |
| | Session | .000 | | Screen | .014 | | Tame | .000 |
| | Lemonade | .000 | | Popcorn | .000 | | Animals | .000 |
| | Hotter | .000 | | Actor | .034 | | Pounce | .000 |
| | Warmer | .000 | | Arts | .000 | | Cub | .000 |
| | Autumn | .000 | | Ballet | .000 | | Claw | .000 |
| | Beach | .071 | | Lobby | .000 | | Lamb | .000 |
| | Endless | .000 | | Mystery | .000 | | Paw | .000 |
| | Fall | .020 | | Performance | .000 | | Prey | .000 |
| | Mosquito | .000 | | Balcony | .000 | | Leo | .000 |
| | Term | .000 | | Diner | .000 | | Predator | .000 |
| | Olympics | .000 | | Play | .162 | | Safari | .000 |
| | Breeze | .000 | | Portray | .000 | | Bear | .000 |

# APPENDIX D. FREQUENCY WITH WHICH WORDS WERE USED IN HIGH SEMANTIC (HS) AND HIGH ASSOCIATE (HA) CONDITIONS IN ALPHABETICAL ORDER FOR EXPERIMENT 2

**Chair List** **Cold List** **Music List** **Sleep List**

| | HS | HA | | HS | HA | | HS | HA | | HS | HA |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Bench** | 8 | 0 | **Air** | 4 | 0 | **Art** | 12 | 0 | **Awake** | 0 | 27 |
| **Couch** | 0 | 14 | **Arctic** | 14 | 0 | **Band** | 0 | 3 | **Bed** | 0 | 6 |
| **Cushion** | 9 | 0 | **Chilly** | 0 | 8 | **Concert** | 22 | 0 | **Blanket** | 0 | 9 |
| **Desk** | 0 | 10 | **Freeze** | 8 | 0 | **Horn** | 0 | 8 | **Doze** | 3 | 0 |
| **Legs** | 0 | 8 | **Frigid** | 0 | 5 | **Instrument** | 17 | 0 | **Dream** | 0 | 4 |
| **Recliner** | 0 | 9 | **Frost** | 9 | 0 | **Jazz** | 18 | 0 | **Drowsy** | 6 | 1 |
| **Rocking** | 12 | 0 | **Hot** | 0 | 25 | **Melody** | 0 | 5 | **Nap** | 11 | 0 |
| **Seat** | 0 | 1 | **Ice** | 0 | 1 | **Note** | 0 | 4 | **Peace** | 3 | 0 |
| **Sitting** | 14 | 0 | **Shiver** | 9 | 0 | **Orchestra** | 10 | 0 | **Rest** | 0 | 4 |
| **Sofa** | 0 | 10 | **Warm** | 0 | 22 | **Piano** | 0 | 2 | **Slumber** | 5 | 1 |
| **Stool** | 10 | 0 | **Weather** | 4 | 0 | **Radio** | 0 | 2 | **Snooze** | 0 | 2 |
| **Swivel** | 7 | 0 | **Wet** | 0 | 14 | **Rhythm** | 8 | 0 | **Snore** | 7 | 1 |
| **Table** | 0 | 7 | | | | **Sing** | 0 | 1 | **Tired** | 0 | 4 |
| **Wood** | 3 | 1 | | | | **Sound** | 0 | 2 | **Wake** | 0 | 25 |
| | | | | | | **Symphony** | 9 | 0 | **Yawn** | 4 | 0 |

## APPENDIX E. FREQUENCY WITH WHICH WORDS WERE USED IN HIGH SEMANTIC (HS), LOW SEMANTIC (LS) AND CONTROL (C) CONDITIONS IN ALPHABETICAL ORDER FOR EXPERIMENT 3

**Breakfast List** **Cotton List** **Sticky List**

| | HS | LS | C | | HS | LS | C | | HS | LS | C |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Bacon** | 6 | 2 | | **Cloth** | 6 | 1 | | **Cobweb** | 0 | 1 | |
| **Bagel** | 2 | 4 | | **Cloud** | 1 | 1 | | **Glue** | 10 | 2 | 11 |
| **Biscuit** | 1 | 1 | | **Denim** | 2 | 9 | | **Goo** | 1 | 7 | |
| **Cereal** | 1 | 2 | 12 | **Dress** | 3 | 0 | | **Gum** | 8 | 5 | |
| **Doughnut** | 8 | 3 | | **Fabric** | 9 | 2 | 13 | **Hairspray** | 4 | 9 | 11 |
| **Eggs** | 10 | 1 | | **Fluffy** | 4 | 5 | | **Humid** | 2 | 8 | |
| **Grits** | 7 | 3 | | **Fuzz** | 2 | 5 | | **Paste** | 4 | 7 | 11 |
| **Instant** | 4 | 1 | | **Gauze** | 4 | 6 | | **Peanut Butter** | 0 | 6 | |
| **Oats** | 3 | 1 | | **Nylon** | 1 | 9 | | **Prickly** | 0 | 10 | |
| **Oatmeal** | 6 | 1 | 12 | **Polyester** | 2 | 5 | 13 | **Sap** | 1 | 0 | |
| **Omelet** | 10 | 2 | | **Puff** | 3 | 4 | 13 | **Slimy** | 2 | 9 | |
| **Orange Juice** | 9 | 0 | 12 | **Qtips** | 8 | 4 | 13 | **Sticker** | 7 | 3 | 11 |
| **Pancakes** | 9 | 0 | 12 | **Soft** | 5 | 1 | | **Stuck** | 6 | 2 | |
| **Sausage** | 8 | 1 | 12 | **Silk** | 3 | 8 | | **Syrup** | 8 | 1 | 11 |
| **Toast** | 7 | 2 | | **Swabs** | 5 | 2 | 13 | **Tape** | 8 | 1 | 11 |
| **Waffles** | 5 | 0 | 12 | **Wool** | 8 | 4 | 13 | **Tar** | 5 | 5 | |
| | | | | | | | | **Vaseline** | 0 | 8 | |

**Summer List** **Theater List** **Tiger List**

| | HS | LS | C | | HS | LS | C | | HS | LS | C |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Autumn** | 2 | 10 | | **Act** | 4 | 6 | 12 | **Animals** | 6 | 4 | |
| **Beach** | 3 | 4 | | **Actor** | 4 | 6 | | **Claw** | 5 | 3 | |
| **Camp** | 5 | 3 | 11 | **Arts** | 7 | 5 | | **Cougar** | 3 | 8 | 13 |
| **Endless** | 0 | 5 | | **Audience** | 5 | 6 | 12 | **Cub** | 2 | 6 | |
| **Fall** | 1 | 4 | | **Ballet** | 2 | 5 | | **Fierce** | 5 | 3 | 13 |
| **Heat** | 10 | 2 | | **Cinema** | 9 | 2 | 12 | **Lamb** | 0 | 4 | |
| **Hotter** | 5 | 3 | | **Drama** | 5 | 5 | 12 | **Leopard** | 2 | 9 | 13 |
| **Lemonade** | 4 | 3 | | **Film** | 8 | 4 | | **Lion** | 3 | 6 | 13 |
| **Sandals** | 7 | 0 | | **Movie** | 6 | 0 | 12 | **Panther** | 4 | 8 | 13 |
| **Season** | 5 | 6 | 11 | **Mystery** | 2 | 2 | | **Paw** | 2 | 0 | |
| **Session** | 1 | 8 | | **Performance** | 4 | 1 | | **Pounce** | 6 | 5 | |
| **Shorts** | 7 | 2 | 11 | **Popcorn** | 5 | 5 | | **Prey** | 1 | 4 | |
| **Spring** | 2 | 8 | 11 | **Production** | 5 | 7 | 12 | **Roar** | 5 | 4 | |
| **Term** | 0 | 1 | | **Screen** | 5 | 3 | | **Saber** | 2 | 8 | 13 |
| **Vacation** | 7 | 2 | 11 | **Stage** | 8 | 2 | | **Stripe** | 7 | 2 | |
| **Warmer** | 3 | 1 | | | | | | **Tame** | 1 | 10 | |
| **Winter** | 4 | 10 | 11 | | | | | | | | |

## APPENDIX F. FREQUENCY WITH WHICH WORDS WERE USED IN HIGH SEMANTIC (HS), LOW SEMANTIC (LS) AND CONTROL (C) CONDITIONS IN ALPHABETICAL ORDER FOR EXPERIMENT 4

**Breakfast List** **Cotton List** **Sticky List**

| | HS | LS | C | | HS | LS | C | | HS | LS | C |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Bacon** | 7 | 2 | | **Cloth** | 6 | 4 | | **Glue** | 4 | 0 | 10 |
| **Bagel** | 8 | 2 | | **Cloud** | 4 | 4 | | **Goo** | 5 | 4 | |
| **Biscuit** | 3 | 4 | | **Denim** | 4 | 7 | | **Gum** | 3 | 4 | |
| **Cereal** | 2 | 1 | 10 | **Dress** | 6 | 4 | | **Hairspray** | 2 | 5 | 10 |
| **Doughnut** | 5 | 2 | | **Fabric** | 13 | 2 | 10 | **Humid** | 3 | 8 | |
| **Eggs** | 9 | 1 | | **Fluffy** | 10 | 2 | | **Paste** | 8 | 4 | 10 |
| **Grits** | 5 | 2 | | **Fuzz** | 6 | 4 | | **Peanut Butter** | 1 | 9 | |
| **Instant** | 2 | 4 | | **Gauze** | 4 | 5 | | **Prickly** | 2 | 11 | |
| **Oatmeal** | 3 | 1 | 10 | **Nylon** | 0 | 9 | | **Sap** | 7 | 3 | |
| **Omelet** | 10 | | | **Polyester** | 3 | 7 | 10 | **Slimy** | 1 | 8 | |
| **OJ** | 9 | 5 | 10 | **Puff** | 7 | 2 | 10 | **Sticker** | 5 | 2 | 10 |
| **Pancakes** | 7 | 1 | 10 | **Qtips** | 9 | 2 | 10 | **Stuck** | 7 | 2 | |
| **Sausage** | 6 | 3 | 10 | **Silk** | 2 | 8 | | **Syrup** | 6 | 4 | 10 |
| **Toast** | 10 | 1 | | **Soft** | 9 | 1 | | **Tape** | 4 | 1 | 10 |
| **Waffles** | 6 | 3 | 10 | **Swabs** | 7 | 2 | 10 | **Tar** | 1 | 4 | |
| **Oats** | 4 | 4 | | **Wool** | 6 | 3 | 10 | **Vaseline** | 1 | 9 | |

**Summer List** **Theater List** **Tiger List**

| | HS | LS | C | | HS | LS | C | | HS | LS | C |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Autumn** | 0 | 13 | | **Act** | 4 | 5 | 16 | **Animals** | 4 | 2 | |
| **Beach** | 6 | 7 | | **Actor** | 4 | 2 | | **Claw** | 5 | 3 | |
| **Camp** | 5 | 8 | 11 | **Arts** | 3 | 5 | | **Cougar** | 0 | 7 | 15 |
| **Endless** | 1 | 10 | | **Audience** | 3 | 4 | 16 | **Cub** | 3 | 1 | |
| **Fall** | 0 | 12 | | **Ballet** | 3 | 4 | | **Fierce** | 7 | 1 | 15 |
| **Heat** | 5 | 0 | | **Cinema** | 2 | 1 | 16 | **Lamb** | 1 | 6 | |
| **Hotter** | 8 | 1 | | **Drama** | 6 | 6 | 16 | **Leopard** | 1 | 4 | 15 |
| **Lemonade** | 5 | 3 | | **Film** | 8 | 1 | | **Lion** | 1 | 4 | 15 |
| **Sandals** | 8 | 1 | | **Lobby** | 0 | 4 | | **Panther** | 0 | 3 | 15 |
| **Season** | 5 | 0 | 11 | **Movie** | 3 | 2 | 16 | **Paw** | 7 | 1 | |
| **Session** | 2 | 12 | | **Mystery** | 3 | 5 | | **Pounce** | 6 | 4 | |
| **Shorts** | 6 | 1 | 11 | **Performance** | 6 | 2 | | **Prey** | 0 | 4 | |
| **Spring** | 1 | 6 | 11 | **Popcorn** | 3 | 0 | | **Roar** | 8 | 4 | |
| **Vacation** | 4 | 0 | 11 | **Production** | 7 | 5 | 16 | **Saber** | 3 | 6 | 15 |
| **Warmer** | 3 | 2 | | **Screen** | 6 | 4 | | **Stripe** | 7 | 2 | |
| **Winter** | 1 | 6 | 11 | **Stage** | 5 | 3 | | **Tame** | 0 | 8 | |

## APPENDIX G. CORRELATIONS OF SEMANTIC AND ASSOCIATIVE RATINGS FOR EACH PARTICIPANT AND EACH LIST FOR EXPERIMENT 4

| Participant | Breakfast | Cotton | Sticky | Summer | Theater | Tiger | Average |
|---|---|---|---|---|---|---|---|
| **1** | .609* | .722* | | .602* | .632* | | .641* |
| **2** | .726* | | | .778* | .677* | .689* | .718* |
| **3** | .724* | .773* | | | .767* | .761* | .756* |
| **4** | .565* | .510* | | | .541* | .258* | .469* |
| **5** | .749* | .814* | .722* | | .724* | | .752* |
| **6** | .544* | .589* | .649* | .611* | | | .598* |
| **7** | | | .754* | .732* | .693* | .588* | .692* |
| **8** | .721* | | .589* | .643* | .630* | | .646* |
| **9** | | .649* | .706* | .749* | | .639* | .686* |
| **10** | .659* | | | .512* | .558* | .639* | .592* |
| **11** | .567* | .656* | | | .472* | .693* | .597* |
| **12** | .666* | .751* | .706* | | .565* | | .672* |
| **13** | .874* | .725* | | .785* | | .773* | .789* |
| **14** | | | .564* | .399* | .574* | .469* | .502* |
| **15** | | .645* | | .729* | .618* | .525* | .629* |
| **16** | .583* | | .661* | .615* | .773* | | .658* |
| **17** | .595* | .659* | | .526* | .589* | | .592* |
| **18** | .401* | .522* | .509* | | .285* | | .429* |
| **19** | | .695* | .593* | .507* | .553* | | .587* |
| **20** | .660* | | .792* | .043 | .061 | | .389* |
| **21** | .598* | | .629* | .540* | | .680* | .612* |
| **22** | | .615* | .795* | | .731* | .734* | .719* |
| **23** | .596* | | .655* | .738* | | .644* | .658* |
| **24** | .657* | .410* | | | .629* | .437* | .533* |
| **25** | .567* | | | .622* | .558* | .672* | .605* |
| **26** | .679* | .772* | | .808* | | .767* | .757* |
| **Average** | .637* | .657* | .666* | .608* | .582* | .623* | .628* |

Note: Each participant made associative ratings for 4 (randomly determined) of the 6 lists

*Correlations significant at the .01 level