12-14-2018

False memories in recognition memory: Recollection or familiarity?

Alexis E. Payne

False memories in recognition memory: Recollection or familiarity?

By

Alexis E. Payne

A Dissertation
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
in Cognitive Science
in the Department of Psychology

Mississippi State, Mississippi

December 2018

False memories in recognition memory: Recollection or familiarity?

By

Alexis E. Payne

Approved:

Deborah K. Eakin
(Major Professor)

Michael S. Pratte
(Co-Major Professor)

Jarrod Moss
(Committee Member)

Eric Samuel Winer
(Committee Member)

Kevin J. Armstrong
(Graduate Coordinator)

Rick Travis
Dean
College of Arts & Sciences

Name: Alexis E. Payne

Date of Degree: December 14, 2018

Institution: Mississippi State University

Major Field: Cognitive Science

Major Professor: Deborah K. Eakin

Title of Study: False memories in recognition memory: Recollection or familiarity?

Pages in Study 222

Candidate for Degree of Doctor of Philosophy

False recollection refers to the retrieval of contextual information associated with an event that has not occurred. For instance, during a recognition task, one might identify a *nonstudied* word presented at test as old because she remembers the font color of the word during study. Although instances such as this are rare and typically occur at a varying rate of 0-5%, current models of recognition such as the Complementary Learning Systems (CLS) model and the Dual-Process Signal-Detection (DPSD) model do not contain a mechanism to account for their occurrence. Although both the CLS and DPSD models have support from studies demonstrating functional dissociations, neurophysiological dissociations, and behavioral findings of process dissociation, their ability to explain false memories has been more elusive; neither theory specifically addresses false recollection. Instead, such models have ignored false recollection as inconsequential noise in the data.

The purpose of this dissertation was to determine whether the false recognition effect obtained by the Payne-Eakin paradigm was due to false recollection or familiarity. The Payne-Eakin paradigm is based on the PIER2 model, which theorizes that targets implicitly activated during study lead to the falser recognition of a false-target pair. Using

a modified version of the Payne-Eakin paradigm, we investigated the nature of the false recognition effect using *a priori* behavioral analyses and statistical modeling.

The findings of this dissertation provide a step toward a more solid understanding of the cognitive mechanisms involved in the recognition of nonstudied items. This dissertation demonstrates that modeling false recollection is possible. The results of this dissertation suggest that, because current models of recognition do not provide a mechanism to account for false recollection, our understanding of recognition is not fully understood. The results highlight that the current understanding of how false recollection contributes to recognition performance is an area in need of further development.

DEDICATION

I would like to dedicate this dissertation to my younger self. Despite the years of uncertainty, lack of support, and feelings of worthlessness, you were able to build resourcefulness and resilience in a way that have served us well on this journey. You did it, kid!

ACKNOWLEDGEMENTS

I would like to thank my advisor, Deborah "Doc" Eakin. Dr. Eakin was my first exposure to memory research and has spent the last ten years molding me into the scientist that I am today. I would like to thank her for her mentorship and unfailing faith in my abilities.

Second, I would like to thank my very loving, and very patient, husband. He has provided unwavering encouragement, faith, and support during these last five years. The sacrifices he has made to ensure that my educational and professional goals were met are truly appreciated.

I have been fortunate to have shared graduate school with two exceptional colleagues, Dr. Elaine W. Tan and Dr. Karla B. Antonelli. I am so grateful to have been given the opportunity to do this part of life with them and gain two life-long friends. Thank you both for your support and for making graduate school a wonderful memory.

Finally, I would like to thank the undergraduate research assistants from the Memory and Metamemory (M&M) Lab. Without their help, I never could have finished this dissertation. I have enjoyed working with each of them and appreciate the effort that was put into this project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Recognition refers to a specific type of memory that involves identifying stimuli presented at test as having been previously experienced (Norman & O'Reilly, 2003). Different from recall, which occurs when specific context can be retrieved from long-term memory, recognition is unique in that a discrimination response can be given to a stimulus with or without retrieval of the specific context. For instance, one can recognize a person seen on a bus as being familiar without being able to recall from where one knows that person (Mandler, 1980). Early explanations of recognition stemmed from a communication engineering framework that discriminated signals from noise, called the signal-detection framework (Green & Swets, 1966). This framework was later adopted by recognition researchers to interpret old/new recognition data (Peterson, Birdsall, & Fox, 1954). The signal-detection framework was expanded in a set of theories referred to as dual-process models (Atkinson & Juola, 1973; Mandler, 1980), which combined an additional recollection process with the familiarity-based signal detection process. However, the early frameworks were limited in terms of identifying the underlying cognitive processes involved in recognition memory. More recently, approaches to identifying and understanding these cognitive processes have included both neuroscience (e.g., Diana, Yonelinas, & Ranganath, 2007; Norman & O'Reilly, 2003; O'Reilly &

Rudy, 2001) and statistical modeling techniques (e.g., Slotnick & Dodson, 2005; Wixted, 2007; Yonelinas, 1994, 1999, 2002).

Models of Recognition

Neuroscience techniques such as neuroimaging and brain models have provided researchers tools to examine how the theorized cognitive processes in recognition are implemented by the brain. One model that has helped provide a brain-based explanation of recognition is Norman and O'Reilly's (2003) Complementary Learning Systems (CLS) model. According to their model, recognition involves activity occurring in both the hippocampus and the medial temporal lobe cortex (MTLC), with each brain region performing functionally different evaluative recognition processes. In their model, recognition consists of two processes, familiarity and recall. The familiarity process is described as a signal-detection process of the overall activation in the MTLC and recall is described as a representational matching process that involves the retrieval of contextual information. The CLS model has been supported through behavioral studies that have validated the model's predictions (Kim & Yassa, 2013; Stark, Bayley, & Squire, 2002). For instance, Stark et al. (2002) found that patients with hippocampal lesions had significant deficits in their recognition performance, demonstrating the role of the hippocampus in recognition. Although the CLS model provides explanation for how recognition processes are implemented in the brain, the model is limited in explaining how—or whether—the two processes work together to produce recognition. In addition, the assumptions of the model are supported mainly by simulations and behavioral findings. Other computational models of recognition have proved to provide additional information about the cognitive processes of recognition.

One of the predominant computational models of recognition is Yonelinas's (1994, 2002) dual-process signal-detection (DPSD) model. In accordance with traditional dual-process theory, the DPSD model also describes recognition as being composed of two processes, familiarity and recollection. Familiarity is defined as a signal-detection process that includes an evaluation of the overall memory signal of an item and recollection is described as a threshold-like process that operates like the CLS model's (Norman & O'Reilly, 2003) recall process. Consistent with the CLS model, the DPSD model requires the retrieval of study context for recollection to occur. Therefore, the model posits that accurate recollection can only occur for studied items.

Although both the DPSD and the CLS models have support from studies demonstrating functional dissociations, neurophysiological dissociations, and behavioral findings of process dissociation (Jacoby, 1991), their ability to explain false memories has been more elusive (Atkinson & Juola, 1973; Heathcote, Raymond, & Dunn, 2006; Mandler, 1980); neither theory specifically addresses false recollection. Instead, both models have ignored false recollection as inconsequential noise in the data (Elfman & Yonelinas, 2015; Higham & Vokey, 2004; Yonelinas, Aly, Wang, & Koen, 2010), citing the low rates at which these false recollections occur as evidence for this decision (Yonelinas et al., 2010). Yonelinas et al. (2010) cite false recollection as violating the boundary conditions of recognition and, therefore, irrelevant to modeling recognition.

False Recognition and False Recollection

Despite their lack of explanation by current recognition models, evidence for false recollection in recognition memory consistently has been reported by many different researchers (e.g., Gallo, Foster, & Johnson, 2009; Higham, 1998; Higham & Vokey, 2004; Lane & Zaragoza, 1995; Jones, 2013; Mather, Henkel, & Johnson, 1997; Norman, Schacter, & Reder, 1996; Roediger & McDermott, 1995; Roediger, 1996; Schacter, Koutstaal, & Johnson, 1997; Strange, Garry, Bernstein, & Lindsay, 2011; Whittlesea, 2002). Evidence for false recollection has been obtained as a result of implicit associations among words (Roediger & McDermott, 1995), misattributions of source (Lane & Zaragoza, 1995), the influence of prior knowledge (Jones, 2013), and stimulus fluency (Higham & Vokey, 2004), as will be discussed.

False recollection has not been specified by either the CLS or DPSD models. According to the CLS model, processes in the hippocampus protect against false recollection through a two-step matching process. When an item is studied, a sparse representation of the features of the event is stored in long-term memory. At test, when an item is presented for recognition, the item's representation of its features is compared against the stored, sparse representation of the studied item's features; a high degree of match between the features of the two representations is required for the hippocampus to allow for retrieval of the studied context and for recollection to occur. This conservative matching process should reject items that have not been studied, making the likelihood that items are falsely recollected improbable. In particular, the fact that context is not retrieved without a match serves the assumption of the DPSD model that retrieval of the context from the study event is required for recollection. False recollection should not occur when there is no study context associated with a new test item to inform recollection. However, false recollection has been reported in the literature and, although the DPSD model has been criticized for its inability to account for false recollection (Higham & Vokey, 2004), even Yonelinas et al. (2010) has acknowledged that if a paradigm could be designed to increase the number of false recollections from the small numbers found using typical recognition paradigms, then they could be better understood as more than a boundary condition violation. Armed with such data, perhaps the model could be amended to account for them differently than treating their rare occurrence as random error (Yonelinas et al., 2010).

Payne-Eakin paradigm

Payne and Eakin (2017) capitalized on the principles of the Processing Implicit and Explicit Representations (PIER2) theory (Nelson, McKinney, Gee, & Janczura, 1998) and findings by Roediger and McDermott (1995) regarding the impact of implicitly activated associates on false memory to develop a paradigm for increasing the potential for false recollection with the goal of providing data to inform the DPSD model. The Payne-Eakin paradigm applies theories specified by PIER2 (Nelson et al., 1998) regarding the associative nature of words in semantic memory. According to the PIER2 model (Nelson et al., 1998), when a related cue-target paired-associate word pair—the type of stimuli used in the paradigm—is studied, the cue and target are activated in semantic memory. This activation then spreads to other words that are associated with both the studied cue and target, resulting in the implicit activation of those words as well. The critical piece is that activation does not spread to all associates of the cue and all associates of the target; rather, only associates that are shared by both the cue and target are implicitly activated. This activation of shared implicit associates at encoding results in those implicit associates being part of the stored representation of the explicitly studied paired associate in memory and, as such, they can impact recognition (e.g., McEvoy, Nelson, & Komatsu, 1999; Nelson et al., 1998; Nelson, Zhang, & McKinney, 2001).

For the Payne-Eakin paradigm, a list of cue-target paired associates was created for which each studied cue-target pair shared an associate that theoretically would be implicitly activated when the pair was studied together (studied pairs). False-target test pairs were created by recombining the studied cue with the implicitly activated associate as the target. Additionally, new test pairs were created by creating new, control cue-target test pairs that were unrelated in any way to studied pairs. The theoretical supposition was that, because the false target was implicitly activated during study and shared the study context, false recollection would be higher for these test pairs than for control test pairs. The results obtained in two pilot studies showed that people confidently, but falsely, recognized the false-target pairs as old more often than the control pairs, indicating that they were certain that they had studied those pairs. Not only was the occurrence of false recognition outside the bounds that would be considered random error but also, the kind of high-confidence responses given to these false recognitions suggests that they were due to recollection, not familiarity. However, that conclusion cannot be definitively made using the pilot data alone.

Dissertation Experiment

The purpose of this dissertation was to extend the findings of the pilot experiments to determine whether the false recognition obtained by the Payne-Eakin paradigm was due to recollection or familiarity. Based on PIER2 (Nelson et al., 1998), false recollection was theorized to have occurred using this paradigm due to the implicit activation of the false target during study of a cue-target paired associate. According to the CLS model, this implicit activation could have led to features of the implicitly activated associate to be part of the sparse representation of the study pair in the hippocampus. Then, when the implicitly activated associate is paired with a studied cue as the target in a test pair, the feature overlap between the test and study pairs should be enough to trigger a match in the hippocampus, resulting in retrieval of the study context and recollection. Therefore, based on predictions from the CLS model, the false recognition of the false-target test pairs obtained by the Payne-Eakin paradigm should be due to recollection and not familiarity. However, this supposition is in direct contrast with assumptions of the DPSD model that do not allow for recollection of nonstudied items. The DPSD model requires the retrieval of study context for recollection to occur and, because a new test pair should not be associated with the study context, any error in calling a new pair old—even a false-target test pair—should only be due to familiarity, never recollection. However, because the implicit associate shared study context with the explicitly studied cue-target pair, it is possible in the Payne-Eakin paradigm for the context to be retrieved, resulting in a false recollection of those new test pairs.

Determining whether false recognition in the Payne-Eakin paradigm was due to recollection or familiarity was done by taking both a behavioral and statistical modeling approach. Results from both approaches served to inform assumptions of both the CLS and DPSD models. The Payne-Eakin paradigm was modified to include five different types of test pairs. Along with the original false-target test pairs (FTi) and the control test pairs (R-Control) used in the original paradigm, three new types of test pairs were added. FTx test pairs included a studied cue with a new, associated false target; however, this false target was not implicitly activated during study of the study pair. This condition was added to provide a direct test of the impact of the implicit target on recognition. Re-Paired test pairs took cues from a set of unrelated study pairs and repaired them with a target from a different studied pair. This condition was added to include a condition of high familiarity; both the cue and target were previously studied, although not together. Finally, a set of unrelated, unstudied test pairs were created to serve as control pairs for the Re-Paired condition (U-control test pairs). Following the convention of experiments testing the DPSD model, recognition was measured using a 6-point confidence recognition scale. The 6-point confidence recognition scale is favorable to old/new judgments because different theories, such as DPSD, make specific predictions regarding how these confidence ratings map on to process parameters such as recollection and familiarity.

Aim 1

The first aim of this dissertation was to demonstrate whether the false memory effect obtained using the Payne-Eakin paradigm is due to the activation of an implicitly associated word during study of an explicitly presented cue-target pair. To accomplish this goal, we compared the false recognition rate for FTi test pairs to R-control and FTx test pairs. The purpose of these comparisons was to allow us to determine whether the presence of the implicit associate facilitated false recognition. The comparison between the FTi and FTx test pairs was especially important because the only difference between the two types of pairs was whether the false target was an implicit associate of the studied pair. In addition to comparing the false recognition rate between the test pairs, we also compared the rate of high-confidence old responses for FTi pairs and FTx test pairs, operationalized as false recollection. This comparison was done to determine whether the implicitly activated false target facilitated false recollection.

Aim 2

The second aim of this dissertation was to examine the false memory effect obtained by the Payne-Eakin paradigm using the distribution of response patterns to determine whether the responses were produced by recollection or familiarity. According to Elfman, Parks, and Yonelinas (2008), response patterns stemming from recollection should produce different distributions than response patterns stemming from familiarity. Because recollection is characterized as an all-or-none process, responses stemming from recollection should result in a dichotomous distribution whereas responses stemming from familiarity should result in a Gaussian-shaped distribution (Diana, Yonelinas, & Ranganath, 2008; Elfman et al., 2008). Therefore, we compared the response distributions of FTi test pairs to R-control and FTx test pairs using both histogram comparisons and receiver-operator characteristic (ROC) curves to determine whether an impact of recollection could be observed.

Aim 3

The third aim of this dissertation was to determine whether the re-presentation of both a studied cue and target, though not together, in the Re-Paired test pairs increased the level of familiarity associated with the presented test pair. Because FTi and FTx pairs both included a studied cue, the impact of familiarity on those pairs could be similar. Therefore, to fully evaluate the impact of high familiarity on recognition, we included the Re-Paired test pairs and compared the false recognition rate of those pairs to both FTi and U-control pairs. However, because high familiarity can contribute to high-confidence responses like recollection, we also compared the response distributions for the Re-Paired pairs to U-control and FTi test pairs using both histogram comparisons and ROC curves to detect differences between the two processes.

Aim 4

Finally, the last aim of this dissertation was to evaluate whether the prediction of the DPSD model that familiarity alone could account for the false memory effect obtained by the Payne-Eakin paradigm or if the inclusion of a false recollection parameter in the model would be required to account for the data. According to the DPSD model, recollection does not occur for nonstudied items (Yonelinas, 1994; 2002; Yonelinas et al., 2010); despite behavioral evidence to suggest otherwise (e.g., Higham & Vokey, 2004; Lane & Zaragoza, 1995; Payne & Eakin, 2017; Roediger & McDermott, 1995). Unfortunately, the behavioral results are limited in their ability to tease apart the contribution of the two processes due to similar predictions, resulting in difficulty disentangling the two processes. Therefore, we developed two models that included a false recollection parameter and fit them to the behavioral data produced by the Payne-Eakin paradigm. We then compared the fits of the two models to the DPSD model to determine whether the false memory effect obtained by the Payne-Eakin paradigm was due to recollection, or solely to familiarity.

CHAPTER II

RECOGNITION

People often find themselves unable to recall the answer to a question; however, when the answer is presented among a list of choices, they find they can accurately recognize the answer. The cognitive process responsible for this phenomenon is recognition memory. Recognition memory is a specific type of memory involving the ability to determine whether an item has been previously encountered in the environment (Medina, 2008). In a traditional recognition memory paradigm, participants study a list of words. After study, participants then take a recognition test. At test, participants are presented with a list of words that includes studied words mixed with new words that have not been studied, or foils. As each word is presented, participants are asked to answer either *old*, indicating that the word was studied or *new*, indicating that the word was not studied. Recognition accuracy is measured by calculating the number of correctly identified studied items as old and nonstudied items as new tempered by errors of identifying a studied item new or a nonstudied item old. The theory underlying this manner of evaluating recognition memory is signal-detection theory (Green & Swets, 1966; Peterson et al., 1954). Signal-detection theory has been highly influential in the development of theories of recognition and has provided a foundation for several computational models of recognition. Therefore, understanding the underlying assumptions of the theory and how it relates to recognition research is essential toward understanding current theories of recognition, including the DPSD model examined in this dissertation.

Signal-Detection Theory

Signal-detection theory was originally developed by engineers as a framework for discriminating electrical signals from background noise for communication purposes (Peterson et al., 1954). The theory was later adapted by recognition researchers because of the similarity of the theory's core assumptions to recognition memory (Green & Swets, 1966). Old/new responses are interpreted in terms of whether the item presented is actually old or new. The interaction of the word's actual state of being old or new and the participant's responses of old or new creates four possible outcomes: two types of correct—hits and correct rejections—and two types of incorrect—misses and false alarms— responses. As shown in Table 1, an old response given to a studied word is categorized as a hit and a new response given to a nonstudied word is categorized as a correct rejection. A new response given to a studied word is categorized as a miss and an old response given to a nonstudied word is categorized as a false alarm. These four categories provide the foundation for the signal-detection framework and are the factors on which all subsequent measures essential for fully evaluating recognition memory are based.

Table 1

Signal-detection Categorization of Recognition Responses

Study Condition:	*Old:*	*New:*
Studied	Hit	Miss
Not Studied	False Alarm	Correct Rejection

At the core of signal-detection theory is the assumption that all words on a recognition test have varying degrees of memory strength (e.g., Gillund & Shiffrin, 1984; Kintsch, 1970; Murdock, 1982; Wickelgren & Norman, 1966). Whether the underlying signal is based on frequency or familiarity strength varies. However, one commonality among theories is that the signal strength of each item is based on the frequency with which a word has been encountered; words encountered frequently have a higher level of strength than words that have been encountered rarely (e.g., Gillund & Shiffrin, 1984; Kintsch, 1970). When a word has been recently encountered—for instance, by being studied on a list—the strength of that word increases from its baseline strength. These varying levels of strength are assumed to be normally distributed, producing a Gaussian distribution. According to signal-detection theory, because words that have been recently studied increase in strength, the combination of studied and nonstudied words on a recognition test produces two separate distributions along a continuum of memory strength; the distribution of nonstudied foil words have a mean strength that is lower than that of the distribution of studied words, as shown in Figure 1. This assumption that new words are a separate distribution of strength from old words is fundamental to signal-detection theory and is essential in calculating response bias and discriminability, two other factors that contribute to the interpretation of recognition memory.

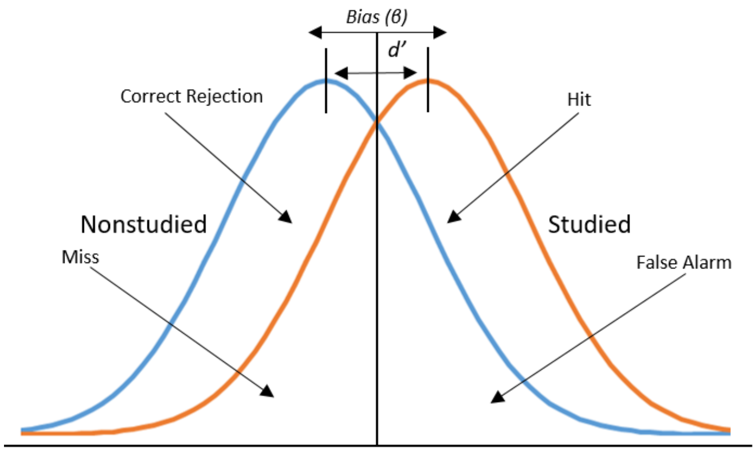


Figure 1. A graphical representation of signal-detection theory's word distribution assumptions, discriminability, and bias.

Response bias, or β, is the likelihood that a word will elicit either an old or new response. Theoretically, response bias—also called criterion bias—is representative of a threshold, predetermined by either within-person factors or due to manipulations using the instructions of a recognition test (Baddeley, Gathercole, & Papagno, 1998, pg. 198), that the memory strength of words must surpass to elicit an old response. If response bias is set to a low value, then a large portion of words, both studied and nonstudied, will be given an old response, resulting in more hits, but also more false alarms. If response bias is set to a high value, then only a small portion of studied words will be given an old response, resulting in fewer hits, but more correct rejections. Response bias is typically denoted with β and is determined by calculating the proportion of hits to false alarms. False alarms and hits are used to calculate bias because the rate at which each of these occurs is indicative of what portion of the studied and nonstudied word distributions surpassed the criterion threshold. However, the degree to which hits and false alarms occur is heavily dependent upon the discriminability of the old and new words presented at test.

Discriminability, or sensitivity, refers to the degree to which studied and nonstudied word distributions are separated from each other on the memory strength continuum. Discriminability is determined by measuring the distance between the mean of the distribution for studied words and the mean of the distribution for nonstudied words. The statistical sensitivity index, *d'* (pronounced d-prime), is typically used when calculating discriminability in signal-detection theory. Because it is assumed that the strength of all words is normally distributed, the means of the distributions for old and new words can be calculated by normalizing, or adjusting the values of the distributions to a comparable scale, which is the rate at which hits and false alarms occur. Then, after the hit and false alarm rates have been normalized, the false alarm rate is subtracted from the hit rate, thus providing a measure of discriminability.

Discriminability, response bias, and the rate at which the four categories of recognition outcomes occur are all used when assessing recognition performance. Typically, performance is first assessed by calculating the hit rate. A high hit rate demonstrates that participants could accurately identify a large portion of studied words as old. However, the hit rate must always be moderated by the false alarm rate. A high hit rate tempered by a high false alarm rate indicates that there was little discrimination between saying old to studied items or to nonstudied items. A combination of a high hit rate and false alarm rate can be indicative of two things: participants either used a liberal response criterion or that studied and nonstudied words were difficult to discriminate between. For instance, foils that are semantically related to studied words are more difficult to distinguish than non-related foils (e.g., McEvoy et al., 1999; Payne & Eakin, 2017). *D'* is typically used when evaluating discriminability, and allows researchers to determine whether participants are operating at chance; a low *d'* value indicates at chance performance whereas a high *d'* value indicates a strong level of discriminability. However, it could also be that a high hit rate combined with a high false-alarm rate is indicative of participants using too liberal of a response criterion in determining whether words were old. Unfortunately, because signal-detection theory uses a dichotomous measure for determining both discriminability and criterion threshold, differentiating between the two is difficult. This issue can be resolved by measuring each of the four parameters in the signal-detection framework and evaluating each parameter in relation to each other.

Although signal-detection theory provides a computational framework for understanding recognition, the theory is limited. One limitation of signal-detection theory is that it fails to describe the processes involved in recognition and how they work. Instead, the theory only provides a framework that can be used to measure and interpret recognition performance. For instance, although the signal-detection framework suggests that list-length effects occur in recognition because of reduced discriminability, the framework cannot explain why increasing the length of a studied list of items decreases discriminability. Additionally, according to the current dominant account of recognition, recognition is described as a dual-process system (Yonelinas, 1994, 1999, 2002) and not a single-process. This view is in direct contrast with signal-detection theory which only describes a single process. As a result, researchers have had to use other methods of evaluating recognition performance to investigate the cognitive processes theorized to be involved. Computational modeling, a mathematical method of theorizing and evaluating how cognitive systems operate, fills this gap.

Computational Modeling

Computational modeling is a mathematical method used to examine how cognitive systems, such as recognition memory, operate. Computational modeling has become an important tool in the examination of recognition memory in that they provide researchers a means of examining cognition in ways that are beyond the capabilities of current behavioral and neuroimaging methods. For instance, current neuroimaging techniques such as fMRI provide methods for examining the underlying neural correlates of the theorized cognitive processes; however, these methods do not provide insight into the nature of the processing. Computational modeling, on the other hand, requires that cognitive processes be not only defined, but also computationalized, resulting in a higher degree of specificity. For instance, computational models of recognition such as Gillund and Shiffrin's (1984) model require that a specific number is assigned to the overall memory signal of to-be-retrieved items as well as to a threshold which memory signals must surpass to be recognized as old. This requirement allows models to make specific predictions as to what items will and will not be retrieved. These model predictions can then be statistically compared to human behavior to determine the accuracy of the model, or the *fit* of the model to the behavioral data.

Although computational models provide a more specific method of evaluating how cognitive processes work, most recognition models have derived directly from the signal-detection framework. For instance, both Norman and O'Reilly's (2003) CLS model and Yonelinas's (1994, 2002) DPSD model use the signal-detection framework to characterize their recognition process of familiarity. However, the unique aspect of these model's signal-detection characterization is that they provide additional specificity to recognition by including a second retrieval process that is responsible for the retrieval of contextual information associated with a studied item. In so doing, both models provide a distinction between the types of information that can be retrieved and how the manipulation of this information impacts recognition. These two models are discussed in detail next because of their relevance to the research questions addressed in this dissertation.

CLS Model

One recognition computational model is Norman and O'Reilly's (2003) CLS neural-network model. The CLS model is different from other recognition models in that it focuses on identifying the neural mechanisms theorized to support the cognitive processes involved in recognition in addition to defining how those processes work. According to the CLS model, recognition consists of two separate retrieval processes, familiarity and recall, with each process having its own underlying neural mechanisms. As a result, each process is implemented within its own neural network model which work together to produce overall recognition performance.

In the CLS model, familiarity is the product of cognitive processing in the medial temporal lateral cortex, or MTLC. Processing in the MTLC is implemented via a cortical neural-network model consisting of multiple layers. One layer is an input layer that projects in a feedforward fashion to other layers. These layers are theorized to be analogous to lower level cortical activity. As items are first being learned, a large number of nodes are activated at an overall lower level, as shown in Figure 2A. When items are repeated, the number of activated nodes is reduced, but activation in each node is increased, making an overall higher level of activation for repeated items, as shown in Figure 2B. This process of specifying which nodes become activated within the presence of a specific cue occurs through Hebbian learning in the model and is referred to as a process of *sharpening*. Norman (2010) validates this sharpening process of cortical nodes by citing that this process is consistent with neurophysiological findings that some cortical neurons show decreased response as a function of how often a stimulus has been presented.

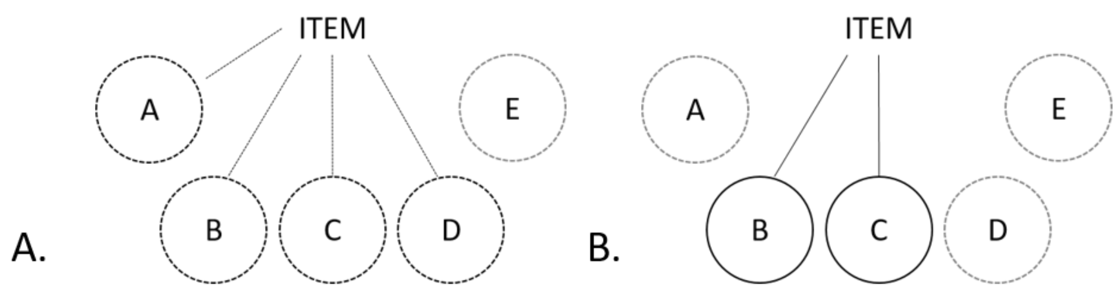


Figure 2. A) A pictorial representation of a large portion of nodes being activated at an overall lower level in the MTLC model. B) A pictorial representation of nodes being sharpened with activation increased and constrained to fewer nodes in the MTCL model.

The process of sharpening highlights the purpose of the cortical model in the CLS model: to extract statistical regularities in the environment to make for quicker and more generalizable learning. However, this aspect of the cortical model also influences the way in which items are retrieved. Because the cortical model concentrates on capturing statistical regularities in the environment, the model cannot isolate specific learning instances, or target events. Instead, the ability of the model to sharpen overall activation patterns is used to discriminate between studied and nonstudied stimuli. Specifically, when an item is presented at test, studied items result in a sharper activation pattern in the cortex whereas nonstudied items result in a "duller" activation pattern. Stemming directly from signal-detection theory, the cortical model then uses the sharpness of the activation to produce an overall value, or memory signal. This value is then compared against a pre-determined criterion threshold to determine whether the item is studied or nonstudied. If the value is greater than the threshold value, it is recognized as old; if it is less than the threshold value, it is recognized as new. This aspect of the model is important for recognition in that it allows the model to quickly determine whether an item has been encountered previously (i.e., is familiar). However, the cortical model is limited in its ability to identify unique items. Therefore, the CLS model includes a recall process to account for the identification of unique stimuli.

In the CLS model, recall is instantiated via the hippocampal sub-model. The hippocampal model is different from the cortical model in that it makes recognition decisions based on the content of the information retrieved rather than the amount of information retrieved. In the hippocampal model, when an item is studied, the representation of that item is stored in two different ways. One representation is a sparse representation, concentrating only on features that are unique to that item. The second representation is more elaborate and the co-occurring context from the study event. When an item is presented at test, the test item's representation is compared to that of the stored item's sparse representation. If the test item's representation has enough overlap with the stored item's sparse representation, then a process referred to as pattern completion begins. During pattern completion, all information related to the test item is retrieved, including contextual information from the studied event. The retrieved information is subsequently compared to that of the presented test item for a second time. If the test item’s representation does not match the stored item’s representation, then that item is rejected as new and is not recognized. However, if the test item’s representation does match the stored item’s representation, then that item is accepted and recognized as old. This process of matching in the hippocampal model allows the model to discriminate between unique studied and nonstudied items, as well as to provide a more stringent criterion for determining whether an item has been studied.

Norman and O’Reilly (2003) tested the strength of their hippocampal model in its ability to discriminate between studied and nonstudied items by simulating the models’ ability to discriminate between studied items and related lures. In this simulation, both the cortical and hippocampal model studied a list of items. At test, a portion of the studied items were re-presented along with a set of both related and unrelated nonstudied lures. The goal of this simulation was to determine whether the hippocampal model was better at discriminating between the studied and related-lure items, highlighting the strength of the hippocampal model at identifying unique study events. Norman and O’Reilly (2003) predicted that the cortical model would have a higher false alarm rate for the related lures than the hippocampal model due to its lower level of discrimination using a signal-detection-like process. Norman and O’Reilly (2003) posit that in the cortical model related lures would activate a large number of studied item nodes due to their overlapping representational patterns, resulting in a higher degree of activation. Therefore, the overall memory signal for related lures should be higher and result in a higher likelihood of surpassing the criterion threshold for those items. However, in the hippocampal model, the level of overlap required between a test item's representation and a stored item's representation was very high—greater than 80%—due to the unique sparse representation of the studied event, resulting in a lower likelihood that a related lure would match. The results of the simulation were consistent with their predictions. The hippocampal model was superior at discriminating between studied items and nonstudied items compared to the cortical model with the cortical model having a higher false alarm rate for related lures (Norman & O'Reilly, 2003).

Elfman et al. (2008) tested the predictions of the CLS model by comparing the data produced by the model to behavioral data using receiver operating characteristic (ROC) curves. To understand their findings, a brief detour to understand ROC curves is required. A ROC curve of recognition data is created by graphing the proportion of hits—studied items that were correctly identified as old—to false-alarms—nonstudied items that were incorrectly identified as old. However, rather than using a dichotomous old/new scale, a 6-point confidence recognition scale is used to extend the number of points that can be plotted on the curve. Items can receive confidence ratings of 6, 5, or 4 with 6 indicating a high degree of confidence that the item was old or confidence ratings of 3, 2, or 1 with 1 indicating a high degree of confidence that the item was new. Points on the ROC curve are plotted as a function of confidence: the first point includes only the most confidently recognized items (i.e., words that elicited a 6 response on the confidence scale). Subsequent points on the scale are cumulative: the second point includes words that received 6 responses plus those that received 5 responses. Each of the subsequent points on the ROC curve plots the cumulative hits as a function of the false-alarms until the final point which represents the total cumulative count for all of the responses on the scale. Discriminability can be extracted using normalized ROC curves. Discriminability is the degree to which studied items are distinguished from nonstudied items and is provided by the space of the curve from the probability line (i.e., the further the curve is from the line, the higher the level of discriminability). Figure 3 shows an example of a typical ROC curve for behavioral recognition data from Yonelinas (1994).

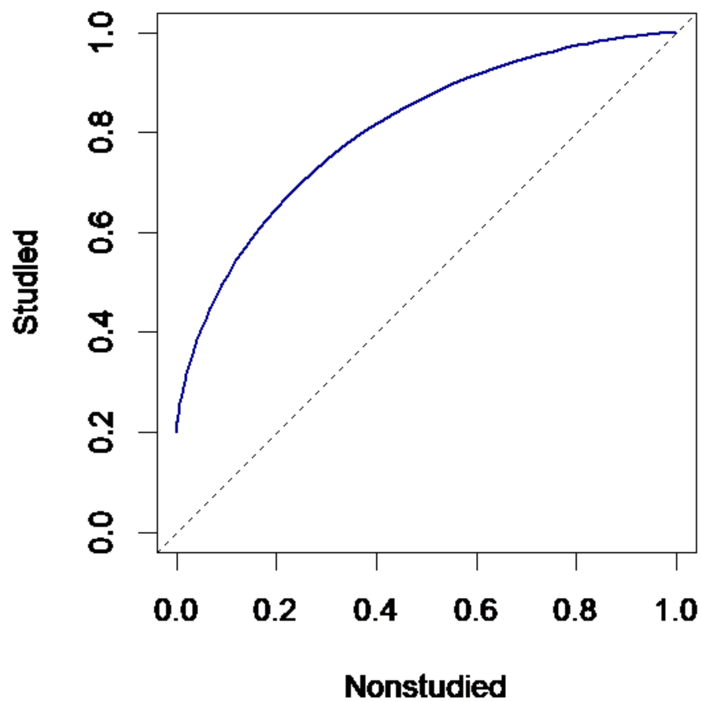


Figure 3. ROC curve for typical recognition data.

ROC curves are often used to evaluate recognition data because of the distinct shape that they produce for standard behavioral recognition paradigms. Typically, item recognition ROCs produce a noticeable curve in probability space with a slight bump in the y-axis, as shown in Figure 3 (Egan, 1975; Murdock, 1974; Yonelinas 1994). It is theorized that the reason item recognition ROC curves are curved in probability space is due to the influence of familiarity which produces a Gaussian-like distribution of responses. The observed bump in the y-axis is due to the influence of recollection (Yonelinas, 1994, 2002).

Because the CLS model characterizes familiarity as a signal-detection process of an overall memory signal in the MTLC and recall as an all-or-none process using representational matching in the hippocampus, predictions about the shape of the receiver operating characteristics (ROC) curves produced by each of these processes can be made. Specifically, items recognized by the cortical sub-model should produce a Gaussian distribution of responses whereas items recognized by the hippocampal sub-model should produce a high-confidence dichotomous response distribution. Elfman et al. (2008) tested this model prediction that cortical ROC curves would be curvilinear and hippocampal ROC curves would be linear by conducting a series of simulations with each model and comparing them to ROC curves observed in behavioral data.

Each simulation done by Elfman et al. (2008) consisted of 500 simulated subjects. For each simulation, a list of 10 items was presented on the left of the screen and 10 items on the right of the screen in random order during study. By pairing items with source data, Elfman et al. (2008) created a requirement that distinct representations of the studied event must be retrieved to ensure accuracy. This source requirement necessitates that the hippocampal model be used for retrieval. New items were generated for each list as well as for each subject. Different learning rates were implemented for each simulated subject to be more comparable with behavioral data. They also included an encoding variability variable to mimic varying levels of encoding for the simulated objects during study (i.e., some objects were encoded better than others). As expected, the hippocampal model produced threshold-like output, and the cortical model produced two overlapping Gaussian-shaped strength distributions, consistent with signal detection models of familiarity. That is, in the cortical model, studied items produced a distinct second distribution in which the overall familiarity signal of those items was larger than nonstudied items causing the studied-item distribution to separate from the nonstudied-item distribution. The distributions produced by both the cortical and hippocampal models were converted into ROC curves. Elfman et al. (2008) found that the ROC curves produced by the cortical model were curvilinear and symmetrical whereas the ROC curves produced by the hippocampal model were linear, supporting their predictions. This ability of the models to produce ROC curves that are consistent with the current theorized recognition processes provides support for the CLS model.

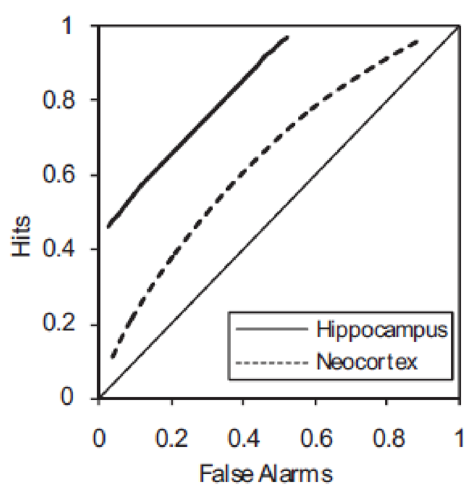


Figure 4. Elfman et al.'s (2008) simulation ROC curves generated by the CLS hippocampal and cortical models.

Finally, additional support for the CLS model comes from studies investigating recognition performance in patients suffering from hippocampal lesions (Broadbent, Gaskin, Squire, & Clark, 2010; Broadbent, Squire, & Clark, 2004; Clark, Zola, & Squire, 2000; Stark et al., 2002; Zola et al., 2000). Stark et al. (2002) compared recognition performance between healthy control patients and amnesic patients with bilateral lesions thought to be limited primarily to the hippocampal region. According to the CLS model, because recognition performance is based on retrieval processing that occurs in both the MTLC and hippocampal regions, if one region is impaired then overall recognition performance also should be impaired. A total of four amnesic patients studied a list of 10 houses paired with a unique face; participants were encouraged to form an association between each house-face pair. Stark et al. (2002) used paired associates to specifically target the ability of the hippocampus to create unique representations of the studied event, predicting that amnesic patients would have a significantly lower recognition accuracy than healthy controls. At test, participants were shown house-face pairs and were asked to determine whether the pair was an intact pair shown at study or a recombination of a house and a face that had been studied, but as part of different pairs.

The addition of recombined house-face pairs tested the influence of the MTLC. Because the recombined pairs contained houses and faces that both had been previously studied, those recombined pairs should activate a large number of overlapping, sharpened nodes. Therefore, if the MTLC is the only neural construct responsible for overall recognition performance, then patients would not be able to discriminate between studied house-face pairs and recombined pairs. However, because the recombined house-face pairs were not studied together, those items should be recognized as new. If the hippocampus stores unique representational features associated with a studied event, as suggested by the CLS model, then the representation of the recombined house-face pairs should create a mismatch in the hippocampus due to having not been studied together, resulting in the rejection of those pairs. Therefore, when the recombined pairs are presented at test, if the hippocampus contributes to overall recognition performance, then those recombined items should be rejected as new, resulting in low false alarm rate. However, because the amnesic patients have hippocampal lesions, those patients should have a lower hit rate as well as a higher false alarm rate.

Stark et al. (2002) found that healthy controls correctly identified 74% of the studied house-face pairs as compared to 59% by the amnesic patients. Healthy controls also had a significantly lower false-alarm rate of 18% than the 45% false-alarm rate of the amnesic patients. These results are consistent with the predictions of the CLS model. Because the amnesiac patients cannot identify the unique studied event of the recombined house-face pairs, they are unable to reject those pairs as new, resulting in a higher false alarm rate, as predicted by the CLS model. However, because the amnesic patients performed near chance at identifying house-face pairs, Stark et al. (2002) performed a second experiment to boost their performance from possible floor effects by increasing the number of times each pair was studied from one to eight times. Although the purpose of increasing the number of study presentations was to prevent possible floor effects, it also served as a second test of the CLS model. Because the CLS model predicts that repeated study presentations results in a more sharpened activation pattern in the MTLC, increasing the number of study presentations should result in identification of a larger number of house-face pairs by the MTLC. However, the increased hit rate in the MTLC should be accompanied by fewer false alarms to recombined pairs; because the items in the recombined pairs were not studied together, their representation at test should not evoke as high of a level of activation as studied, sharpened pairs. Stark et al. (2002) found that increasing the number of study presentations significantly increased the amnesic patients' hit rate from 59% to 74% and also significantly decreased their false-alarm rate from 45% to 35%; however, they continued to perform significantly worse than healthy controls who scored 100% for pairs studied eight times. This result further supports the CLS model's (Norman & O'Reilly, 2003) predictions. Because these amnesic patients are unable to rely on their hippocampus, they are unable to form a stored item representation for each house-face pair, leaving them to rely solely on the familiarity of the pair which cannot discriminate between unique events. However, according to the CLS model, as the number of times that an item is studied increases, the activation for those items is sharpened in the MTLC, resulting in a more accurate activation pattern for that item. Therefore, when the MTCL is presented with an item that has a sharpened activation pattern, the MTLC is more likely to accurately recognize that item as old and is less likely to false alarm to similar lures for that item.

Because the CLS model identifies both familiarity and recall to be involved in recognition, with familiarity being tied to the MTLC and recall to the hippocampus, Stark et al.'s (2002) results are not surprising. If recognition was the product of processing in the hippocampus alone, then the amnesic patients would not have been able to identify the house-face pairs at an above-chance rate. On the other hand, if recognition was the product of processing in the MTLC alone, then the amnesic patients would not have had impaired recognition performance when compared to the controls. Therefore, Stark et al.'s (2002) results from their two experiments provide strong evidence to support the CLS's model of separate recognition processes, familiarity and recall.

Although Elfman et al. (2008) and Stark et al. (2002) provide strong evidence to support the CLS model, the model is limited. The CLS model is vague about how the cortical and hippocampal models work together to produce overall recognition decisions (Norman, 2010). The CLS model is clear as to how the cortical model and hippocampal model operate independently; however, the model does not specify the degree to which the two interact on a given single item. For instance, Norman and O'Reilly (2003) do not specify the timing of the two processes; does familiarity occur after recollection fails, vice versa, or do the two processes occur in parallel? Instead, Norman and O'Reilly (2003) treat recognition decision-making, or the process of making a recognition judgment, as a separate process and do not include it in their model. Norman (2010) addresses this weakness by stating that the purpose of the CLS model is to understand the neural mechanisms underlying recognition retrieval processes, not to understand how the processes interact to produce overall recognition performance. This limitation of the CLS model is a hurdle for recognition researchers who are still debating the underlying recognition processes that produce recognition responses. Therefore, recognition researchers have, instead, mostly relied on Yonelinas's DPSD model (1994, 2002).

DPSD Model

Yonelinas's (1994, 2002) DPSD model is different from the CLS model in that it is a more traditional mathematical recognition model. Unlike the CLS model, the DPSD model does not seek to identify the neural correlates that underlie recognition processes. Instead, the DPSD model concentrates on identifying the recognition retrieval processes and how they interact to produce overall recognition performance. According to the DPSD model, recollection and familiarity are the two cognitive processes responsible for recognition. Familiarity can best be described as an assessment of the information related to an item that provides a value that must reach some pre-defined criterion, and recollection is the result of an active retrieval process initiated by an item at test and produces an all-or-none result (Yonelinas, 1994; 2002). Yonelinas (2002) also uses the same equal-variance distribution as signal-detection theory to describe familiarity. Although Yonelinas (2002) draws heavily from signal-detection in terms of describing his familiarity process in the DPSD model, the recollection process is characterized as a threshold process. Recollection is a separate and independent cognitive process from familiarity that operates at the same time during retrieval and is an all-or-none process that requires the retrieval of veridical contextual information associated with a target item.

According to the DPSD model, the recollection process can result in identifying an item as either old or new. If an item is recollected and the contextual information that is retrieved matches the item that is currently being presented, then that item is immediately recognized as old and elicits a high-confidence response on a 6-point confidence recognition scale (i.e., a 6-response). Recollection in the DPSD model is consistent with the CLS model's (Norman & O'Reilly, 2003) recall process. Just as recall in the CLS model involves the retrieval of contextual information associated with a targeted studied item, recollection in the DPSD model also requires the retrieval of contextual information. Additionally, in the CLS model, recall is responsible for retrieving items associated with a studied event by storing information with unique event features. Therefore, for recall to occur, items presented at test must have a high level of overlap between a test item's representation and a stored item's representation, resulting in a high degree of accuracy. Similarly, Yonelinas (1994, 2002) also describes recollection as a highly accurate retrieval process that can only occur for studied items.

However, unlike the CLS model, the DPSD model posits that if an item is not recollected, then that item undergoes the familiarity process for retrieval. In the DPSD model, the signal-detection process of familiarity occurs for both studied and nonstudied items. According to the model, because familiarity is a signal-detection process, retrieved items will yield an equal distribution of varying levels of familiarity with most items associated with a middle level of familiarity. Therefore, because items can have varying levels of familiarity, they also will have varying levels of confidence associated with them as well. For instance, studied items that have a low level of familiarity associated with them might yield a low-confidence unsure response, as shown in Figure 5 below. This unsure response would most likely translate to a 4-response on a 6-point confidence recognition scale. Conversely, nonstudied items might be associated with a higher degree of familiarity and evoke a less-confident unsure new response, as shown in Figure 5 below. This unsure new response would likely translate to a 3-response on a 6-point confidence recognition scale.

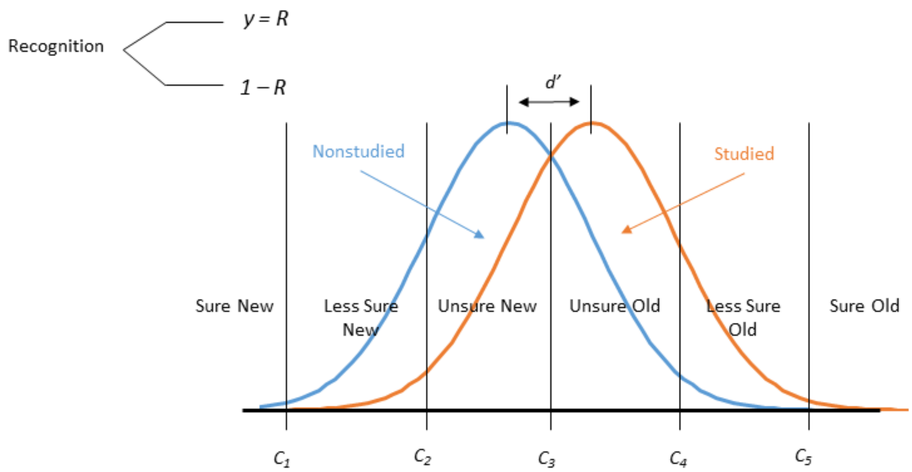


Figure 5. A pictorial representation of the familiarity signal for both studied and nonstudied items divided into varying levels of confidence.

However, rather than using equal variance distributions like shown in Figure 5, Yonelinas (2002) typically represents recognition data using ROC curves to provide evidence for the DPSD model. Yonelinas (2002) argues that ROC curves of recognition data provide the strongest arguments for the presence of recollection. If recognition was based on a single cognitive process—because familiarity is theorized to be a normally distributed measure (Gillund & Shiffrin, 1984; Mickes, Wixted, & Wais, 2007; Wixted, 2007)—then ROC curves of recognition data should be symmetrical, producing a slope equal to 1.0 and a y-intercept equal to 0 when normalized. Specifically, using the 6-point confidence recognition scale should produce mostly ratings of three and four with only a small portion of familiar items categorized as highly familiar 6 responses, and relatively few unfamiliar items categorized as very unfamiliar 1 responses, as shown in Figure 5. However, recognition data typically produces a skewed distribution with approximately

10% of data falling into the high-confidence 6 category, creating an asymmetrical ROC curve and an increased y-intercept, shown in Figure 3 (Egan, 1958; Heathcote, 2003; Mandler & Boeck, 1974; Mickes et al., 2007; Murdock & Dufty, 1972). Other recognition models have attempted to explain this asymmetry by theorizing that studied-item distributions create a larger amount of variance, thus allowing for a larger portion of data to fall into the right tail of the studied item distribution. However, those models have difficulty accounting for as much as 10% of the data falling into the right tail (Mickes et al., 2007).

The DPSD model can account for the increased proportion of high-confidence responses observed. Because the DPSD model includes a recollection process that is independent from familiarity, recollection contributes independently to recognition performance. Also, because recollection involves the active retrieval of contextual information associated with an item, information retrieved via recollection is more likely to induce a stronger feeling of confidence for either accepting the item as old or rejecting it as new. For instance, if information such as the font color of a word during encoding is retrieved, the likelihood that a high-confidence 6 will be given is much more likely than a lower confident response (Mickes, Wais, & Wixted, 2009). Therefore, the inclusion of this recollection process in the DPSD model allows the model to account for the increased number of high-confidence responses that produce an increased y-intercept and asymmetrical shape on the ROC curves.

Yonelinas (1994) performed an experiment to demonstrate the statistical validity of the DPSD model in terms of being able to fit behavioral data. For this experiment, Yonelinas (1994) hypothesized that, if recollection is a separate cognitive process that contributes to overall recognition performance, then manipulations that have been shown to negatively impact the retrieval of contextual information, but not familiarity (e.g., inclusion/exclusion instructions; Jacoby, 1991), should negatively impact recognition performance, evidenced by either a decreased y-intercept or decreased slope. The experiment was a 2 (List Length: short, long) X 2 (Instructions: inclusion, exclusion) experimental design. List length was included because it has been shown to have an inverse relationship with recollection; as the number of to-be-remembered items increases, recollection for those items decreases (Yonelinas & Jacoby, 1996). The instruction manipulation was included to evaluate the differential impact of recollection because of prior research that has demonstrated that exclusion instructions differentially impacted recollection over familiarity (Jacoby, 1991). Both list length and test instructions were manipulated within subjects. The order of the studied list length and test instructions were counterbalanced across participants. Participants completed a total of eight study-test blocks. Each study-test block consisted of two studied lists, either short (i.e., 10 words) or long (i.e., 30 words); each word was presented at a rate of 2 seconds each. Immediately following the study phase of each block, participants then completed two recognition tests using either inclusion instructions or exclusion instructions. In the inclusion instruction condition, the first test asked participants to only respond old (i.e., 4-, 5-, or 6-response) to an item if it was recognized as being in List 1, and the second test asked participants to only respond old (i.e., 4-, 5-, or 6-response) to a word if it was recognized as being in List 2. In the exclusion instruction condition, the first test asked participants to respond new (i.e., 1-, 2-, or 3-response) to a word if it was recognized as being from either List 2 or was never studied, and the second test asked participants to respond new (i.e., 1-, 2-, or 3-response) to a word if it was recognized as being from either List 1 or was never studied. Although these instructions are similar, they are thought to differentially bolster recollection and familiarity.

According to Jacoby (1991), inclusion and exclusion instructions impact recollection and familiarity differently because recollection is an intentional process that can be consciously controlled, whereas familiarity is an automatic process. Inclusion instructions are the same as typical recognition test instructions and are thought to impact both recollection and familiarity by asking participants to recognize words from a specific list as old, allowing for both recollection and familiarity to operate. However, exclusion instructions emphasize recollection by asking participants to exclude words recognized from other lists as new; only words recognized from the specified lists should be identified as old. By asking participants to concentrate on which word belongs to which lists, the controlled process of recollection operates to identify words' source or contextual information (i.e., other words that also appeared on the same list, internal contextual cues felt during the study of this word, etc.), to determine whether the word was on the targeted list or not. Therefore, exclusion instructions emphasize the impact of recollection and minimize the impact of familiarity.

In terms of the behavioral data, Yonelinas (1994) found that for the inclusion instructions, participants were slightly more accurate at recognizing words from a short list of words (M = .78) than from a long list of words (M = .70), demonstrating a list-length effect (Strong, 1912). However, for the exclusion instructions the opposite was found; participants were more accurate at identifying words from a long list of words (M = .30) than a short list of words (M = .22). Yonelinas (1994) then calculated the impact of recollection and familiarity separately. Because the exclusion instructions were thought to emphasize the impact of recollection over familiarity, recollection was calculated by subtracting the probability of accepting a studied word under exclusion instructions from the probability of accepting a studied word under inclusion instructions. Familiarity was calculated by dividing the exclusion score by one minus the estimated probability of recollection. An analysis of variance (ANOVA) performed on both recollection and familiarity in terms of list length found that recollection was negatively impacted by list length. Recollection was significantly higher for short lists (M = .56) than for long lists (M = .40), whereas familiarity was not affected by list length. This finding is consistent with Yonelinas and Jacoby (1996) results' that also demonstrated recollection was worse for longer lists than for shorter lists.

Yonelinas (1994) provided additional evidence of recollection by plotting recognition data using ROC curves. Figure 6 shows the plotted ROC curves for each of the four experimental conditions: List Length (short, long) X Instructions (inclusion, exclusion). For the inclusion conditions, the curve for the long list fell slightly below the curve for the short list, demonstrating that recognition accuracy was higher for the short list than the long list. The inclusion curves also demonstrate a higher y-intercept than the exclusion instructions indicating a higher degree of recollection for those items (i.e., sure old, or 6-responses, on the 6-point recognition confidence scale). However, for the exclusion conditions, both curves had significantly lower y-intercepts and slopes. At first glance, these results suggest that the impact of recollection was significantly lowered rather than increased, as predicted; however, looking at the 1-reponses demonstrates that recollection was increased. The ROC curve for the exclusion instructions show a large increase in high-confident new responses as compared to the inclusion instructions. This increase in high-confident new responses indicates that participants were using recollection to reject information (as evidenced by the increased number of 1 responses) rather than accept information (which would have been evidenced by an increased number of 6 responses). Because participants were asked to *exclude* items recognized from another studied list as new rather than *accept* items recognized as old, recollection is more likely to impact sure new, or 1-, responses more so than sure old, or 6-, responses. Therefore, these results demonstrate that participants could consciously retrieve contextual information associated with a target word to determine whether the word was old or new which provides support for a second cognitive process.

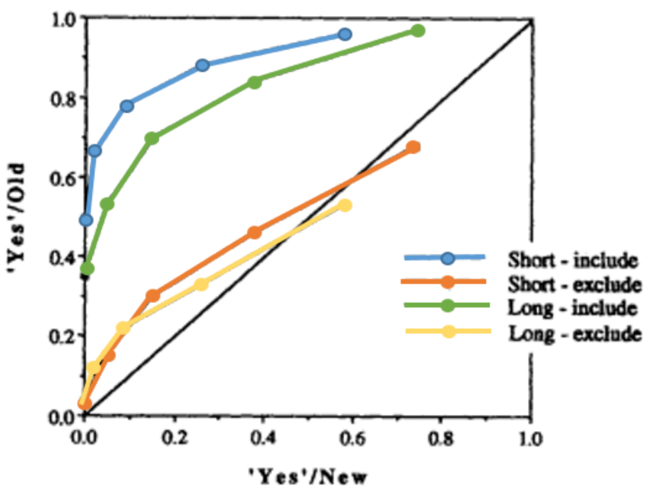


Figure 6. ROC curves for Yonelinas (1994) Experiment 1 results.

In terms of fitting the DPSD model, Yonelinas (1994) created ROC curves using the DPSD model and estimated parameters based on the observations from the behavioral data. From these simulations, Yonelinas (1994) found that the DPSD model accurately predicted the y-intercepts and slopes of the observed ROC curves for both short and long lists under inclusion instructions. The predicted and observed y-intercepts were 1.70 and 1.60, respectively, for the short lists, and the 1.33 and 1.33 for the long lists. The predicted and observed slopes were .64 and .57, respectively, for the short lists, and .76 and .74 for the long lists. Although Yonelinas (1994) found that the DPSD model could accurately predict the observed behavioral data, he did not provide any statistical measures, such as a maximum likelihood estimate, to demonstrate the model's ability to fit the data.

Although the DPSD model provides a good account of recognition data and has allowed researchers to more fully understand the cognitive process involved in recognition, the model has its limitations. One limitation of the DPSD model is that it assumes that recollection can only occur for studied items and not nonstudied items. Therefore, the model cannot account for instances in which contextual information for nonstudied items might be retrieved, or recollected. This aspect of the model has been problematic in that there have been several studies that have shown recollection of nonstudied information (e.g., Dennis, Bowman, & Peterson, 2014; Gallo & Roediger, 2003; Higham & Vokey, 2004; Lane & Zaragoza, 1995). Higham and Vokey (2004) criticized the DPSD model for its inability to account for false recollections, stating that the seriousness of false recollections for models of recognition has been underemphasized. Specifically, Higham and Vokey (2004) argue that the inability of current models of recognition to account for the retrieval of contextual information for nonstudied items suggests that our current understanding of recognition retrieval processes is incorrect.

Yonelinas acknowledged this limitation of the DPSD model by referring to instances in which the model fails as violations of the model's *boundary conditions* (Yonelinas, 2002; Yonelinas et al., 2010). According to Yonelinas, the purpose of the DPSD model is to account for conditions that are typical for recognition; therefore, conditions that are atypical fall outside the purview of the model and violate the model's boundary conditions. An example of an atypical condition in recognition is a high level of semantic and perceptual similarity between a studied item and a nonstudied lure (i.e., RAT and RATS). Under these conditions, recollection has difficulty discriminating between the two items. As a result, this difficulty can result in false recollection of the nonstudied lure (i.e., falsely recollecting RATS). Yonelinas et al. (2010) goes on to explain why this violation occurs using verbiage from the CLS model, stating that the similar semantic and perceptual information results in a high level of feature overlap during test. According to the CLS model, if there is a high enough level of feature overlap (approximately 80%) then the hippocampus will retrieve incorrect contextual information for that item. Therefore, a high level of semantic and perceptual information between a studied and nonstudied item can result in a high level of feature overlap that can lead to the false recollection of the nonstudied item.

Although Yonelinas et al. (2010) acknowledges that there are conditions under which false recollection can occur, he has not amended his DPSD model to account for them. Rather, Yonelinas et al. (2010) defends the DPSD model's (Yonelinas, 1994, 2002) limitation by claiming that false recollection is simply a violation of a boundary condition of the model. Instead, Yonelinas et al. (2010) argues that because false recollection produces such a small effect—from 0% to approximately 5%—in conditions that are typical to recognition, the cost of modeling such a small proportion of data is not worth the benefit of fitting them. As a result, Yonelinas (2010) typically removes any false recollection data and classifies it as erroneous noise. Norman (2010) goes on to defend this limitation of the DPSD model stating that models such as the DPSD and CLS models (Norman & O'Reilly, 2003) are purely measurement models,

> "… a measurement model does not have to be exactly correct to be a useful source of converging evidence, and it does not have to be applicable in every situation… We just need to be aware of the boundary conditions, and mechanistic models like CLS give us a way of developing intuitions about what these boundary conditions will be." (pg. 12)

Although Yonelinas et al. (2010) and Norman (2010) provide a somewhat valid argument, they fail to address that one consequence of their decision to not include a mechanism that can account for false recollection is that our understanding of recognition retrieval processes is incorrect. Although false recollections may occur infrequently in typical experiments used to test the DPSD model, because false recollection has been shown to systematically occur under certain conditions—demonstrating that they are not the byproduct of erroneous error—then nonstudied items should be impacted by the same factors shown to influence studied items. Aim 1 of this dissertation was to implement a paradigm designed to increase the potential for false recollections (discussed in Chapter III) using an experimental manipulation theorized to increase false recollection. With more false recollections in hand, Aim 4 of this dissertation was designed to directly test the DPSD model to determine whether the addition of a false recollection parameter was necessary to fit false recollection data. Before presenting the specific aims of this dissertation, we will discuss the literature on false recollection and the conditions under which they have occurred.

CHAPTER III

FALSE RECOLLECTION

False recollection refers to the retrieval of contextual information associated with an event that has not occurred (e.g., Dennis et al., 2014; Gallo & Roediger, 2003; Higham & Vokey, 2004; Lane & Zaragoza, 1995). For instance, during a recognition task, one might identify a nonstudied word presented at test as old because she remembers the font color of the word during study (Mickes et al., 2009). Although instances such as this are rare and typically occur at a varying rate of 0-5%, they have been found to occur in many different studies under a variety of conditions (e.g., Barba, 1993; Dennis et al., 2014; Dewhurst, 2001; Eakin, Schreiber, & Sergent-Marshall, 2003; Gallo et al., 2009; Higham, 1998; Higham & Vokey, 2004; Johnson, Hashtroudi, & Lindsay, 1993; Lane & Zaragoza, 1995; Jones, 2013; Mather et al., 1997; Norman et al., 1996; Payne, Elie, Blackwell, & Neuschatz, 1996; Roediger & McDermott, 1995; Roediger, 1996; Schacter et al., 1997; Strange et al., 2011; Whittlesea, 2002), demonstrating their robustness.

For the next portion of this dissertation, I will discuss four studies with varying experimental factors that have shown to induce false recollection. Each study will be discussed in terms of the current DPSD model and Yonelinas's recollection boundary condition explanation using the CLS model.

Implicit Associations

One experimental factor that has been shown to influence false recollection is the manipulation of implicitly activated associations among a list of words. In the Deese-Roediger-McDermott (DRM) paradigm, participants are given a list of 15 related words such as BED, REST, AWAKE, TIRED, DREAM, WAKE, SNOOZE, BLANKET, DOZE, SLUMBER, SNORE, NAP, PEACE, YAWN, and DROWSY to study. After a short delay, participants are given an old/new recognition test containing studied words, unrelated nonstudied words (e.g., GOAT), and a single related nonstudied word often referred to as the critical lure (e.g., SLEEP). The critical lure is a word that is associated with each of the words on the studied list but is not presented at study. The DRM paradigm produces a significantly higher false alarm rate for the critical lure on the recognition test than for the unrelated word. Roediger and McDermott (1995) theorized that the increased false alarm rate for the critical lure was due to the item being falsely recollected as being a part of the list. This hypothesis stemmed from the idea that, because the lure was associated with each of the other words on the list, it was implicitly activated during study. This implicit activation of the critical lure resulted in the lure being a part of the studied representation. However, because only old/new judgements were used in initial study, the degree to which participants experienced these critical lures as familiar or recollected cannot be determined. To determine whether participants were experiencing these critical lures as familiar or recollected, Roediger and McDermott (1995) performed a follow-up study that collected Remember/Know judgments in addition to old/new judgments.

Remember/Know judgments are given at the time of recognition to indicate the underlying reason an item is recognized (Tulving, 1983). Participants are instructed to provide a Remember judgment if contextual information is retrieved along with the test item. For instance, if a person remembers an item as being presented in a particular font color of the item, then that item would be given a Remember judgment. Participants are instructed to provide a Know judgment if a test item seems familiar but no contextual information about studying the item can be retrieved. For instance, if the test item seems familiar but the font color in which it was presented cannot be recalled, then that item would be given a Know judgment. As such, Remember judgments are thought to measure the process of recollection, whereas Know judgments are thought to measure familiarity (Tulving, 1983). In Roediger and McDermott's (1995) study, more than half of the nonstudied lures (M = .58) that were categorized as old by participants also received a Remember judgment, suggesting that participants were recollecting contextual details associated with the lure. This finding is surprising given that there should not be any contextual information associated with these lures because they were never part of the study event.

Roediger et al. (1998) explained the false recollection effect obtained by the DRM paradigm using the theory of associative processes. According to this theory, false memories occur because associations with studied list words are implicitly activated along with the explicitly activated studied words (Nelson et al., 1998). Consequently, these implicit associations become part of the studied event. For instance, when studying the DRM list of words associated with the critical lure SLEEP, the word SLEEP is implicitly activated when the list of associates for SLEEP, such as BED, DREAM, and

PILLOW, are studied and becomes part of the stored representation of the studied list. Therefore, during retrieval, the critical lure SLEEP, is falsely recollected as having been studied because contextual information from the study event is also retrieved for that item. Therefore, the contextual information that was part of the studied list is shared by the implicitly activated critical lure, resulting in people to experience the critical lure as having been studied when it was not, producing false recollection for that lure.

Roediger et al.'s (1998) theory of associative processes explanation is in accordance with Yonelinas's boundary condition explanation. According to Yonelinas's boundary condition explanation, when a studied and nonstudied item share a high degree of feature overlap, information associated with the studied item can be incorrectly recalled and associated with the nonstudied item. Therefore, regarding the DRM paradigm, the implicit activation of the critical lure during study of the DRM lists results in the lure being encoded as part of the studied list and subsequently sharing features with studied items. During retrieval, because the lure shares these contextual features with stored list items there is sufficient overlap to activated matching in the hippocampus. This high degree of overlap, then results in the retrieval of the studied event information, including contextual information, resulting in false recollection for that item.

Although Yonelinas would consider the conditions of the DPM paradigm as violations of the boundary conditions for recollection, these false recollections do not occur infrequently. Yonelinas et al. (2010), stated that, because false recollections typically occur at such a low rate, accounting for them is not necessary. However, Roediger and McDermott (1995) reported that as much as 58% of critical lures can be falsely recollected, resulting in a much higher rate than seen in standard recognition paradigms. This finding demonstrates that false recollection effects can be very large, highlighting the need to amend the DPSD model to better account for the retrieval processes responsible for false recollection. One method of doing this would be to amend the model to include a mechanism for recollection of nonstudied items. This amendment was tested as a viable method as part of Aim 4 of this dissertation.

Source Attribution Errors

The studies on the DRM paradigm showed that veridical contextual information from a studied event can be associated with and impact recollection of a nonstudied item. The literature regarding source attribution errors has shown that incorrect contextual information can produce false recollection. A source attribution error, or a source misattribution, is an error that occurs when a person retrieves incorrect source information for an event (Johnson et al., 1993). For instance, a person might study both the word RAT and RATS but in two different contexts, RAT studied in white font and RATS studied in black font. A source attribution error occurs, when someone retrieves the word RAT as having been studied in black font instead of white. Although source information can include information such as font color, but it can also refer to contextual information regarding where the information was encoded (e.g., at a store versus at school), how it was encoded (e.g., in a magazine or a book), or who produced the information (e.g., a friend or a parent). According to the DPSD model, source information, along with all other contextual information, should always be veridical. However, studies demonstrating source attribution errors suggest that this might not always be the case.

Lane and Zaragoza (1995) investigated the subjective experience reported by participants that often accompanies source attribution errors. Participants were presented 96 items—half of the items were presented as pictures, and the other half were presented as words—and asked to state the use for each item (e.g., SCISSORS could be used to cut a sheet of paper). One day later, participants took a recognition test consisting of some items they had seen the previous day as pictures, some items they had seen as words, and 48 nonstudied items. Participants were instructed to only identify words as old that they remembered seeing as a picture the day before. In addition, participants gave Remember/Know judgments. They were instructed to indicate whether they remembered seeing the word as a picture (i.e., retrieved contextual information associated with the word) or only knew they had seen the word as a picture (i.e., the word was familiar but no contextual information was retrieved). Participants were significantly more likely to give a Remember judgment to a word incorrectly identified as a picture than to a new item, indicating that participants retrieved false contextual information of the word being presented as a picture, producing a false recollection due to a source attribution error. Approximately 14% of the items incorrectly identified as having been studied as pictures were given a Remember judgment by participants, demonstrating that participants were experiencing false recollections for those items.

Although Lane and Zaragoza (1995) found that the source information on which participants based their judgments was incorrect, they did show that people could base their recognition judgements on incorrect contextual information retrieved for that item. This finding is problematic for the DPSD model because it shows that veridical contextual information can be misattributed to the wrong studied item, violating the

DPSD model assumption that contextual information associated with the studied item produces recollection. This finding is also problematic for Yonelinas's boundary condition explanation. According to the boundary condition explanation, a nonstudied item must share a large enough portion of features with a studied item for contextual information to be retrieved. However, in Lane and Zaragoza's (1995) study, the items studied as words and the items studied as pictures were not semantically related or perceptually similar. Therefore, items studied as pictures should not have shared enough features with items studied as words to surpass the retrieval criterion. Instead, Lane and Zaragoza's (1995) results could be due to either a binding error or a monitoring error (Fandakova, Shing, & Lindenberger, 2013), but it currently unclear as to the cause of their source misattributions. Regardless, Lane and Zaragoza's (1995) results demonstrate that false recollection due to source misattribution can occur outside of Yonelinas's recollection boundary condition, further highlighting the necessity for the DPSD model to be amended.

Prior Knowledge

Although Lane and Zaragoza (1995) demonstrates that source information within the context of a single study can be misattributed, there is also evidence that prior experience can be result in source misattribution errors. Jones (2013) manipulated the amount of prior knowledge participants demonstrated for the subject of a news headline. Participants completed a pre-screen, consisting of five questions about various celebrity actors, musicians, politicians, and sports figures (i.e., What is this person famous for?), which researchers scored for accuracy. Correct answers were given a score of 1, resulting in each subject receiving a total possible score of 0 to 5 for the five questions. Headlines receiving a score of 4-5 were categorized as high-knowledge, those receiving a score of 2-3 were categorized as middle-knowledge, and those scored 0-1 were categorized as low-knowledge. Based on their results in the pre-screen, researchers created an individualized list of 30 headlines for each participant consisting of 24 true headlines (8 high-knowledge, 8 middle-knowledge, and 8 low-knowledge) and 6 false headlines (2 high-knowledge, 2 middle-knowledge, and 2 low-knowledge). Each headline was presented one at a time for 6 seconds each after which participants gave old/new judgments about the headline and then Remember/Know judgments for headlines given an old judgement. Participants were explicitly told to only provide Remember judgments if they remembered having seen the headline before and could retrieve contextual information for that event. Additionally, they were also told to provide Know judgments if they knew (i.e., it was familiar) that they had seen the headline before but could not recall any contextual information associated with the event. Jones (2013) found that false headlines informed by a high level of prior-knowledge received significantly more Remember judgments, followed by middle-knowledge false headlines, and then low-knowledge false headlines, as shown in Figure 7. This same pattern of results was found for true headlines, demonstrating a relationship between the amount of prior knowledge about the topic of a headline and the likelihood that a headline would be falsely recollected.

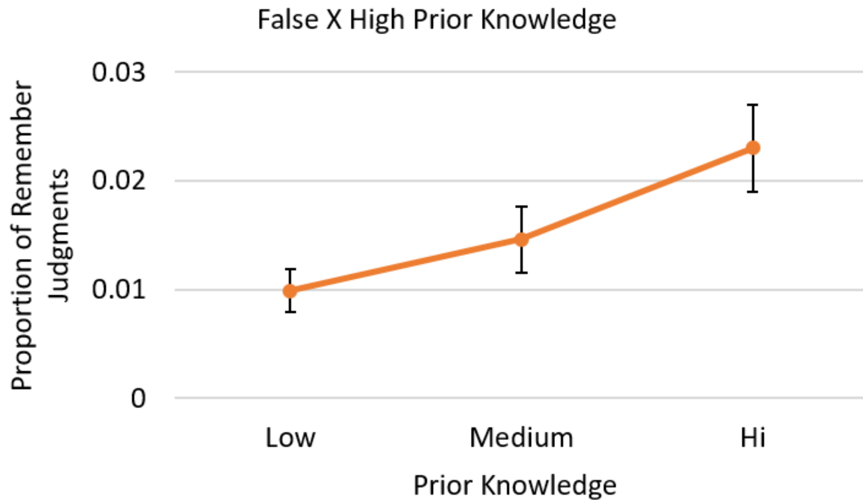


Figure 7. Remember judgments for false headlines increased as a function of level of prior knowledge (Jones, 2013).

Jones (2013) data is problematic for the DPSD model in two ways. The first problem for the model is that Jones (2013) demonstrates that prior knowledge was systematically impacting the likelihood that a headline, either true or false, would be remembered, suggesting a systematic, rather than random, error in the retrieval process. According to the DPSD model, only headlined events that had been previously encountered should have been remembered. However, retrieval of contextual information due to prior knowledge that was related with the topic of the false headline, especially for high-knowledge headlines, resulted in false recollections. This systematic relationship between prior knowledge and false recollections contradicts Yonelinas's (2002) assumption that false recollections are due to random error. The second problem that Jones (2013) creates for the DPSD model is that it defies Yonelinas's recollection boundary condition explanation. According to Yonelinas's boundary condition, false headlined events must share a high degree of feature overlap with a true event for contextual information to be retrieved. However, Jones (2013) chose false headlines that were deemed less plausible and less likely to have occurred based on previous pilot data. Therefore, the false headlines that were used in her study should not have shared a high degree of feature overlap with any true headlines. Instead, Jones (2013) attributes her results to knowledge schemas that influenced retrieval processes. Regardless of her conclusion, her results prove to be problematic for the current DPSD model and falls outside the boundary conditions of recollection. According to Yonelinas et al.'s (2010) boundary condition explanation, a nonstudied item must share enough features with a studied item for contextual information to be retrieved. However, Jones (2013) demonstrates that prior knowledge related to the topic of a nonstudied headline led to false recollection. These false headlines did not share similar perceptual or semantic information with true headlines. The only information that was shared was the topic of the headline. Therefore, high-knowledge false headlines should not have shared enough features with true headlined events to surpass the retrieval criterion (Norman & O'Reilly, 2003). It could be that false recollection of the false headline events could have been due to an over-generalization of knowledge structures, but it is unclear as to how those knowledge structures influenced retrieval. Jones (2013) results demonstrate that false recollection due to prior knowledge can occur.

Stimulus Fluency

False recollections due to implicit associations, source misattributions, and prior knowledge, all occur because of factors that result in the retrieval of contextual information—whether veridical or misattributed—lead to false recollection. However, false recollection has been shown to also occur due to factors that occur during retrieval.

Higham and Vokey (2004) found that increasing the duration of a stimulus during test impacted familiarity and recollection, and also produced false recollections. Although this manipulation does not directly manipulate contextual information, this study does highlight additional conditions that violate Yonelinas's boundary condition explanation of recollection. Higham and Vokey (2004) increased the duration that a target stimulus was presented to hypothetically also increase the fluency of that target stimulus. Fluency of a stimulus has been shown to influence familiarity, but not recollection (Higham & Vokey, 2004). Therefore, Higham and Vokey (2004) hypothesized that increasing the presentation time of a stimulus should impact Know judgments without impacting Remember judgments. Participants studied a list of 74 five-lettered words by reading them aloud. After study, participants were given an old/new recognition test and a Remember/Know test. At test, participants were told to watch the screen for each word. A fixation stimulus was presented on the screen, and participants were instructed to look at the screen. After the researcher confirmed that the participant was looking at the screen, a word was briefly presented and then masked with ampersands. Presentation duration of the stimulus varied: half were presented for a short duration (6-34 ms) and the other half were presented for a long duration (36-64 ms), depending on the refresh rate of the monitor. Participants were instructed to write down the word that was presented, guessing if needed. After writing down the word, participants pressed the space bar to reveal the word that had been presented and indicated whether they had seen the word before (i.e., old/new judgments) as well as whether they remember seeing the word or just knew they had seen it previously (i.e., Remember/Know judgments).

Participants were more likely to false alarm to new items that were presented for a long-than a short-duration, demonstrating an impact of presentation time on recognition. Overall, nonstudied items were more likely to receive Know than Remember judgments; however, the probability that a nonstudied item received a Remember judgment significantly increased when the item was presented for a longer duration. This finding is problematic for the DPSD model because, in its current state, it cannot account for why the systematic manipulation of the perceptual features of an item impacted false recollection when the manipulation of presentation time should have only impacted the process of familiarity. Higham and Vokey (2004) showed that presentation time impacted recollection for both studied and nonstudied items, suggesting that contextual information was retrieved for both items. However, the presentation duration of a stimulus during retrieval should not have had any known impact on access to contextual information related to a studied event. Therefore, it could be that the increased amount of presentation time allowed for the retrieval of contextual information.

Although Higham and Vokey (2004) did not provide an explanation for their finding, they did emphasize the fact that their findings cannot be accounted for by current models of recognition. According to Yonelinas et al. (2010) false recollection is an issue that violates the boundaries of recognition within the DPSD model. However, Higham and Vokey's (2004) findings provide an additional example of the need to modify the model. Because Yonelinas states that only information that shares a high level of feature overlap can evoke false recollection, the impact of stimulus duration on false recollection falls outside of that boundary. Therefore, Higham and Vokey's (2004) findings cannot be explained by the DPSD model in its current state. That the DPSD model cannot account for Higham and Vokey's (2004) findings, further demonstrates that our current understanding of how recognition retrieval processes operate is limited.

Summary

Although evidence from Roediger and McDermott (1995), Lane and Zaragoza (1995), and many others (e.g., Dennis et al., 2014; Gallo et al., 2009; Howe, 2007; Schacter, Koutstaal, Johnson, Gross, & Angell, 1997) has demonstrated that, not only do false recollections occur, they can also occur at a higher frequency than 0-5%. However, their results must be tempered by the fact that they were obtained using Remember/Know judgments and not a 6-point confidence recognition scale. A 6-point confidence recognition scale is the typical scale used to measure recognition in a typical DPSD recognition paradigm. The 6-point confidence recognition scale is necessary because of its ability to provide a finer measure of the degree to which participants recognize items at test than Remember/Know judgments. The 6-point confidence recognition scale also has the advantage of mapping onto a signal-detection distribution. For instance, a 5-response on a 6-point confidence recognition scale demonstrates that the item was recognized as old with a certain level of confidence. However, it also reveals that the item was not recognized with a high-level of confidence, suggesting that it was retrieved via familiarity. It is assumed that Remember/Know judgments measure recollection versus familiarity (Higham & Vokey, 2004; Jones, 2013; Lane & Zaragoza, 1995; Roediger & McDermott, 1995) and that Remember judgments are analogous to a 6-response on a 6-point confidence recognition scale, but the Remember/Know scale has been subject to criticism in terms of their ability to reliably measure recollection and familiarity (e.g., Wais, Mickes, & Wixted, 2008; Wixted & Mickes, 2010).

Although Remember/Know judgments have been theorized to be directly correlated with recollection and familiarity, respectively, several studies have demonstrated that these judgments can sometimes be skewed because of their dichotomous nature (e.g., Gardiner, Ramponi, & Richardson-Klavehn, 1998; Hicks & Marsh, 1999). Specifically, Gardiner et al. (1998) found that when participants were not given a third choice of “Guess” during a Remember/Know task, they were more likely to assign either Remember or Know judgments to items that were not based on a memory for that item (i.e., participants were guessing). However, if participants were given the opportunity to provide a “Guess” response, the accuracy of the Remember/Know judgments increased, as evidenced by verbal reports. Therefore, the validity of the Remember/Know judgments provided in the Roediger & McDermott (1995), Lane and Zaragoza (1995), Jones (2013), and Higham and Vokey (2004) studies can be challenged. It should be noted that participants in Roediger & McDermott (1995) and Lane and Zaragoza (1995) were only asked to provide Remember/Know judgments to items which they had previously identified as old. Therefore, they were given the opportunity to reject nonstudied items as new before providing a Remember/Know judgment. Participants in Jones (2013) study were also given the opportunity to provide a Neither judgment and did provide Neither judgments for approximately 95% of false headlines, demonstrating their willingness to decisively use Remember/Know judgments. Additionally, participants were told to only provide a Remember judgment for items in which they could retrieve contextual information.

Although the scale that was used in Roediger and McDermott (1995), Lane and Zaragoza (1995), Jones (2013), and Higham and Vokey (2004) can be criticized, the results from each of these studies demonstrate that false recollection can occur and be manipulated by certain experimental conditions. Yonelinas et al. (2010) acknowledged one way false recollection could occur was under conditions of high similarity. However, Lane and Zaragoza (1995) demonstrated that not only were people able to recollect veridical contextual source information for studied items, but they were also able to recollect false contextual source information for studied items. This finding is problematic for the DPSD model in that it demonstrates that contextual information retrieved for studied items can be incorrect. Therefore, false recollection can occur in one of two ways: a) retrieval of contextual information for nonstudied items, resulting in recollection for an event that never occurred and b) retrieval of inaccurate contextual information for a studied item. If incorrect contextual information is retrieved for a studied item, then that information is false, resulting in a false recollection. Lane and Zaragoza (1995) demonstrated that false recollection can occur in more than one way, further highlighting the DPSD model's limited ability to explain all aspects of the recollection process.

Although false recollection has been acknowledged as a possible boundary condition (Norman, 2010; Yonelinas et al., 2010), the fact that the DPSD model has not been amended to account for their occurrence is problematic in two ways. First, because the DPSD model has not been amended to include a false recollection parameter, measuring the extent to which false recollection occurs is not possible. Instead, other methods such as Remember/Know judgments must be used to measure the extent to which false recollection occurs. This difference in methodology can be problematic because it leaves open the possibility for measurement differences between the two. The second problem that results from the DPSD model not being amended to include a false recollection parameter is a more global problem. Because the model does not account for false recollection, it limits our understanding of recognition. The purpose of a computational model is to further our understanding of the cognitive processes that they represent. Therefore, by failing to model all parts of a cognitive process, the current model limits our ability to fully understand how recognition operates.

One solution to this problem would be to investigate false memories using the same methodology that is used in traditional recognition paradigms. Specifically, instead of using Remember/Know judgments, we should use a 6-point confidence recognition scale. By using a 6-point confidence recognition scale in a paradigm designed to increase the number of false recollections, we can test the DPSD model's currently ability to account for this data. Unfortunately, most false memory paradigms have not been tested using a 6-point confidence recognition scale and provide few instances of false recollection. The Payne-Eakin paradigm (Payne & Eakin, 2017) was designed to address both of these shortcomings.

The Payne-Eakin Paradigm

Payne and Eakin (2017) developed a paradigm for investigating false memories in recognition memory that capitalized on the finding of the DRM paradigm that implicit associate among words produced false recollections, as suggested by the theory of associative processes (Roediger et al., 1998). However, unlike Roediger et al.'s (1998) theory, Payne and Eakin (2017) quantified the associations in their paradigm using the Processing Implicit and Explicit Representations 2 (PIER2) model; PIER2 specifies how the associative relationships between words operate (Nelson et al., 1998). According to

PIER2, when a cue and target word pair are explicitly studied together, all words that are associated with both the cue and target are activated implicitly through links in long-term working memory (Ericsson & Kintsch, 1995). PIER2 states that presenting two words as pairs not only activates the explicit representation of the two presented words, but also activates words that are implicitly associated with both the cue and target (words that are implicitly associated with only the cue or only the target falls away from activation).

Payne and Eakin (2017) capitalized on these explicit and implicit associative relationships among words to develop a false memory paradigm based on principles described by the PIER2 model. Cue-target word pairs were created by consulting the University of South Florida Word Association Norms database (Nelson et al., 1992). The USF Norms specify the forward and backward association strength of over 10,000 words as well as measures of the number of associates of each word. For the Payne-Eakin paradigm, a word list was created using the USF Norms that consisted of 72 associated cue-target word pairs (e.g., GARLIC-SPICE). These pairs also were required to have a shared associate that had a similar forward association strength to both the presented cue and the presented target (mean characteristics of the actual words are detailed in the method section). According to PIER2, when the studied word pairs were presented, not only were the explicitly presented words activated, but also the shared associates of both words were activated implicitly. Figure 8 depicts the studied, associated cue-target pair, GARLIC-SPICE. Note that two implicit associates, SALT and SEASONING, are also activated during study; because they also are associates of both GARLIC and SPICE, activation spreads to these words as well. Activation does not spread in the same manner to SMELL, because it is not an associate of both GARLIC and SPICE (PIER2; Nelson et al., 1998). Therefore, although SEASONING was not explicitly studied, in an implicit sense, because it is an associate of both the cue-target pair that was explicitly studied, this associate is also "studied" and should be a part of the encoded representation. This factor becomes important for the Payne-Eakin paradigm at recognition.

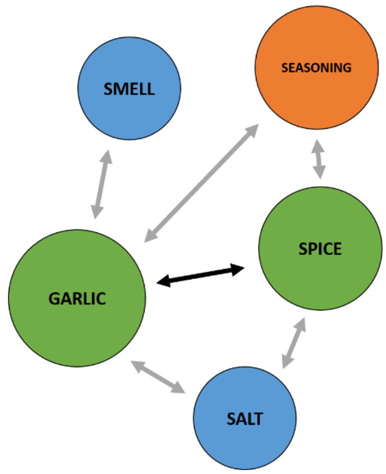


Figure 8. A graphical depiction of the implicit and explicit associations between cue-target word pairs and their associates. The black arrow indicates the association between the explicitly presented cue and target; the gray arrows indicate activated implicit associates between the explicitly presented cue/target.

Three types of pairs are presented for recognition: a) a studied cue-target word pair, b) a false-target word pair, and c) a control word pair. The studied word pair is a re-presentation of one of the studied cue-target pairs. False-target test pairs are created by re-pairing a studied cue with one of its associated words that was activated implicitly during study of the original cue-target word pair. Control pairs were new, nonstudied test pairs and served as the baseline false alarm rate for the study. According to PIER2, presenting a studied word pair for recognition activates the same associations—both explicit and implicit—present at study; the associates activated at study and test are the same, facilitating recognition. For false-target test pairs, because the target was previously implicitly activated during study, presenting it explicitly with the studied cue, creates a similar activation set at recognition. The difference is that the explicitly presented target was only activated implicitly, not explicitly, during study. Payne and Eakin (2017) hypothesized that this explicit presentation of the previously implicitly activated target would produce a false recollection.

In the first test of the Payne-Eakin paradigm (2017a), participants studied the list of 72 related word pairs (e.g., GARLIC – SPICE) at a rate of six seconds each. After a 10-minute distractor task, participants were given a standard old/new recognition test containing a total of 102 word pairs. Of the 102 word pairs, 36 were studied pairs (e.g., GARLIC – SPICE), 36 were false-target pairs (e.g., GARLIC – SEASONING), and 20 were control pairs (e.g., SCHEDULE – MEETING). False-target test pairs were significantly more likely to be falsely recognized (M = .22) than control test pairs (M = .11), demonstrating a false memory effect for the false-target test pairs.

Although the Payne-Eakin paradigm (2017) produced false recognition, the paradigm had not yet been tested in its ability to produce false recollections using either Remember/Know judgments or a 6-point confidence recognition scale. Because of the prior criticism of Remember/Know judgments, a demonstration of the Payne-Eakin paradigm's ability to produce a reliable false recollection effect using a 6-point confidence recognition scale is required and served Aim 1 of this dissertation. Therefore, a series of pilot experiments was conducted to determine whether using a 6-point confidence recognition scale could produce a false recollection effect.

Pilot Experiment 1

Payne and Eakin (2017) tested the paradigm by using a 6-point confidence recognition scale instead of old/new judgments to determine the degree to which people were experiencing recollection or familiarity when assessing recognition of the false-target test pairs. In this study, participants studied a list of 72 related word pairs (e.g., GARLIC-SPIC) at a rate of six seconds each. After a 10-minute distractor task, participants were given a 6-point confidence recognition test consisting of a 102 test word pairs: 36 were studied pairs (e.g., GARLIC – SPICE), 36 were false-target pairs (e.g., GARLIC – SEASONING), and 20 were control pairs (e.g., SCHEDULE – MEETING). Participants falsely recognized control test pairs at a rate of .14 (*SD* = .15); however, they falsely recognized false-target test pairs at a nearly double rate of .27 (*SD* = .19). Table 2 reports all means and standard errors for the conditions. A paired samples *t*-test between the false-target false-alarm rate and control false-alarm rate was significant, $t = 7.88$, $p < .001$. Consistent with Payne and Eakin (2017), participants were significantly more likely to false alarm to the false-target test pairs than to control test pairs. In addition, significantly more 6-responses were given to false-target (*M* = .17, *SD* = .15) than control test pairs (*M* = .04, *SD* = .06), $t(66) = 8.06$, $p < .001$. Table 3 lists the proportions and standard errors. Figure 9 shows the response distribution of the three test pair types. These results demonstrate that the Payne-Eakin paradigm could produce a significant false recollection effect.

Table 2

Pilot Experiment 1 Recognition Rates and Standard Error

Recognition	Mean	SE
Hit Rate	.92	.009
False-Target FA Rate	.27	.02
Control FA Rate	.14	.02

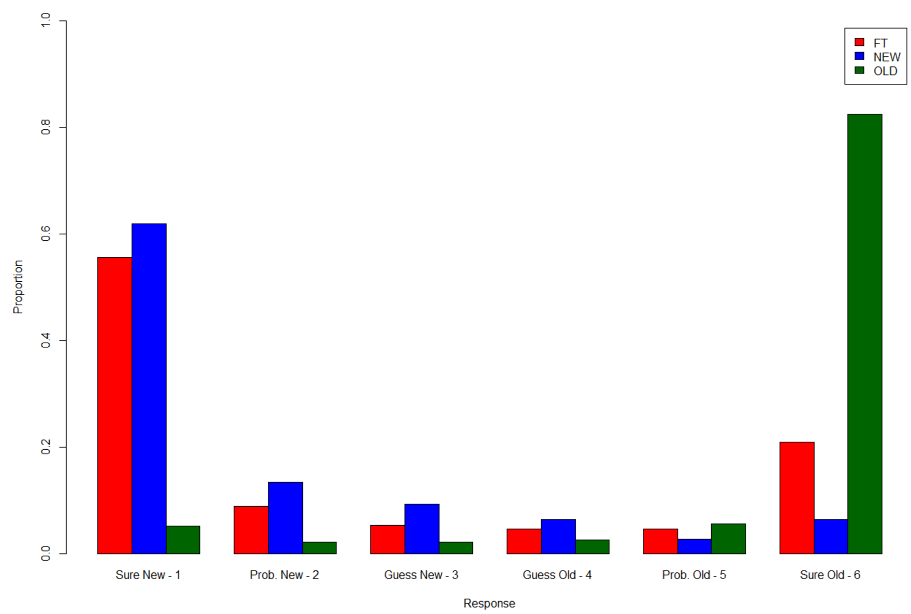


Figure 9. Proportion of responses by word pair type for Pilot Experiment 1.

The goal of Pilot Experiment 1 was to determine whether the Payne-Eakin paradigm can produce a false recollection effect using a 6-point confidence recognition scale. The results from the pilot experiment demonstrated that, overall, the paradigm can produce a reliable false recollection effect; a significantly higher rate of false recollections occurred for false-target test pairs (M = .18, SE = .02) than for control test pairs (*M* = .05, *SE* = .01), *t*(66) = 8.06, *p* < .001. Although Pilot Experiment 1 demonstrated that the Payne-Eakin paradigm could produce false recollections as defined using a six-point confidence recognition scale, the study was limited by lack of variability in responses on the scale; ideally, participants should use the full range of the scale when making their confidence judgments. In fact, to model the recognition data using ROC curves, it is critical that there be a distribution of judgments along the scale. One reason participants tended to use the scale more dichotomously could have been because the recognition task was not difficult; the mean hit rate was .92 (*SD* = .07). To test whether this high hit rate contributed to the lack of variability in responses along the entire scale, a second pilot experiment was conducted to reduce the hit rate by increasing the retention interval between study and test, as well as making the distractor task more difficult.

Pilot Experiment 2

The goal of Pilot Experiment 2 was to determine whether more variation in responses could be obtained with the implementation of a longer and more difficult distractor task between study and the recognition test. Rather than complete the Stroop during the interval, participants completed the operational span (Ospan) task for a longer period (Unsworth, Heitz, Schrock, & Engle, 2005). This task was chosen for its difficulty, which both prevented rehearsal of the pairs as well increasing the length of the interval between study and test.

In this study, participants studied the same list of 72 related word pairs (e.g., GARLIC-SPICE) at a rate of six seconds each. After study, the participants completed the Ospan task (Unsworth et al., 2005). The Ospan task consisted of two parts: memorizing a string of random letters and solving math problems. Participants memorized fourteen lists of random letters ranging from three to eight letters. After participants studied each list of letters, participants were given a math equation to solve. The math equations ranged from simple addition problems such as 2 + 2 = x to more difficult problems such as (2*5) + 4 = x. After participants finished solving each equation, they were asked to recall the string of letters that was shown to them before in the correct order. This task lasted approximately 10 minutes. After the Ospan task, participants were given a 6-point confidence recognition test containing the same total of 102 test pairs used in the first pilot experiment.

Overall, the implementation of the Ospan task was successful at reducing the hit rate for studied pairs from .92 (*SD* = .02) in Pilot Experiment 1 to .88 (*SD* = .11) in this pilot. A larger proportion of nonstudied items were falsely recognized, showing that the longer, more difficult interval also impacted the overall false alarm rate. Participants falsely recognized false-target test pairs at a rate of .36 (*SE* = .03), which was significantly higher than the false recognition of control test pairs of .23 (*SE* = .03), $t(50) = 7.03$, $p < .001$. The false-alarm rate was further divided into 4-, 5-, and 6- responses for both false-target and control test pairs. False-target test pairs produced high-confidence 6-responses at a rate of .22 (*SE* = .02), which was an increase from .18 (*SE* = .02) in Pilot Experiment 1; control test pairs only produced high-confidence 6-response at a rate of .09 (*SE* = .02). A paired samples *t*-test found that the false-target rate of false recollection was significantly higher than the control rate of false recollection, indicating that participants were significantly more likely to falsely recollect the false target than the control test pairs, $t(50) = 7.89$, $p < .001$.

The main goal of this pilot experiment was to determine whether an increased delay filled with a more difficult distractor task could produce more variability in the responses obtained with a 6-point confidence recognition scale using the Payne-Eakin paradigm. Overall, we did find a slight increase in the variability of responses given to each type of word pair (See Figure 10). However, the increase was not sufficient to fulfill Aim 4. Therefore, additional changes were implemented for the dissertation experiment to bring down the hit rate and increase response variability. One change that was implemented is that the duration of the Ospan task was increased to 40 minutes. A second change that was implemented is that several more word pairs were added during study and test, increasing the difficulty of discriminating between studied and nonstudied pairs. These changes are discussed in further detail in the method section.

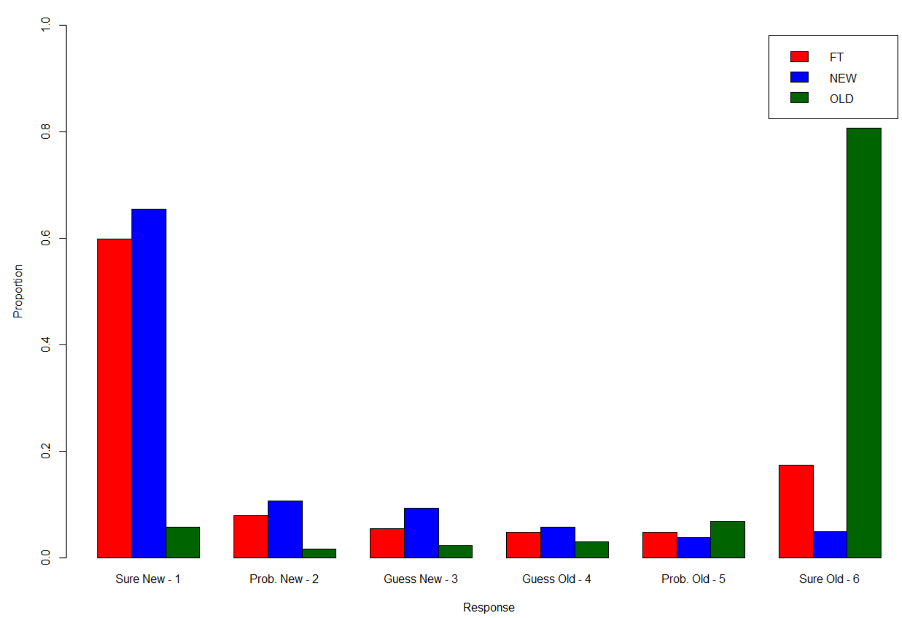


Figure 10. Proportion of recognition responses by word pair type for Pilot Experiment 2.

CHAPTER IV

SPECIFIC AIMS

Aim 1

To demonstrate whether the false memory effect obtained using the Payne-Eakin paradigm is due to the activation of an implicitly associated word during study of an explicitly presented cue-target pair.

One purpose of the dissertation was to determine whether the false memories obtained using the Payne-Eakin paradigm were due to the activation of the implicitly associated word during study of the explicitly presented cue-target paired associate. The typical Payne-Eakin paradigm was modified to compare the false-target test pair—the studied cue along with the implicitly activated false target (FTi)—to a condition for which there should, theoretically, be no implicit associate activated at study (FTx). The FTx test pairs were created by combining the studied cue with a new associated target; the associate was only associated with the studied cue, not the studied target. These pairs, according to PIER2, should not have been implicitly activated when the original cue-target pair was studied; therefore, the test target should be completely new.

The associative recognition of the FTi, FTx, and control test pairs was measured using a 6-point confidence recognition scale with a 6-reponse indicating that the test pair was studied with high confidence and a 1-response indicating that it was not studied with high confidence. One reason for using this scale was to determine whether the false memories obtained were the result of recollection or high familiarity, as defined by Yonelinas et al. (2010). The comparison between the proportion of 6-responses for FTi and control test pairs as well as the comparison between the proportion of 6-responses for FTi and FTx test pairs will test both the assumptions of the PIER2 and CLS model. According to the PIER2 model, FTi test pairs should be falsely recognized at a significantly higher rate than FTx and control test pairs because of the activation of the implicit associate at study. According to the CLS model, the result of this implicit activation is that FTi test pairs should be falsely recognized at a higher rate than FTx and control test pairs due to a match in the hippocampus which allows for recollection. Therefore, FTi pairs should also receive a higher proportion of 6-responses on the scale than FTx or Control test pairs.

H1: FTi test pairs will be falsely recognized at a higher rate than R-control test pairs, evidenced by a significant *a priori* planned comparison *t*-test comparison between the proportion of false alarms for the two test conditions.

H2: FTx test pairs will be falsely recognized at the same rate as R-control test pairs, evidenced by a nonsignificant *a priori* planned comparison *t*-test comparison between the proportion of false alarms for the two test conditions.

H3: FTi test pairs will be falsely recognized at a higher rate than FTx test pairs, evidenced by a significant *a priori* planned comparison *t*-test comparison between the proportion of false alarms for the two test conditions, due to the implicit associate in FTi test pairs.

H4: FTi test pairs will be falsely recollected at a higher rate than R-control test pairs, evidenced by a significant *a priori* planned comparison *t*-test comparison between the proportion of 6-responses for the two test conditions.

H5: FTx test pairs will be falsely recollected at the same rate as R-control test pairs, evidenced by a nonsignificant *a priori* planned comparison *t*-test comparison between the proportion of 6-responses for the two test conditions.

H6: FTi test pairs will be falsely recollected at a higher rate than FTx test pairs, evidenced by a significant *a priori* planned comparison *t*-test comparison between the proportion of 6-responses for the two test conditions, due to the implicit associate in FTi test pairs.

Aim 2

To examine whether the false memories obtained using the Payne-Eakin paradigm are due solely to familiarity or a combination of familiarity and recollection by examining the pattern of given responses.

According to Elfman et al. (2008), responses resulting from recollection should produce a different distribution than responses resulting from familiarity. Because recollection is characterized as an all-or-none process, items impacted by this process should lead to a disproportional increase in the number of high-confidence old responses. However, because recollection has also been characterized to aid in the rejection of items during paired associate recognition (Norman & O'Reilly, 2003; Rotello & Heit, 2000), this rejection should also lead to a disproportional increase in the number of high-confidence new responses, creating a more dichotomous response distribution. In contrast, because familiarity is described as a signal-detection process, it should produce a more Gaussian-shaped distribution (Elfman et al., 2008). Therefore, the following hypotheses compare the shape of response distributions across test pair conditions by comparing histograms and ROC curves of the responses.

H7: Response distributions for FTi test pairs will be more dichotomous, with most responses falling into either the 1-response category or the 6-response category on the 6-point confidence recognition scale, than response distributions for R-control test pairs. The response distribution shape for the FTi test pairs will be examined using histograms and ROC curves of the responses. We predict that the ROC curve with FTi test pairs serving as the false alarm rate will be shifted to the right and more linear than the ROC curve with R-control test pairs serving as the false alarm rate.

H8: Response distributions for FTx test pairs will be comparable to R-control test pairs, with more responses falling into the 2-, 3-, 4-, and 5-response categories on the 6-point confidence recognition scale because of the larger contribution of familiarity. The response distribution shape for the FTx test pairs will be examined using ROC curves of the responses. We predict that the ROC curve for FTx and R-control test pairs will be similarly curvilinear.

Aim 3

To determine whether the re-presentation of both a studied cue and target, though not together, increases the level of familiarity associated with the pair presented at test.

Both the FTi and FTx test pairs have in common that they repeat a studied cue. Because the cue was previously studied, familiarity could play a role in recognition of these pairs over control pairs. However, the unique role of familiarity is difficult to determine because the FTi and FTx pairs would make similar predictions regarding familiarity. Therefore, a third type of word pair was included to test the potentially differential impact of familiarity on recognition. A Re-Paired test pair consisting of a studied cue and studied target, that were not studied together, was created. Because both the cue and target were studied—albeit not together—their overall level of familiarity should be higher than FTi or FTx pairs whose target was never studied. However, one practical issue arose when creating these test pairs. All of the studied pairs were related cues and targets and re-pairing these study pairs would have resulted in unrelated test pairs, making them easily identifiable as new because no unrelated test pairs were studied. Therefore, a new set of unrelated test pairs were added to the studied list and only these unrelated pairs were re-sorted to serve as the Re-Paired test pairs. In addition, a new set of nonstudied unrelated test pairs (U-control) were created to serve as controls for these Re-Paired test pairs.

Although familiarity should be high because both the cue and target of the Re-Paired test pairs was previously studied, these words should not be recollected. Because they are unrelated, they do not share any associates regardless of how they were recombined; therefore, the Re-Paired target should never be an associate of the studied cue-target pair or of the test cue. Any familiarity associated with the Re-Paired test pair should stem exclusively from having been seen at study. Additionally, according to the CLS model, because these Re-Paired test pairs were never studied together, the pairs should be easily rejected by the hippocampus as new due to a mismatch; the presented pair's representation should have no overlap with the studied pair's representation in the hippocampus. However, due to their potential for high familiarity, the Re-Paired test pairs should create a high level of activation in the MTLC, leading them to be more likely to be falsely recognized due to familiarity.

H9: Re-Paired test pairs will be falsely recognized at a higher rate than U-control test pairs, as evidenced by a significant *a priori* planned comparison *t*-test comparison between the proportion of false alarms for the two conditions, due to the increased familiarity with the cue and target in the pair.

H10: Response distributions for Re-Paired test pairs will be more Gaussian than U-control test pairs, with more responses falling into the 2-, 3-, 4-, and 5-response categories on the 6-point confidence recognition scale, indicating familiarity. The response distribution shape for the FTi test pairs will be examined using ROC curves of the responses. We predict that the ROC curve with Re-Paired test pairs serving as the false alarm rate will be more curvilinear than the ROC curve with U-control test pairs serving as the false alarm rate.

Aim 4

To determine whether the inclusion of a false recollection parameter is necessary for fitting the false recognition produced by the behavioral data.

According to the DPSD model, recollection does not occur for nonstudied items (Yonelinas, 1994, 2002; Yonelinas et al., 2010); therefore, false recollection cannot occur. Although there is behavioral evidence to suggest otherwise (e.g., Higham & Vokey, 2004; Lane & Zaragoza, 1995; Roediger & McDermott, 1995), modeling the data using the DPSD model allows for a direct test of the unique contribution of recollection versus familiarity. Therefore, in addition to the behavioral analyses, we will also be using statistical modeling to determine whether any false memory effects obtained in the experiment were due to recollection or familiarity. The standard DPSD model will be fit to the data from the experiment. We will also fit two modified versions of the DPSD model that allow for false recollection. Using model comparison tests, such as the Akaike Information Criterion (AIC; Akaike, 1974) and Bayesian Information Criterion (BIC; Schwartz, 1978), we will compare the fit of each model to determine which model provides the best account of the data.

H11: A model allowing for false recollection of the FTi test pairs will provide a better account of the behavioral data better than the standard DPSD model as evidenced by AIC and BIC tests.

CHAPTER V

DISSERTATION EXPERIMENT

The aim of this experiment was to determine whether the false memory effect obtained by the Payne-Eakin paradigm is due to recollection or familiarity using a modified version of the Payne-Eakin paradigm. The Payne-Eakin paradigm is a recognition paradigm designed to induce false memories by capitalizing on pre-existing explicit and implicit associative relationships between words described by the PIER2 model (Nelson et al., 1998). In the paradigm, cue-target word pairs were created using the University of South Florida Word Association Norms database. In the original Payne-Eakin paradigm, implicit associative relationships between words were manipulated to create three types of stimuli: studied word pairs, false-target test pairs, and control test pairs. However, the original paradigm failed to provide experimental conditions that could test whether the false memory effects obtained by the paradigm are due to errors in recollection or familiarity. Therefore, for the current study, we added three additional word pair types for a total of six word pair types: Studied, false target implicit (FTi), false target no implicit (FTx), Re-Paired (Re-Paired), related control (R-control), and unrelated control (U-control). In the new paradigm, studied word pairs served as our baseline hit condition, or the rate at which participants were able to identify previously studied word pairs; FTi test pairs capitalized on the implicit associative relationship among words, testing for false recollection; FTx test pairs provided a contrast condition to the FTi condition by using test pairs that had a repeated studied cue Re-Paired with a nonstudied target that was not implicitly related to the original studied target; Re-Paired test pairs tested for the influence of familiarity by re-pairing the targets of unrelated test pairs to different studied cues, thus testing for the influence of overall familiarity from having studied both words previously; R-control test pairs provided a baseline false alarm rate for related word pairs; and U-control test pairs provided a baseline false alarm rate for unrelated word pairs.

Method

Participants and Design

A total of 54 Mississippi State University undergraduate students recruited via the Psychology Research Program SONA-system website participated in this study. The sample size for this study was generated according to a sample size calculation conducted using the G*Power analysis program (Faul, Erdfelder, Lang, & Buchner, 2007), which resulted in the requirement of 15 participants. However, prior studies that used this paradigm had a minimum of 50 participants. As a result, we collected 54 participants. Participants were at least 18 years old and native English speakers. The experiment was a single-factor (Word Pair Type: Studied, FTi, FTx, Re-Paired, R-control, U-control) within-subjects design.

Materials

Word list. A word list was created that consisted of six different types of word pairs: studied, false target implicit (FTi), false target no implicit (FTx), Re-Paired, related control (R-control), and unrelated control (U-control). The list was created using the

University of South Florida Word Association Norms (Nelson, McEvoy, & Schreiber, 2004). The associations between the pairs were confirmed using ListChecker Pro 1.2 (Eakin, 2010), a program designed to examine word associations using the University of South Florida Word Association Norms (Nelson et al., 2004) as its database. ListChecker Pro 1.2 (Eakin, 2010) was also used to test the requirement that each word was associated only with its intended pair and not with any of the other words on the list. There were 108 related word pairs and 30 unrelated word pairs for a total of 138 word pairs presented at study. At test, there were 36 studied test pairs, 36 FTi test pairs, 36 FTx test pairs, 30 Re-Paired test pairs, 20 R-control test pairs, and 10 U-control test pairs, for a total of 168 test pairs. A list of each word pair type and examples of each are listed in Table 3 below.

Table 3

List of Word Pair Types and Examples for the Dissertation Experiment.

Study	**Test**	
	Example:	Word Pair Type:
DISCREET – PRIVATE	DISCREET – PRIVATE	Studied
FLANNEL – MATERIAL	FLANNEL – CLOTH	FTi
AIRCRAFT – PILOT	AIRCRAFT – CARRIER	FTx
MARSHMALLOW – HUT	MARSHMALLOW – ADDITION	Re-Paired
ADDITION – SHELL	ATMOSPHERE – OZONE	Related Control
-----	QUILL – TRICYCLE	Unrelated Control

False Target Implicit (FTi). A list of 72 related word pairs (e.g., GARLIC – SPICE) was created for the study phase. For each word pair in the list, a second target, associated with both the cue and target, was selected (e.g., GARLIC – SPICE; GARLIC –

SEASONING; see Appendix A for full list of word pairs and characteristics). The forward associative strength (FAS; likelihood of producing the target, given the cue) from the cue to both targets was controlled for both targets. The FAS mean difference between the two targets was .02 (*SD* = .13), and the BAS mean difference was .01 (*SD* = .11), both of which are allowable differences according to typical list-construction constraints. (These associative factors were controlled for although association strength is less likely to impact recognition of cue-target pairs than cued recall of the target, given the cue.) Each cue was associated only with its two targets and not to any other cue on the list. Each of the targets were associated only with their cue and with each other, not to any other target on the list.

The list of 72 word pairs and their corresponding implicit false target was divided into two separate lists resulting in two lists of 36 studied word pairs and two lists of 36 FTi word pairs; having two lists allowed for an examination of whether any effects obtained on one list could generalize to a different list. List 1 and List 2 were counterbalanced across participants at study. The studied target versus the implicit false target used within each list was not counterbalanced within each list; one target always served as the studied target and the other target always served as the false target. The reason for this was that, although the implicit false targets selected were always an implicit associate of both the studied cue and target, the studied target was not always an implicit associate of the false target. Therefore, the false target could not always serve as a studied target because it would not implicitly activate that target during study. Also, as shown in Appendix A, the words were organized as studied and implicit false targets in such a way as to allow that the mean characteristics of set size, concreteness, printed word frequency, the number of mediators and the number of shared associates to be equated for the two target types; each characteristic is defined in Appendix A. The targets' mean characteristics were also equated with the mean characteristics of the cue. Having tight control over these characteristics necessitated that they serve in static positions on the list; this method is conventional for list construction (Eakin, 2010).

False Target No Implicit (FTx). A list of 36 related word pairs (e.g., AIRCRAFT – PILOT) was created for the study phase. For each word pair in the list, a second target, associated with only the cue, was selected (e.g., AIRCRAFT – PILOT; AIRCRAFT – CARRIER; see Appendix B for full list of word pairs). The forward associative strength (FAS; likelihood of producing the target, given the cue) from the cue to both targets were controlled for both targets (see Appendix A for a full list of mean word pair characteristics by word pair type). The FAS mean difference between the two targets was .02 (*SD* = .09), and the BAS mean difference was .01 (*SD* = .11), both of which are allowable differences according to typical list-construction constraints. Each cue was associated only with its two targets and not to any other cue on the list. Each of the targets were associated only with their cue and not to any other target on the list.

The list of 36 word pairs and their corresponding FTx was counterbalanced across participants at study. Half of the participants studied AIRCRAFT – PILOT and received AIRCRAFT – CARRIER at test while the other half studied AIRCRAFT – CARRIER and received AIRCRAFT – PILOT at test. Because the studied target and FTx were always associated with the cue, and not each other, they could be counterbalanced between serving as the studied target or the FTx. Also, as shown in Appendix A, the words were organized in such a way as to allow that the mean characteristics of set size, concreteness, printed word frequency, the number of mediators and the number of shared associates to be equated for the two target types; each characteristic is defined in Appendix A.

Re-Paired. A list of 60 unrelated words was created for the study phase. This list of 60 words was randomly arranged and paired with another word from the list to create list of 30 unrelated word pairs (e.g., NEST – PRINCE; JOKE – CONTEST). At test, the studied target was randomly Re-Paired with another studied cue to create a new test pair (e.g., NEST – CONTEST). Each cue and target were not associated with each other or with any other word on the list. The list of 30 test pairs was counterbalanced across participants at study. Half of the participants studied NEST – PRINCE and received NEST – CONTEST at test while the other half of participants studied NEST – CONTEST and received NEST – PRINCE at test.

Related Control (R-control). A list of 20 related test pairs (e.g., APPOINTMENT – SCHEDULE) was created for the test phase. These pairs were not studied; they appeared on the test to serve as new items. Each of the cues and targets were associated only with their intended cue and/or target, not to any other word on the list.

Unrelated Control (U-control). A list of 20 unrelated words (e.g., PRINCIPAL – MELODY) was created for the test phase. This list of 20 words was randomly arranged and paired with another word from the list to create list of 10 unrelated test pairs. These pairs were also not studied, but served as new, unrelated pairs on the test.

Recognition Test. Half of the FTi word pairs were selected to serve as the studied cue-target test pairs on the recognition test. The other half of the FTi word pairs were modified to serve as the FTi test pairs on the recognition test; the cue was Re-Paired with its false target. The recognition test consisted of 36 studied test pairs, 36 FTi test pairs, 36 FTx test pairs, 30 Re-Paired test pairs, 20 R-control test pairs, and 10 U-control test pairs resulting in a total of 168 test pairs.

6-point recognition confidence scale. The 6-point recognition confidence scale was used as the response on the recognition test. This scale is a 6-point scale ranging from 1 (Sure New) to 6 (Sure Old) responses. Responses 1 to 3 were located on the left side of the screen with the label "NEW" above the responses, and the labels "Sure" underneath the 1-response and "Unsure" underneath the 3-response. Responses 4 to 6 were located on the right side of the screen with the label "OLD" above those responses, and the labels "Unsure" written underneath the 4-response and "Sure" underneath the 6-response. See Figure 11 for a view of the scale. Participants were instructed to type in their responses using the keyboard.

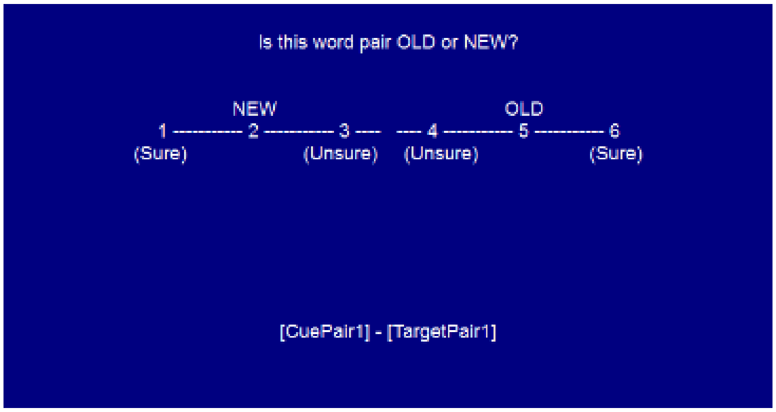


Figure 11. The 6-point recognition confidence scale presented during the recognition test.

Ospan Task. The operational span (Ospan) task consisted of two parts: memorizing a string of random letters and solving math problems. Participants memorized twenty-five lists of random letters ranging from three to eight letters. After participants studied each string of letters, participants were shown a math equation to solve. The math equations ranged from simple addition problems such as 2 + 2 = x to more difficult problems such as (2*5) + 4^2 = x. After participants finished solving each equation, they were asked to recall the string of letters in the correct order that previously was shown to them.

Consent Form. An IRB-approved and stamped informed consent form was used to inform participants about the general aims and procedures of the experiment. It also explained the minimal risks involved as well as the benefits for participating in this study. Participants were informed that their participation in the study was voluntary and that they may choose to cease participation at any time during the study without penalty (see Appendix C).

Debriefing Form. A debriefing form was given to participants with more detailed post-experiment information. Participants were told about the purpose of the study and the experimental manipulation, as well as given contact information in case they have any further questions (see Appendix D).

Procedure

Participants were tested one at a time. Each participant sat at a testing station equipped with a computer, chair, consent forms, and white noise machine to block out any extraneous noise. After the consent forms were read, signed, and collected, participants were told to direct their attention to the computer screen and to hit "ENTER" to begin reading the instructions. The experiment was programmed using E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA) and was presented on a standard PC computer.

Study Phase. The study phase consisted of 108 related word pairs and 30 unrelated word pairs. Each of the 138 word pairs was presented one at a time at a rate of 4 seconds each in random order. The presentation time was selected based on prior research showing that longer presentation times produced ceiling effects in recognition memory (Payne & Eakin, 2017). To encourage participants to engage in deep processing of each pair, after each pair was presented, participants rated each for relatedness using a scale of 1 (not at all related) to 7 (completely related).

Distractor Phase. After the study phase, participants completed the Ospan task (Unsworth et al., 2005) for the distractor task. This task allowed for a filled interval to both prevent rehearsal of the pairs and to allow time to pass; both were necessary to prevent ceiling effects in recognition and to increase the variability of responses on the recognition confidence scale.

Recognition Phase. The recognition phase began with a practice phase. During this practice phase, the ten word pairs presented in the study practice phase were presented again. Participants practiced using the 6-point confidence recognition scale by rating on a scale of 1 (Sure New) to 6 (Sure Old) how confident they were in having seen the presented word pair during study. The 6-point scale remained at the top of the screen for each word pair, as shown in Figure 11, and rating was self-paced.

After the practice phase, participants were presented with the recognition test. Each of the word pairs were randomly presented and confidence ratings were made using the 6-point recognition confidence scale to indicate the degree to which participants thought the pair was old or new. Each word pair was presented one at a time in the bottom center of the screen with the 6-point recognition scale at the top; rating was self-paced. After rating all of the word pairs, participants were given a debriefing form and assigned course credit.

CHAPTER VI

BEHAVIORAL RESULTS

The design was a single-factor (Word Pair Type: studied, FTi, FTx, Re-Paired, R-control, U-control) within-subjects design. Old responses were operationalized as either a 4-, 5-, or 6- response, and New responses were operationalized as either a 1-, 2-, or 3-response on the 6-point recognition confidence scale. All responses were converted to a signal-detection framework: hits, misses, false alarms, and correct rejections. Hits were operationalized as studied word pairs that received either a 4-, 5-, or 6-response. Misses were operationalized as studied word pairs that received either a 1-, 2-, or 3-response. False alarms were operationalized as FTi, FTx, Re-Paired, R-control, or U-control test pairs that received either a 4-, 5-, or 6-response. Correct rejections were operationalized as FTi, FTx, Re-Paired, R-control, or U-control test pairs that received either a 1-, 2-, or 3-response. For FTi test pairs, there were 36 false alarms possible; for FTx test pairs, there were 36 false alarms possible; for Re-Paired test pairs, there were 30 false alarms possible; for R-control test pairs there were 20 false alarms possible; and for U-control test pairs there were 10 false alarms possible. Because counts were not equal across all word pair conditions, all recognition counts were converted to proportions for analysis; the number of hits was divided by the total number of hits possible and the number of false alarms was divided by the total number of false alarms possible for each word pair type.

A total of 64 participants participated in this study for course credit. Six participants were removed from the analysis due to incomplete data; two participants were removed due to a chance level of identifying new items as old or vice versa; and two participants were removed due to a low rate of recognizing studied items as old, identified by a boxplot analysis. Therefore, all behavioral analyses included a total of 54 participants. All alpha levels were set to .05.

To ensure that there were no differences in the dependent variables among the four counterbalanced list conditions, the hit rate and false alarm rate for R-control test pairs were compared across lists using a univariate analysis of variance (ANOVA). There were no significant effects in the hit rates among List 1, List 2, List 3, and List 4, $F(3, 50) = .464$, $p = .71$, $\eta_p^2 = .027$. Additionally, there were no significant effects in the false alarm rates among List 1, List 2, List 3, and List 4, $F(3, 50) = .149$, $p = .93$, $\eta_p^2 = .009$. Therefore, all further analyses were collapsed across the list of counterbalance conditions for the remainder of the results.

Overall, recognition for studied word pairs was high; the hit rate for the studied word pairs was .85 (*SD* = .11). The false alarm rate was divided into separate rates for FTi, FTx, Re-Paired, R-control, and U-control test pairs. Participants were most likely to falsely recognize Re-Paired test pairs (*M* = .34, *SD* = .13), followed by FTi test pairs (*M* = .33, *SD* = .17), FTx test pairs (*M* = .29, *SD* = .17), R-control test pairs (*M* = .17, *SD* = .13), and U-control test pairs (*M* = .07, *SD* = .10). Table 4 reports all means and standards errors for the conditions.

Table 4

Recognition Rates and Standard Error

Recognition	Mean	SE
Hit Rate	.85	.01
FTi FA Rate	.33	.02
FTx FA Rate	.29	.02
Re-Paired FA Rate	.34	.02
R-control FA Rate	.17	.02
U-control FA Rate	.07	.01

Aim 1: Hypothesis Testing for Recognition Performance

The first three hypotheses of this dissertation sought to inform the overall recognition aspect of Aim 1: to investigate whether the false memory effect obtained using the Payne-Eakin paradigm is due to the activation of an implicitly associated word during the study of an explicitly presented cue-target word pair. Therefore, the first set of behavioral hypotheses compared only old/new judgments, by categorizing 1-, 2-, and 3-responses as new and 4-, 5-, and 6-responses as old, using *a priori* planned comparison *t*-tests to test critical conditions for significant differences.

Hypothesis 1

Hypothesis 1 predicted that FTi test pairs would be falsely recognized at a higher rate than R-control test pairs. An *a priori* planned comparison *t*-test comparison was conducted between the proportion of false alarms for both word pair types. As shown in Figure 12, the paired samples *t*-test between the FTi false-alarm rate and the R-control false alarm rate was significant, $t(53) = 8.66$, $p < .001$, 95% CI [.13, .20], supporting our hypothesis. As predicted, participants were significantly more likely to falsely recognize FTi than R-control test pairs as old. Table 4 lists mean recognition rates and standard error for each word pair type. This finding suggests that activation of the implicitly associated false target during the explicit presentation of the studied word pair led to the false recognition of the FTi test pair.

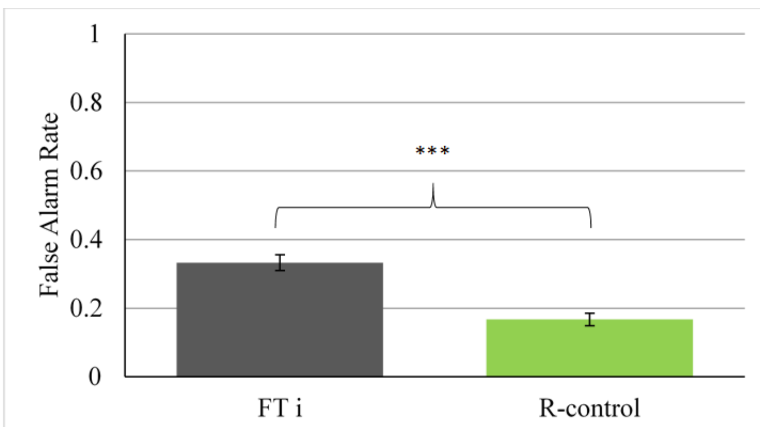


Figure 12. Hypothesis 1 results: comparison between the rate of false recognition for FTi test pairs and R-control test pairs.

Hypothesis 2

Hypothesis 2 predicted that FTx test pairs would be falsely recognized at the same rate as R-control test pairs. An *a priori* planned comparison *t*-test comparison was conducted between the proportion of false alarms for both word pair types. As Figure 13 shows, the paired samples *t*-test between the FTx false-alarm rate and the R-control false alarm rate was significant, $t(53) = 7.29$, $p < .001$, 95% CI [.09, .16]. Contrary to our prediction, participants were significantly more likely to falsely recognize FTx than R-control test pairs as old. This increased false alarm rate could be due to the repetition of the presented cue. This result was surprising given a previous pilot study that found FTx and R-control test pairs were falsely recognized at the same rate (Payne & Eakin, 2014), which led us to predict that FTx test pairs would be falsely recognized at the same rate as R-control test pairs due to the lack of implicit association between the studied word pair and the FTx target. It could be that this false recognition effect is due to familiarity with the cue. Because FTx test pairs re-present the same cue presented at study, it is likely that familiarity with that cue drives the level of familiarity associated with the FTx test pair, thus increasing the likelihood that they were falsely recognized.

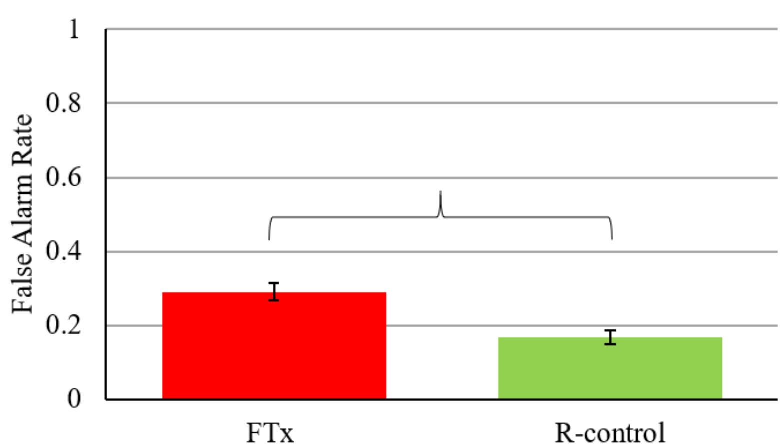


Figure 13. Hypothesis 2 results: comparison between the rate of false recognition for FTx test pairs and R-control test pairs.

It also could be that the re-presentation of the studied cue during test for FTx test pairs was responsible for the increased false-alarm rate observed for the FTi test pairs.

However, we anticipate that FTi test pairs are significantly more likely to be falsely recognized than FTx test pairs due to activation of the implicit FTi target during study. This hypothesis was tested in Hypothesis 3.

Hypothesis 3

Hypothesis 3 stated that FTi test pairs would be falsely recognized at a higher rate than FTx test pairs. An *a priori* planned comparison *t*-test comparison was conducted between the proportion of false alarms for both word pair types. As shown in Figure 14, the paired samples *t*-test between the FTx and FTi test pairs was significant, $t(53) = 2.59$, $p = .01$, 95% CI [.01, .07]. Participants were significantly more likely to falsely recognize the FTi test pairs than the FTx test pairs. This result supports the hypothesis that the implicit activation of the FTi target during study resulted in false recollection for those pairs. Although our results support the hypothesis that the increased false alarm rate for FTi test pairs is due to recollection, it is possible that a portion of the increased false alarm rate is due to familiarity with the re-presented cue. This possibility is further explored in the next set of analyses.

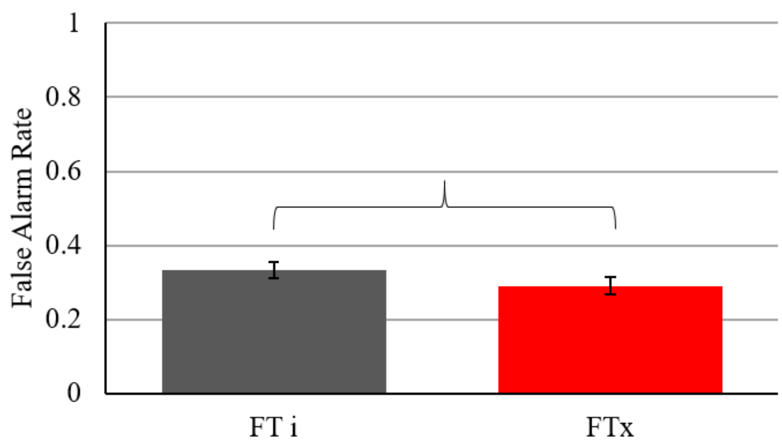


Figure 14. Hypothesis 3 results: comparison between the rate of false recognition for FTi test pairs and FTx test pairs.

Aim 1: Hypothesis Testing for High-Confidence Recognition Performance

The next two hypotheses sought to inform the recollection aspect of Aim 1, which was to investigate whether the false memory effect obtained using the Payne-Eakin paradigm was due to the activation of an implicitly associated word during the study of an explicitly presented cue-target word pair. One defining aspect of recollection is the presence of context; this definition is consistent across the CLS and DPSD model. According to the CLS model, when information that was present during encoding is presented during test, contextual information related to that studied event will be retrieved. Therefore, we predicted that implicitly activated associates present during encoding would result in the false recollection of that word pair when presented at test both due to its presence during study and its shared context with the studied cue-target word pairs.

According to Yonelinas (1994, 2002), a disproportional number of high-confidence old responses, or 6-responses, is indicative of an all-or-none retrieval process called recollection. Yonelinas et al. (2010) does state, though, that 6-response are not a process pure measure of recollection; high-confidence familiarity can also contribute to these responses. Instead, it is the disproportionate number of high-confidence response that is indicative of recollection. Although Yonelinas et al. (2010) states that 6-responses are not a pure process measure of recollection, for the purposes of the next set of behavioral analyses, we operationalized 6-responses to nonstudied pairs as false recollection and compared the critical word pair types using *a priori* planned comparison *t*-tests. In the discussion, we temper this language to be more appropriate of 6-responses can indicate high familiarity or recollection. The implications of these results and their underlying process are discussed further in the analysis.

Hypothesis 4

Hypothesis 4 predicted that FTi test pairs would be falsely recollected at a higher rate than R-control test pairs, as evidenced by more 6-responses for FTi than R-control test pairs. An *a priori* planned comparison *t*-test comparison was conducted between the proportion of false-alarm 6-responses for both word pair types. The paired samples *t*-test between the FTi 6-response rate and the R-control 6-response rate was significant, $t(53) = 9.82$, $p < .001$, 95% CI [.12, .18]. Participants were significantly more likely to be highly confident that they recognized FTi test pairs than R-control test pairs; a higher rate of recollection for those word pairs was obtained. Six-responses on the 6-point recognition confidence scale indicate a high degree of confidence that a word pair has been seen before. Because participants were significantly more likely to indicate a high degree of confidence in seeing FTi test pairs than R-control test pairs, suggests that the implicit activation of the false targets for FTi test pairs might have led them to falsely recollect having seen those word pairs during study. The false recollection of these word pairs is likely due to the retrieval of contextual information from the study event resulting from a match in the hippocampus (Norman & O'Reilly, 2003). Because the false target for FTi test pairs was implicitly activated during study, the presentation of the cue and implicitly activated target from study on the test resulted in the retrieval of contextual information associated with that event, resulting in false recollection of that FTi pair (Norman & O'Reilly, 2003). However, it also could be that the implicitly activated information during study could have contributed to an increased familiarity signal which would be consistent with the predictions of the DPSD model. This possibility is explored further in the modeling analysis. Table 5 lists mean recollection and standard error for each of the word pair types.

Table 5

Mean High-confidence Recognition Rates and Standard Error

Recollection	Mean	SE
Studied	.72	.19
FTi	.20	.02
FTx	.16	.02
Re-Paired	.19	.02
R-control	.06	.01
U-control	.02	.01

Hypothesis 5

Hypothesis 5 stated that FTx test pairs would be falsely recollected at the same rate as R-control test pairs. An *a priori* planned comparison *t*-test comparison was conducted between the proportion of false alarms for both word pair types. The paired samples *t*-test between the FTx 6-response rate and the R-control 6-response rate was significant, $t(53) = 7.20$, $p < .001$, 95% CI [.08, .14]. Contrary to our prediction, participants were significantly more likely to be highly confident that they recognized FTx test pairs than related control test pairs, suggesting a possible higher rate of recollection for FTx test pairs. However, it could be that because FTx test pairs involved the re-presentation of the studied cue, a higher degree of familiarity was evoked for those test pairs than R-control pairs, resulting in an increased number of 6-responses. According to Yonelinas et al. (2010), high-confidence old responses are not necessarily an indication of recollection. The DPSD model allows for high-confidence old responses of nonstudied items, but the model predicts that they should occur infrequently, if based on a normal distribution. What is evidence of recollection, as stated by Yonelinas (1994, 2002) is the disproportional rate at which high-confidence old responses occurs for studied items. Specifically, Yonelinas (1994, 2002) has demonstrated that a higher proportion of high-confidence responses are given to studied items than can be predicted by a standard signal-detection model, and this disproportionate rate is evidence of a second cognitive process. Therefore, it could be that the high rate of 6-responses to FTx test pairs is being driven by familiarity with the studied cue and not recollection. Theoretically, recollection is not likely to have occurred for these pairs because there should not be any shared context between the studied cue-target pair and the FTx test pair at test to result in recollection. The reason for this is that the test target for FTx pairs is not an associate of both the cue and target and, therefore, was not implicitly activated during study of the original cue-target pair. This lack of activation during study should result in the FTx pair being rejected by the hippocampus, not allowing for recollection. Therefore, 6-responses for FTx test pairs should be evidence of a high familiarity experience, not recollection.

One way to test this hypothesis is to compare the 6-response rate of FTx and FTi test pairs. Both FTx and FTi test pairs should evoke familiarity processes due to the re-presentation of the studied cue. However, we predict that FTi test pairs will have a higher probability of receiving high-confidence old responses due to the presence of an implicitly activated target during study, suggesting an influence from recollection. Table 6 demonstrates each word pair type and the retrieval information that each type presents during encoding. Because of this implicit activation, FTi test pairs should evoke recollection processes, resulting in shared context with the studied cue-target word pair. Because FTi test pairs include this one additional piece of information, they should have a higher likelihood of triggering recollection through pattern completion. Therefore, FTi test pairs should be more likely to receive a 6-response than FTx test pairs, due to recollection of the FTi pairs. This comparison was completed in Hypothesis 6.

Table 6

Word Pair Type and Information Present During Test

Word Pair Type	Studied Cue	Studied Target	Context	Implicit Associate
Studied	**Yes**	**Yes**	**Yes**	No
FTi	**Yes**	No	**Yes**	**Yes**
FTx	**Yes**	No	No	No
Re-Paired	**Yes**	**Yes**	No	No
R-control	No	No	No	No
U-control	No	No	No	No

Hypothesis 6

Hypothesis 6 stated that FTi test pairs would be falsely recollected at a higher rate than FTx test pairs. An *a priori* planned comparison *t*-test comparison was used between the proportion of false alarms for both test pair types. The paired samples *t*-test between the FTi 6-response rate and the FTx 6-response rate was significant, $t(53) = 3.06$, $p < .01$, 95% CI [.01, .07]. Consistent with our prediction, participants were significantly more likely to be highly confident that they recognized FTi test pairs than FTx test pairs. Because FTi test pairs consist of a re-presentation of the studied cue plus a false target, if the same cognitive process (familiarity) is occurring for these test pairs, then they should receive the same rate of 6-responses. However, this was not found to be the case; FTi test pairs were significantly more likely to receive 6-responses than FTx test pairs. This finding provides support that FTi pairs were retrieved via recollection.

Summary

The findings from the first six hypotheses support the prediction of the CLS model, that FTi test pairs would be falsely recognized at a higher rate than FTx or control test pairs. This prediction was due to the hypothesis that the activated implicit associate present at encoding was stored as part of the studied pair's representation in the hippocampus, resulting in a match between the stored and presented representations during test. This match then allowed for the retrieval of the studied item's representation that included context from the studied event. This retrieved information, then resulted in recollection for the FTi pairs. Therefore, these findings indicate that the false memory effect obtained for the FTi test pairs is due to false recollection. This supposition is supported by the finding that FTi test pairs were significantly more likely to receive 6-responses than FTx or R-control test pairs.

However, the results obtained for the first six hypotheses also reveal that the false memory effect obtained in the Payne-Eakin (2017) paradigm could be partly due to familiarity, as predicted by the DPSD model. According to the DPSD model, both FTi and FTx test pairs should receive more false alarms than R-control test pairs due to the re-presentation of the cue at test and the semantic similarity between the studied target and the false targets, which they did. However, the results from the first six hypotheses also provide data that are problematic for the DPSD model. One limitation of the DPSD theory (Yonelinas, 1994, 2002) is that it does not provide concrete explanations for how an item is determined as more or less familiar. However, based on previous research (Kim & Yassa, 2013; Mandler, 2008; Yonelinas, 2002), these predictions regarding the familiarity of FTi and FTx test pairs can be made. The first prediction is that FTi and FTx test pairs should have similar levels of familiarity because the false target for FTi and FTx test pairs are semantically similar (Kim & Yassa, 2013). The second prediction is that both FTi and FTx test pairs should have similar levels of familiarity due to the re-presentation of a studied cue during test (Mandler, 2008; Yonelinas, 2002). Therefore, according to the DPSD theory, both types of word pairs should have the same level of familiarity associated with them, which the DPSD model then would predict equal false recognition rates for both types of word pairs, which they did not. Instead, FTi test pairs were significantly more likely to be falsely recognized than FTx pairs.

Aim 2: Hypothesis Testing of Response Distributions

Another method to determine the degree to which the false memories obtained by the Payne-Eakin (2017) paradigm are solely due to familiarity or familiarity and false recollection is to examine the distribution of responses along the 6-point confidence scale. Two hypotheses were developed to inform Aim 2, which sought to examine whether the response distributions for the word pair types were different. According to Elfman et al. (2008), response patterns resulting from a recollection process should produce a different distribution shape than response patterns resulting from a familiarity process. Because recollection is characterized as an all-or-none cognitive process, it should lead to a disproportional increase in the number of high-confidence old responses, or 6-responses. Additionally, because recollection has been characterized to aid in the rejection of items during paired-associate learning (Norman & O’Reilly, 2003; Rotello & Heit, 2000), this rejection should also lead to a disproportional increase in the number of high-confidence new responses, or 1-responses. Therefore, items influenced by recollection should produce response distributions that have more responses in the tails than in the middle when compared to items that are influenced by familiarity. In contrast, because familiarity is described as a signal-detection process, responses resulting from this process should create a more Gaussian-shaped distribution (Elfman et al., 2008). Therefore, hypotheses 7 and 8 predicted that the shape of the response distribution among the different types of word pairs would vary according to the recognition process hypothesized to influence that word-pair type.

Hypothesis 7

Hypothesis 7 predicted that response distributions for FTi test pairs would be more dichotomous with a higher proportion of responses landing in the 1- and 6-response categories, as compared to R-control test pairs. The proportion of responses given to both FTi and R-control test pairs was plotted as a histogram (Figure 15) and an ROC curve (Figure 16). The histogram demonstrates that R-control test pairs appear to have received more 1-, 2-, 3-, and 4-responses than FTi test pairs; whereas FTi test pairs received more 5- and 6-responses. The finding that R-control test pairs possibly received more 1-, 2-, and 3-responses than FTi test pairs is not surprising given that R-control test pairs were not studied and were unrelated to any of the other words presented at either study or test. Although the number of middle responses is low for both groups, R-control test pairs appear to have received more middle responses than FTi test pairs.

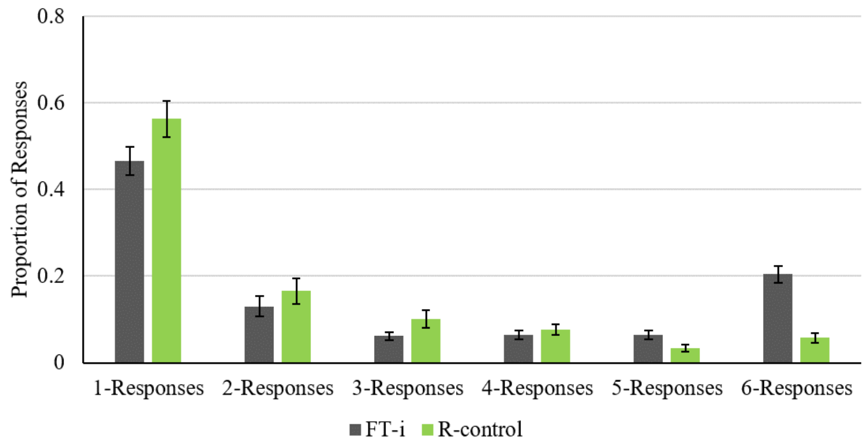


Figure 15. Histogram for Hypothesis 7: comparison of responses between FTi and R-control test pairs.

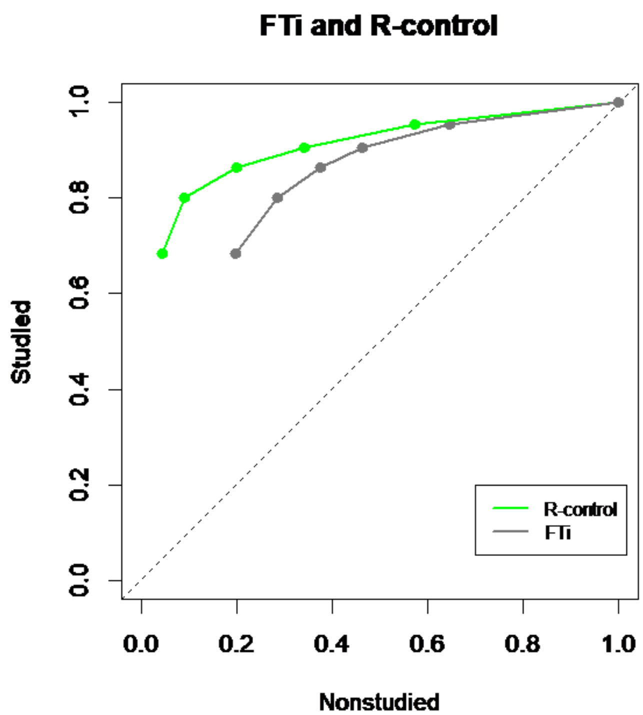


Figure 16. ROC curves for Hypothesis 7: comparison of responses for FTi and R-control test pairs.

We predicted that the x-intercept of the ROC curve for FTi test pairs would be shifted further to the right than the x-intercept of the ROC curve for R-control test pairs, demonstrating that those word pairs received more 6-responses. We also predicted that the ROC curve for FTi test pairs would be more linear (i.e., flatter curve) than the R-control test pairs, demonstrating the influence of recollection rejection. Figure 16 shows that the ROC for the FTi test pairs is shifted to the right, consistent with our hypothesis. However, inconsistent with our hypothesis, the ROC for the FTi test pairs is not more linear, demonstrating that FTi test pairs did not receive significantly fewer middle responses than R-control test pairs. This finding is likely due to the overall low rate of middle responses obtained across all word pair types, resulting in a floor effect. Although the finding that response distributions for FTi test pairs was not more dichotomous than R-control test pairs was inconsistent with our prediction, the ROC analysis does not provide further insight into whether familiarity alone or a combination of familiarity and recollection occurred during retrieval. The forthcoming modeling analysis will be used to answer this question.

Hypothesis 8

Hypothesis 8 predicted that response distributions for FTx and R-control test pairs should be comparable. Because FTx test pairs do not contain an implicitly activated target, according to the CLS model, those word pairs should not be recollected but, instead, should be rejected and recognized as new, similar to R-control test pairs. Therefore, FTx test pairs should be retrieved via the same familiarity process as R-control test pairs. Because familiarity produces a more Gaussian-shaped distribution than recollection, we predicted that response distributions for FTx and R-control test pairs would produce similar Gaussian shapes, with more responses falling into the left tail.

The proportion of responses given to both FTx and R-control test pairs was plotted as both a histogram (Figure 17) and a ROC curve (Figure 18). The histogram of responses to FTx and R-control test pairs demonstrates that R-control test pairs appear to have received more 1-, 2-, and 3-responses than FTx test pairs; whereas FTx test pairs appear to have received more 5- and 6-responses, demonstrating that FTx test pairs were more likely to be falsely recognized than R-control test pairs. As discussed in the results for Hypothesis 2, it is likely that FTx test pairs received more 6- responses than R-control test pairs due to the re-presentation of the studied cue. Although we did not predict that the repetition of the cue for FTx test pairs would influence 6-responses, these results in addition to the results from Hypothesis 2 and 5—and observation of the histogram distribution—suggest that it did.

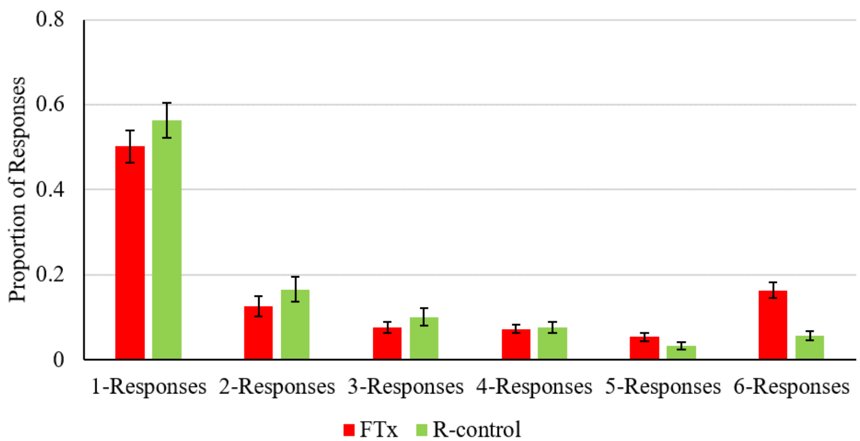


Figure 17. Histogram for Hypothesis 8: comparison of responses for FTx and R-control test pairs.

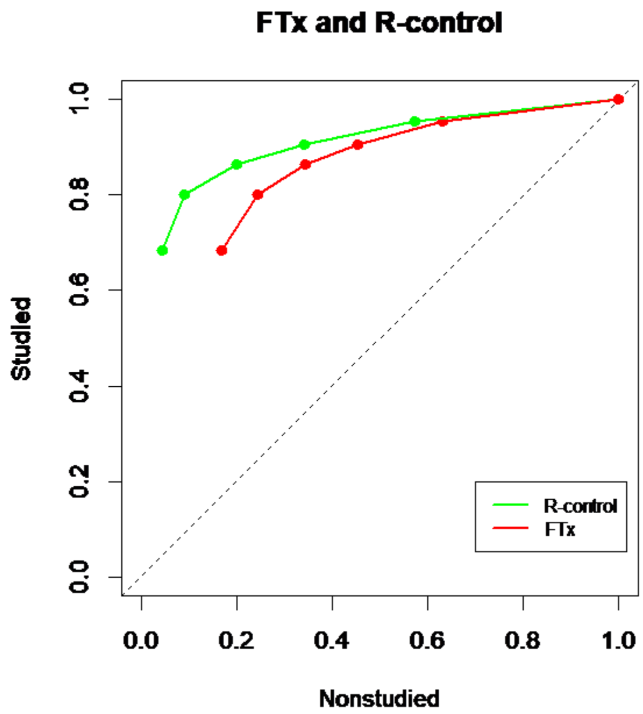


Figure 18. ROC curve for Hypothesis 8: comparison of responses for FTx and R-control test pairs.

We predicted that the x-intercept of the ROC curve for FTx test pairs would be equal to the x-intercept of the ROC curve for R-control test pairs. However, Figure 18 reveals that the x-intercept for FTx test pairs is shifted further right than R-control pairs, demonstrating more 6-responses for FTx than R-control test pairs. It should also be noted that the ROC curve for FTx test pairs is closer to the diagonal than the ROC curve for R-control pairs. This shift indicates reduced discriminability for FTx pairs; FTx pairs were more difficult to discriminate from studied words than R-control test pairs. This shift suggests an increased influence of familiarity and adds to the evidence that suggests the repetition of the studied cue for FTx test pairs produced a higher level of familiarity that impacted false recognition. According to the CLS model, the repetition of the cue likely created more activation in the MTLC, resulting in a higher level of familiarity for those word pairs. This finding is counter to our prediction that FTx test pairs and R-control test pairs would be retrieved via the same process which was based on the notion that FTx pairs would be rejected as new by the hippocampus. Finally, we also found that, consistent with our hypothesis, the ROC for the FTx test pairs was similarly curvilinear to the ROC curve for R-control test pairs. This finding was consistent with our prediction and suggests that FTx and R-control pairs were both retrieved via familiarity.

Follow-Up Distribution Analysis

In addition to Hypothesis 7 and Hypothesis 8, we also performed a follow-up analysis of the response distribution for FTx and FTi test pairs. Because we predicted that FTx test pairs would be retrieved via familiarity and FTi test pairs would be retrieved via recollection, we predicted that FTx test pairs would have more of a Gaussian-shaped response distribution and FTi test pairs would have a more dichotomous response distribution. Because both FTi and FTx test pairs produced a ROC curve that was more shifted to the right than R-control test pairs, we included a comparison of the ROC curves for FTx and FTi test pairs. If FTi test pairs were more likely to be recollected than FTx test pairs, then FTi test pairs—falsely recognized due to familiarity—should produce a more dichotomous distribution of responses, resulting in a more linear ROC shape and more 6-responses, producing a ROC curve with an increased x-intercept.

The proportion of responses given to both FTx and FTi test pairs was plotted as both a histogram (Figure 19) and a ROC curve (Figure 20). The histogram shows that the distribution of responses was similar for both FTx and FTi test pairs. Therefore, it could be that both FTx and FTi test pairs were retrieved via familiarity. However, as demonstrated by the findings for Hypothesis 3, FTi test pairs were significantly more likely to receive a high-confidence response, or 6-response, than FTx test pairs. This disproportionate rate of 6-responses for FTi test pairs compared to FTx test pairs suggest that either another process or factor was influencing the retrieval of those pairs. It could be that FTx test pairs and some of the FTi test pairs were retrieved via familiarity, with some additional FTi pairs being retrieved via recollection, but the distribution of responses is inconclusive. This explanation seems plausible given that recollection is theorized to only influence high-confidence responses, which would predict the significantly higher number of 6-responses for FTi pairs. However, it also could be that the implicit activation of the FTi target provided an additional influence on familiarity. This question will be further explored in the modeling analysis.

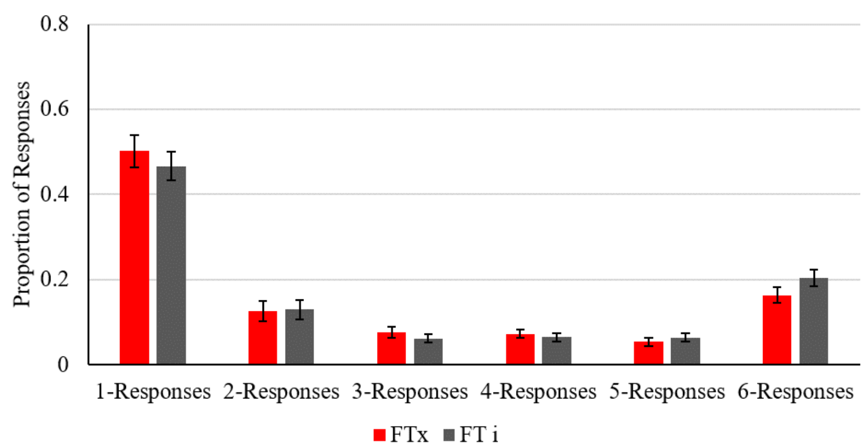


Figure 19. Histogram for follow-up analysis: comparison of responses for FTx and FTi test pairs.

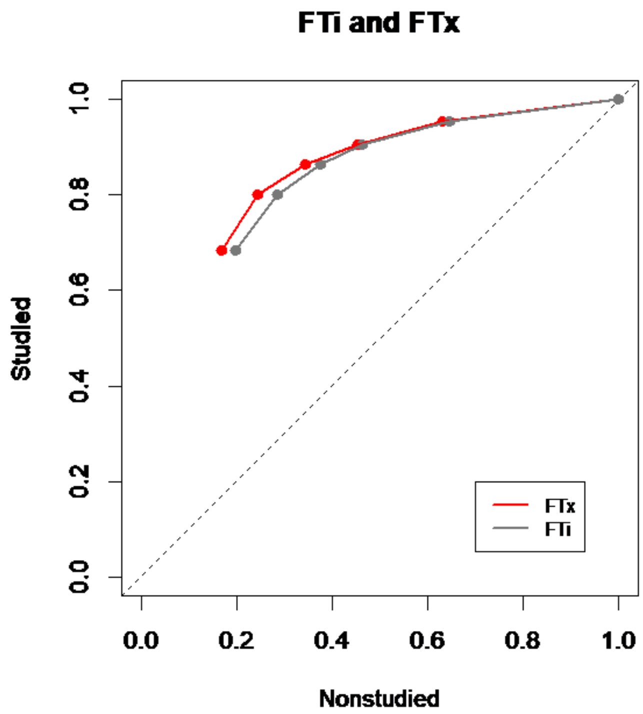


Figure 20. ROC curve for follow-up analysis: comparison of responses for FTi and FTx test pairs.

Summary

The findings from Hypotheses 7 and 8 as well as the follow-up analysis were inconclusive. Hypotheses 7 and 8 were designed to serve Aim 2, which was to examine whether the false memories obtained by the Payne-Eakin paradigm were due solely to familiarity or a combination of familiarity and recollection. According to the CLS model, we predicted that FTi test pairs would have a more dichotomous response distribution compared to that of R-control test pairs. This prediction was based on the hypothesis that the activated implicit associate present during study in the FTi test pairs was stored as part of the studied pair's representation in the hippocampus, resulting in a match between the stored pair's and presented pair's representations allowing for the additional contribution of recollection due to the retrieved context for those pairs. Although the results from Hypothesis 7 revealed that response distribution for FTi test pairs was somewhat dichotomous, people were not more likely to provide more middle responses for the R-control pairs. Instead, the main difference in the type of responses provided for the two types of word pair is that FTi test pairs were more likely to be falsely recognized than R-control pairs. Additionally, the results for Hypothesis 8, that FTx test pairs would receive a similar response distribution to the R-control pairs was not supported. People were more likely to falsely recognize FTx test pairs than R-control test pairs, resulting in an imbalanced response distribution. Although FTx test pairs were more likely to be falsely recognized, we attribute this increase to familiarity with the cue.

In a follow-up analysis, we compared the response distribution for FTi and FTx test pairs. We predicted that FTi test pairs would receive a more dichotomous distribution of responses than FTx test pairs due to a greater influence of recollection. Based on the CLS model, we predicted that FTi test pairs would have a greater influence on recollection than familiarity, resulting in a dichotomous distribution, and that FTx test pairs would have a more Gaussian shaped distribution of responses due to the influence of familiarity. The follow up comparison did not support this prediction. We found no difference in the distribution of responses for FTi and FTx test pairs. FTi test pairs did receive more high-confidence old responses than FTx test pairs, resulting in a greater number of responses in the right tail; however, this increase did not produce a large enough difference to separate the rest of the two distributions. Therefore, it is unclear if our hypothesis that FTi test pairs were retrieved via recollection and FTx test pairs were retrieved via familiarity is correct.

Although our findings did not support our predictions, they do provide some insight into the retrieval processes that were used for FTi and FTx test pairs. It could be that our original predictions, based on the CLS model, were incorrect. The biggest weakness of the CLS model is its lack of explanation as to how the contribution of the hippocampal and MTLC model work together to make recognition judgments. Based on theoretical assumptions, we assumed that because FTx test pairs do not contain an implicit associate, they would be rejected as new by the hippocampus, resulting in only some activation from the MTLC. This prediction was also supported by a previous study by Payne and Eakin (2014) that found FTx test pairs were falsely recognized at the same rate as R-control test pairs. However, the results of the current study indicate that FTx test pairs had a greater impact on familiarity than expected. Additionally, the results pertaining to the FTi test pairs suggest that those word pairs are being retrieved from recollection. The implications of these findings will be discussed in the forthcoming discussion section.

Aim 3: Hypothesis Testing of Re-Paired Test pairs

The next set of hypotheses sought to inform Aim 3, which was to determine whether increasing familiarity by re-presenting a studied cue and target at test—although not in their original pairing—affected false recognition. As shown in the results for Hypotheses 2 and 3, participants were significantly more likely to falsely recognize FTx over control test pairs. In addition, false recognition was more frequent for FTi than FTx test pairs. Although the FTi and FTx pairs have in common that they have a repeated studied cue, as shown in Table 3, producing similar levels of familiarity, apparently another process other than familiarity with the cue contributed to the increased false recognition of FTi test pairs. We hypothesized that this process was recollection, resulting from the implicit activation of the presented target during study of the original cue-target pair and the shared context with the studied pair. Although we anticipated this finding—we expected FTx pairs to be similar to controls, based on previous studies—we included a third type of word pair to further test the potentially differential impact of familiarity on recognition by presenting not only familiar cues, but also familiar targets.

The Re-Paired test pair consists of a cue and target that were both studied but not together as a pair. Each Re-Paired test pair was created by shuffling studied cue-target pairs to create a new set of test pairs. Because the studied word pairs were semantically associated, recombining them into new pairs created a set of unrelated test pairs, providing a simple heuristic for people to easily identify these word pairs as new. Therefore, we added a set of unrelated cue-target pairs during study to serve as the Re-Paired test pairs. The purpose of this recombination was to test familiarity. Because both the cue and target were studied—albeit not together—their overall level of familiarity should be higher than that for FTi or FTx pairs, whose target was never studied. However, these words should not be recollected for several theoretical reasons. First, according to the PIER2 model, because the cues and targets used for this condition are not associated, no target should ever be implicitly activated during study. Second, according to the CLS model, these Re-Paired pairs should not provide a close enough match to the original studied pair to allow for pattern completion which would result in the retrieval of contextual information necessary for recollection. Therefore, identification of the Re-Paired test pairs should not be influenced by recollection. Instead, the identification of the Re-Paired test pairs should only be influenced by familiarity. Because of this feature, these pairs provide a comparison condition for the impact of high familiarity in the absence of any influence of recollection. As comparison, a new set of controls word pairs consisting of unstudied new unassociated cue-target word pairs were created to serve as unrelated control—U-control—pairs. These pairs served as controls for the Re-Paired test pairs for the next two hypotheses.

Hypothesis 9

Hypothesis 9 stated that Re-Paired test pairs would be falsely recognized at a higher rate than U-control test pairs. An *a priori* planned comparison *t*-test comparison was used between the proportion of false alarms for both word pair types. As Figure 21 shows, the paired samples *t*-test between the false-alarm rate for Re-Paired and U-control test pairs was significant, $t(53) = 14.46$, $p < .001$, 95% CI [.23, .31]. Consistent with our prediction, participants were significantly more likely to falsely recognize Re-Paired test pairs than U-control test pairs. This finding is not surprising given that both the cue and target for the Re-Paired test pairs were studied; therefore, the level of familiarity for those word pairs should be higher.

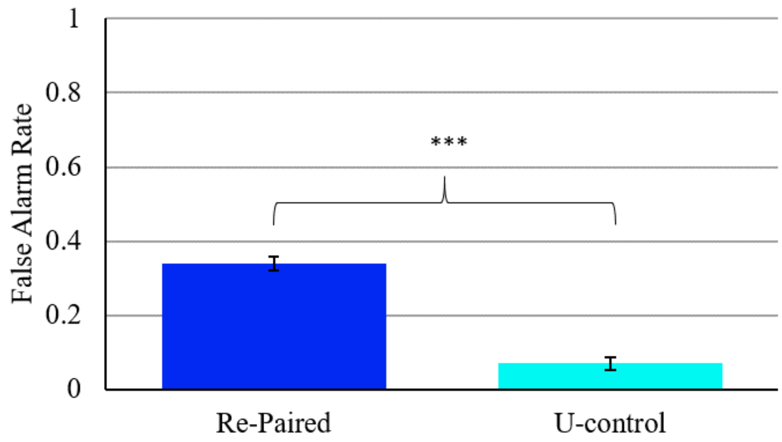


Figure 21. Hypothesis 9 results: comparison between the rate of false recognition for Re-Paired test pairs and U-control test pairs.

Hypothesis 10

Hypothesis 10 predicted that response distributions for Re-Paired test pairs would be more Gaussian than U-control test pairs, with more responses falling into the middle responses. The proportion of responses given to both Re-Paired and U-control test pairs was plotted as both a histogram (Figure 22) and a ROC curve (Figure 23). The histogram of responses for Re-Paired and U-control test pairs shows that, overall, responses for Re-Paired test pairs appear to be more distributed with more responses falling into the 4-, 5-, and 6-response category than U-control test pairs. U-control test pairs appear to have more responses falling into the 1- and 2-response category. The finding that U-control test pairs received for 1- and 2-responses than Re-Paired controls pairs is not surprising given that U-control test pairs were not studied. It should be noted, though, that Re-Paired test pairs received significantly more 6-responses (seen in Figure 22 and Figure 23), suggesting an influence of high familiarity.

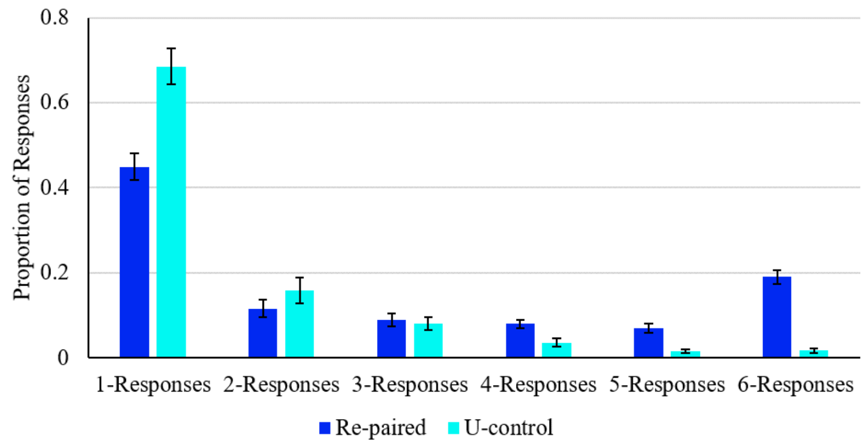


Figure 22. Histogram for Hypothesis 10: comparison of responses between Re-Paired and U-control test pairs.

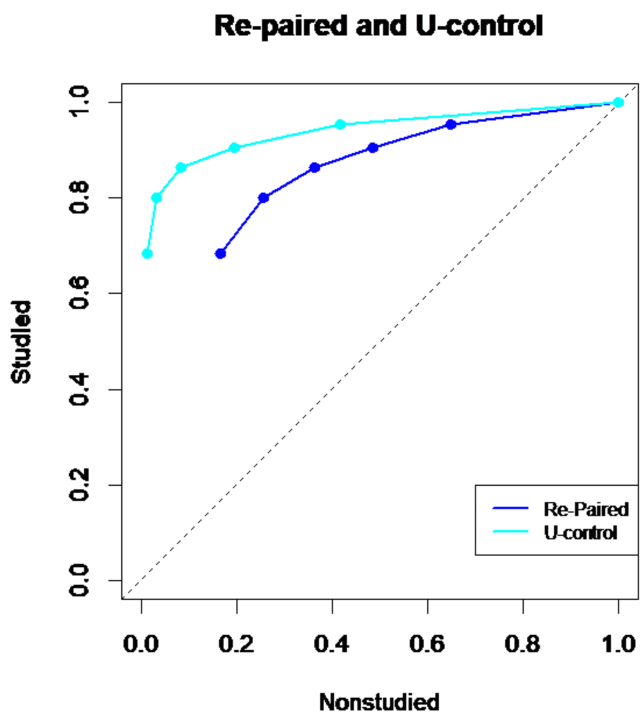


Figure 23. ROC curve for Hypothesis 10: comparison of responses between Re-Paired and U-control test pairs.

Follow-Up Analysis

To further investigate the nature of the Re-Paired test pairs, we performed a follow-up analysis. We predicted that Re-Paired test pairs would have a more Gaussian-shaped response distribution than FTi test pairs due to a greater influence of familiarity (Elfman et al., 2008). If Re-Paired test pairs were being impacted by familiarity, and not recollection, then the Re-Paired test pairs should receive more middle-responses than FTi test pairs. We also included FTx test pairs which were predicted to have a more Gaussian-shaped response distribution. If the impact of familiarity is greater for Re-Paired test pairs than FTx test pairs, then we would expect that the magnitude of the number of middle-responses would be greater for Re-Paired test pairs than FTx. Each of these predictions were based on the CLS model. According to the CLS model, a familiarity signal for both the studied cue and target in the Re-Paired condition should result in an overall stronger familiarity signal. Although the Re-Paired test pairs were not studied together, both the cue and target were studied one time. Because the cue and target will have only been studied once, the process of sharpening (i.e., the process of refining and strengthening the familiarity signal for items studied together) should not have any influence on the pair. Therefore, we predicted that the influence of familiarity should be greatest for Re-Paired test pairs. Figure 24 demonstrates the comparison of the response distributions for Re-Paired, FTi, and FTx test pairs.

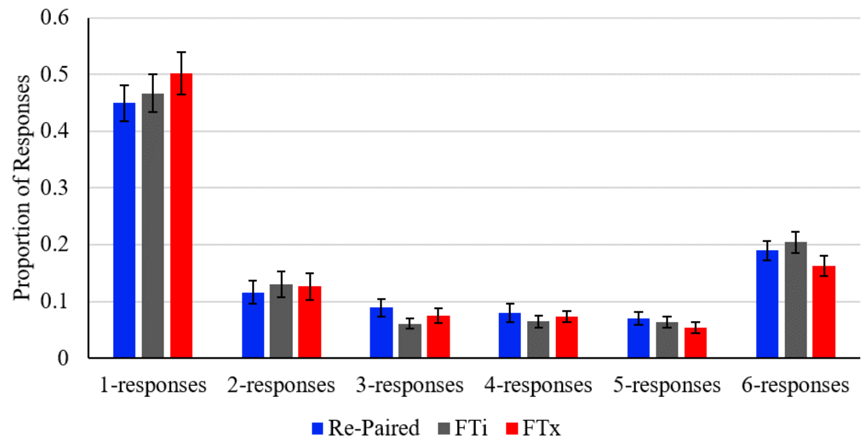


Figure 24. Comparison of responses among Re-Paired, FTi, and FTx test pairs.

Contrary to our predictions, the response distribution for Re-Paired, FTi, and FTx appear to be similar. Although the histogram revealed no significant differences between the response distributions, it appears that the number of 6-responses given to each of the three types of test pairs varied. Therefore, we conducted a one-way repeated measures ANOVA on the proportion of 6-responses given to each word pair type. From the ANOVA, we found that there were significant differences in the number of 6-responses given, $F(1,53) = 9.33$, $p < .05$ $\eta_p^2 = .07$; FTi and Re-Paired test pairs received the most, followed by FTx test pairs. Figure 25 shows the relationship between the three types of test pairs. Interestingly, this result demonstrates that people provided an equally high number of high-confidence 6-responses to FTi and Re-Paired test pairs rather than to just Re-Paired test pairs. The purpose of the Re-Paired test pairs was to push the influence of familiarity. Because both the cue and target in the Re-Paired condition had been previously studied as part of a different pair, those word pairs should have the highest level of familiarity associated with them. According to the CLS model, the familiarity signal for Re-Paired pairs should be higher than the familiarity signal for FTx pairs due to both the studied cue and target being recently studied. However, we predicted that 6-responses for FTi pairs would be higher than FTx pairs due to recollection. Therefore, the finding that FTi and Re-Paired test pairs received an equal number of 6-responses is not surprising. However, because the DPSD model would not predict any contribution of recollection for FTi pairs, the model would attribute the increased number of 6-responses to familiarity. Therefore, the results of the follow-up analysis do not provide any discriminating evidence for the two processes.

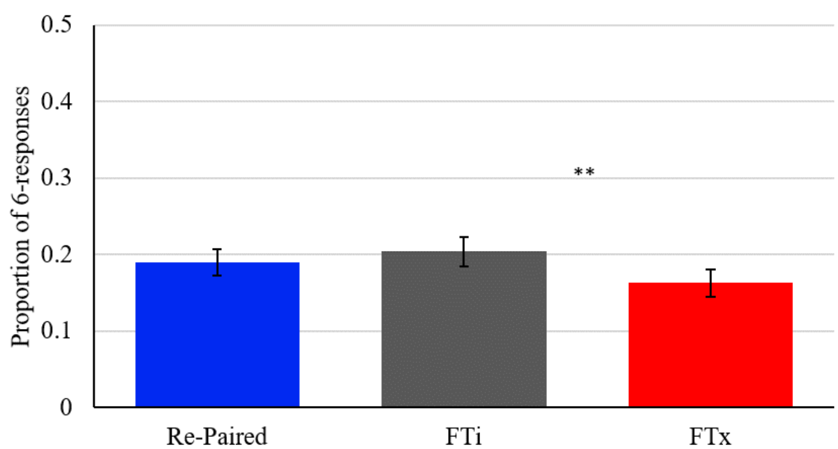


Figure 25. Comparison of 6-responses among Re-Paired, FTi, and FTx test pairs.

Discussion

Each of the word pair types in the dissertation experiment was designed to manipulate one of the two recognition retrieval processes, familiarity or recollection. Table 7 presents a list of each word pair type and the recognition process it was designed to measure. The word pair types of interest were the FTi, FTx, and Re-Paired test pairs.

FTi test pairs were designed to measure recollection by presenting a false target that was implicitly activated during study along with a studied cue. The purpose of presenting a studied cue re-paired with an implicit associate was to determine whether this implicit activation during study led to false recollection at test. We predicted, based on the CLS model, that because the FTi test pairs included a studied cue and an implicitly associated false target, the representation of those pairs presented during test would allow for a match with the studied item's stored representation, allowing for recollection to occur. Therefore, we theorized that recognition of FTi test pairs would be influenced by recollection.

Table 7

Word Pair Type and Its Corresponding Recognition Process.

Word Pair Type:	Process:
Studied	Familiarity + Recollection
FTi	Recollection
FTx	Familiarity
Re-Paired	Familiarity
R-control	Familiarity + Recollection
U-control	Familiarity + Recollection

FTx test pairs were designed to measure familiarity and also served as a comparison for FTi test pairs. The only difference between FTx and FTi test pairs was that FTi pairs were chosen to ensure implicit activation of associated false target whereas

FTx pairs were not. Therefore, based on predictions from the CLS model, when the FTx pairs were presented during test, a mismatch between the FTx pair's representation and the studied item's stored representation should have create a mismatch, resulting in the pair being rejected as new by the hippocampus. However, recognition of the FTx test pairs could have been based on familiarity. The cue for the FTx test pairs was a re-presentation of a studied cue with a new associated target. Recognition of FTx test pairs could be based activation in the MTLC due to repetition of this cue. Therefore, we predicted that any recognition of FTx test pairs would be based on familiarity.

Finally, Re-Paired test pairs were designed to test the limits of familiarity. Re-Paired test pairs included the re-presentation of both a studied cue and target that were not studied together. It was predicted, based on the CLS model, that both the studied cue and target would elicit a stronger familiarity signal from the MTLC than the FTx, FTi, or control test pairs. Because the Re-Paired test pairs repeated both a studied cue and target, they should have been more familiar than either the FTi or FTx test pairs for which only the cue was repeated. Additionally, recognition of these pairs was not thought to be based on recollection. Because the cue and target in the Re-Paired pairs were not studied together, according to the CLS model, they should be rejected by the hippocampus resulting in a failure of recollection.

Aim 1

The results showed that recognition was high, which was expected and is typical in the Payne-Eakin paradigm (e.g., Payne & Eakin, 2017, Pilot Experiments 1 & 2). Despite having 108 word pairs to study, people correctly identified studied word pairs as old 85% of the time. To avoid ceiling effects, we increased the number of studied items as well as the amount of time to complete the distractor task, O-Span (Unsworth et al., 2005) used in a previous pilot study. As a result of these changes, we were able to bring the hit rate down to 85% from 88%. Although the magnitude of the hit rate was not drastically reduced, it was lowered. It should be noted, though, that recognition memory is notoriously accurate (Heathcote et al., 2006; Pratte & Rouder, 2011; Yonelinas, 1994) and, as a result, a high hit rate is difficult to avoid without including a large number of stimuli. The typical technique is to use 250 or more stimuli to avoid ceiling effects. Unfortunately, due to the manipulation of the associative relationship between word pairs, finding 250 or more stimuli that met the criteria was not possible.

The critical comparisons related to the hypotheses delineated in Aim 1 of this dissertation were among the proportion of false alarms across the five test pair conditions. Hypothesis 1 predicted that FTi test pairs would be falsely recognized at a significantly higher rate than R-control test pairs. Because the target presented in those pairs was implicitly activated during study of the original pair, this target theoretically was part of the representation of the studied pair. Therefore, the CLS model would predict that when these pairs were presented explicitly with the studied cue at test, this prior implicit activation should generate enough overlap between the stored representation of the studied event and the FTi test pair representation to allow for pattern completion. If pattern completion occurred, then contextual information associated with the studied event would be retrieved and recollection would occur, resulting in a higher proportion of high-confidence old responses for FTi test pairs. This hypothesis was supported in that participants were significantly more likely to say old to and to give more 6-responses to FTi than to R-control test pairs (i.e., Hypotheses 1 and 4).

Hypothesis 2 predicted that FTx test pairs would be falsely recognized at the same rate as R-control test pairs. This hypothesis was based on a prior finding that false recognition was similar between FTx and R-control test pairs (Payne & Eakin, 2014) as well as predictions from the CLS model. Because the false target used for FTx test pairs was not activated within the context of the studied word pair, it should be rejected by the hippocampus as new and not result in recollection. This hypothesis was not supported; FTx test pairs were falsely recognized at a significantly higher rate than control test pairs. It could be that the re-presentation of the studied cue influenced familiarity in a way that we did not predict. Although recollection, processed in the hippocampus, should fail, according to the CLS model, the re-presentation of the studied cue as part of the FTx test pair could have increased activation in the MTLC. This increased activation could have produced a higher level of familiarity for those test pairs over R-control test pairs, resulting in an increased likelihood of being falsely recognized.

We did not originally predict this finding due to one of the limitations in the CLS model discussed in Chapter II. In the CLS model's (Norman & O'Reilly, 2003) current state, the relationship between the hippocampus and MTLC is not explicitly stated. Specifically, the model does not state how the hippocampal signal and MTLC familiarity signal are used to produce overall recognition judgments (e.g., does the MTLC takeover if the hippocampus rejects? Is a combination of both the hippocampal and MTLC results used?). Therefore, it is difficult to predict what role the MTLC will have during recognition if an item is rejected by the hippocampus. As a result of this, we originally theorized that FTx test pairs would be rejected due to a mismatch in the hippocampus, similarly to the R-control test pairs. However, it appears from the result for Hypothesis 2 that, despite rejection by the hippocampus, familiarity from the repeated cue resulted in initiation of the MTLC leading to false recognition due to familiarity. This finding requires refinement of the assumptions of the CLS model and warrants further exploration.

Hypothesis 3 predicted that FTi test pairs would be falsely recognized at a higher rate than FTx test pairs due to the implicit activation of the FTi false target during study. This hypothesis was supported in that participants were significantly more likely to give an old response to FTi than to FTx test pairs. According to the PIER2 model, the implicit activation of the false target during encoding resulted in the target being incorporated into the representation of the studied cue-target pair (Nelson et al., 1998). The effect of the implicit activation of this false target in producing false recollection is explained by the CLS model, which predicted that the FTi test pairs would be falsely recalled by the hippocampal model due to a match between the presented and stored representations (Norman & O'Reilly, 2003). This match would then result in the retrieval of contextual information associated with the studied event—which is necessary for recollection—and, because the test target was implicitly activated during study, the context would be associated with that target, resulting in a high-confidence recollection response. Therefore, the results from Hypothesis 3 were predicted and can be explained by the CLS model. This finding is, however, problematic for the DPSD model (Yonelinas, 2002).

According to the DPSD theory, any recognition of new test items must be due to familiarity; the DPSD model does not allow for recollection of new test items. Therefore, the DPSD would not attribute the higher false recognition rate of the FTi over the FTx test pairs to recollection, only to familiarity. Both the FTx and FTi test pairs included a studied cue and a semantically similar false target. Therefore, both types of test pairs should have elicited a similar level of familiarity (Kim & Yassa, 2013; Mandler, 2008; Yonelinas, 2002). The results from Hypothesis 3 require that FTi test pairs were more familiar than FTx test pairs to be explainable by the DSPD model. Alternatively, the DPSD model could allow for FTi and FTx test pairs to have different thresholds for old responses. Specifically, participants would have had to have a more liberal threshold for saying old to FTi pairs than FTx. However, the possibility of these two types of test pairs having different old thresholds within the same experiment is unlikely. Specifically, having different thresholds would require participants to be able to consciously differentiate between FTi and FTx test pairs. The likelihood of this occurring is minimal given that the only difference between the two types of word pairs is the implicit activation of the FTi false target; this information should not be consciously available. Therefore, the only theoretical explanation for the results is that recollection is responsible for the increased false recognition rate of FTi test pairs as predicted by the CLS model and that familiarity is responsible for the false recognition rate of FTx pairs as predicted by both the CLS and DPSD models. This prediction was further investigated in Hypotheses 4 through 6.

According to the DPSD model, a high-confidence old response can be indicative of either a high familiarity signal or recollection; however, a disproportionate number of high-confidence old responses is indicative of recollection (Yonelinas et al., 2010). Hypothesis 4 predicted that FTi test pairs would be falsely recollected at a significantly higher rate than R-control test pairs as indicated by more 6-responses for FTi than R-control pairs. We predicted that the explicit presentation of the implicitly activated false target along with the studied cue would lead to a match between study representations in the hippocampus and test representations, allowing for pattern completion. As a result of the pattern completion, contextual information associated with the studied event would then be retrieved—which is required for recollection (DPSD model, Yonelinas, 1994, 2002)—and would lead people to experience a false recollection for those pairs. The result would be significantly more high-confidence old responses for FTi than R-control test pairs, which was obtained. However, an alternative explanation could be possible. Both the CLS and DPSD models would have predicted that, because the FTi test pairs included a studied cue, these pairs would have a higher degree of familiarity associated with them than the R-control pairs which was comprised of new cues and new targets. We originally did not predict that familiarity could also play a role in the false recognition of FTi test pairs because of the vague explanation by the CLS model as to how familiarity and recollection work together. It could be that an increased level of familiarity due to the re-presentation of a studied cue could have contributed to the increased number of high-confidence old responses for some or all of the FTi test pairs. This possibility and its implications were further explored in the follow-up analysis.

Hypothesis 5 predicted that FTx test pairs would be falsely recollected at the same rate as R-control test pairs. Because the target for FTx test pairs was not implicitly activated during study, those pairs should not result in a match in the hippocampus when presented during test, resulting in a failure to retrieve the context necessary for recollection. This lack of context should have resulted in the rejection of the FTx test pair as old. This hypothesis was not supported in that participants were significantly more likely to give FTx test pairs a 6-response than R-control test pairs. Originally, it was predicted that FTx test pairs would be falsely recollected at the same rate as control test pairs due to predictions based on the CLS model and a prior study that showed an equal false alarm rate for FTx and control test pairs. However, the results suggest that FTx word pairs might have been influenced by familiarity instead, as predicted by the DPSD model. According to the DPSD model, FTx pairs should have more familiarity associated with them due to the use of a studied cue as the cue in the FTx test pair. This increased familiarity should result in a greater likelihood of false recognition for pairs with a repeated cue than those without a repetition of either the cue or target, which it did. Hypothesis 6 was designed to investigate whether the increased likelihood of receiving a high-confidence old response for FTi word pairs was also due to familiarity.

Hypothesis 6 predicted that FTi word pairs would be falsely recollected at a higher rate than FTx word pairs. This prediction was based on the CLS model's (Norman & O'Reilly, 2003) predictions that the implicit activation of the false-target for FTi word pairs would facilitate false recollection due to a match in the hippocampus for FTi, but not FTx, test pairs. This hypothesis was supported in that participants were significantly more likely to give FTi test pairs a high-confidence old response than FTx test pairs. This finding could be problematic for the DPSD model. According to the DPSD model, FTi and FTx test pairs should only be recognized due to familiarity. Because the FTi and FTx test pairs both include a studied cue as part of their pairs as well as a semantically similar false target, the DPSD theory would not predict that one test pair would produce a higher familiarity signal than the other. According to the CLS model, the only reason FTi test pairs should receive more high-confidence old responses is because of the influence of recollection, which the DPSD model does not allow for.

Although the CLS model suggests that the increased number of high-confidence old responses for FTi pairs is due to recollection rather than familiarity, it is difficult to disentangle the influence of the two processes based on standard behavioral analyses. According to the DPSD model, a high-confidence old response can be due to both high familiarity and recollection. Conversely, it could be that FTi test pairs received more high-confidence responses than FTx pairs due to the implicit activation of the FTi target. It could be that the activation of the implicit target during study contributed to activation in the cortical areas, resulting in a stronger familiarity signal. Therefore, disentangling what proportion of high-confidence responses is due to familiarity and what proportion is due to recollection is difficult, even with theoretically based test-pair manipulations. Elfman et al. (2008) suggested that looking at the overall distribution of responses should provide some insight into which process is responsible for the results produced. Therefore, hypotheses 7 through 10 were included to examine the overall response distribution for FTi and FTx test pairs.

Aim 2

According to Elfman et al. (2008), response distributions resulting from familiarity should produce a Gaussian-shaped distribution whereas recollection should produce a dichotomous distribution. Therefore, Aim 2 was included to examine whether the false memories obtained by the Payne-Eakin paradigm were due to familiarity or recollection by examining the pattern of given responses. Unfortunately, none of the predictions about the response distributions or ROC curve shapes were supported. Although the previous analyses suggested that the responses for FTi test pairs would be significantly less distributed than those for R-control test pairs, their overall response distributions did not appear different. Therefore, Hypothesis 7 was not supported. This pattern of results was obtained for Hypothesis 8 as well; the response distribution for FTx test pairs was not significantly different from R-control test pairs, indicating the same cognitive process was responsible for their recognition. Additionally, the shape of the ROC curves was similar for the two test-pair types, further supporting that the same retrieval process was used. FTx test pairs did receive significantly more 5- and 6-responses than R-control test pairs; however, this finding could have been due to familiarity with the test cue, as predicted by the DPSD model. The increase in high-familiarity responses observed for the FTx pairs also resulted in a higher x-intercept on the ROC curve.

An additional analysis of the response distribution for both FTi and FTx test pairs was done to determine whether the response distribution for FTi test pairs was more dichotomous than FTx word pairs, as would be expected if there was a differential influence of recollection between the pairs. However, the response distributions for FTi and FTx test pairs were not different from each other, suggesting that the same retrieval process(es) was responsible for falsely recognizing both FTx and FTi test pairs. This supposition was further supported by the ROC analysis which revealed that FTi test pairs were less discriminable than FTx test pairs, evidenced by a reduced *d'*, which produced a curve closer to the diagonal. These findings are consistent with the DPSD model's (Yonelinas, 1994, 2002) predictions and provides support for the assumption that only one retrieval process occurs for nonstudied items. Although these findings provide support for the DPDS model (Yonelinas, 1994, 2002), they are inconsistent with the predictions of the CLS model.

According to the CLS model, FTi test pairs should have had a higher probability of being falsely recognized than FTx test pairs due to recollection. This prediction was supported by the results for Hypotheses 3 and 6 finding higher false recognition and more 6-responses for FTi than FTx test pairs. Alternatively, it could be that FTi test pairs were influenced by both recollection and familiarity. Because the CLS model is vague about whether familiarity and recollection work together in sequence or parallel to produce overall recognition judgments, it was hypothesized that a match between the studied and test representations would occur in the hippocampus for FTi pairs, resulting in recollection, and that FTx pairs would be rejected due to a mismatch, resulting in rejected FTx pairs being identified as new. However, as suggested by the results for Hypotheses 2 and 5, familiarity could have played a role. False recognition of the FTx test pairs was due to familiarity because the re-presentation of a studied cue resulted in familiarity increasing the false-alarm rate for those test pairs over Control test pairs. However, a studied cue was also re-presented as part of the FTi test pairs. Therefore, it could be that recollection was not the only process that occurred for FTi pairs. Although the CLS model does not specify this possibility, some of the FTi pairs could have been rejected by the hippocampus due to a mismatch between the presented representation and the stored representation, resulting in the false recognition of the rejected pairs due to familiarity, similar to the FTx pairs.

This possibility is supported by assumption of the DPSD model that recollection and familiarity occur in parallel. If recollection returns contextual information related to a studied event, then the item is recognized as old without any consideration of the familiarity signal (Yonelinas et al., 2010). However, if recollection does not occur, then recognition is based solely on familiarity, which always occurs. Therefore, if some of the FTi pairs were rejected by the hippocampus due to a mismatch, then FTi pairs might have been comparable to FTx pairs and processed according to the principles of the MTLC model of the CLS model. Although the CLS model does not specify this two-step process between hippocampal and MTLC processing, it seems reasonable. The results from the follow-up analysis comparing FTi to FTx response distributions show that the distributions and ROCs are similar between FTi and FTx test pairs, indicating that they were based on the same cognitive process. However, the results from Hypothesis 3 showed that FTi test pairs had a higher rate of false recognition and Hypothesis 6 showed that more of these were high-confidence 6-responses. These findings show that the underlying cognitive processes might not be the same. Taken together, these findings could be explained by the assumption that some FTi pairs were rejected by the hippocampus and were falsely recognized due to familiarity, similar to FTx pairs, whereas other FTi pairs produced a match in the hippocampus, resulting in retrieval of the study context, and were recognized due to false recollection.

Although the results for Hypotheses 3 and 6 provide support for the hypothesis that both familiarity and recollection influenced the recognition of FTi test pairs, the results from the response distribution analysis suggested that the same recognition processes were used for both FTi and FTx test pairs. One possible reason that the results of the distribution analysis were not consistent with the results for Hypotheses 3 and 6 is that fewer FTi test pairs were recognized due to recollection which could have resulted in too few data points to detect its presence. This possibility is supported by the fact that the false-alarm and false-recollection difference for FTx and FTi pairs was .04. Because there were 36 FTi test pairs, a difference of 4% would mean that, on average, only one test pair would have had a response due to recollection. This small difference likely resulted in a weak influence from recollection which could not be detected in the response distributions and is consistent with Yonelinas et al.'s (2010) argument that false recollection occurrence is too infrequent to model. Overall, this result suggests that 36 FTi and FTx test pairs did not provide enough data points to detect differences in the overall response distributions. Typically, recognition studies use a minimum of 250 stimuli to investigate the influence of familiarity and recollection (e.g., Heathcote, 2003; Pratte, Rouder, & Morey, 2010; Yonelinas, 1994). However, given the test-pair characteristic constraints of the Payne-Eakin paradigm, creating 250 or more test pairs was not possible. In addition to the limited number of test pairs, the lack of difference in the response distributions also could be due to problems with using the 6-point confidence recognition scale.

Recognition is typically characterized as producing a dichotomous result (i.e., old or new). However, the 6-point scale requires participants to not only provide a dichotomous response but to also break down their response into degrees of confidence. Requiring participants to simultaneously perform two tasks might have made the task too difficult or confusing, resulting in participants relying only on their initial dichotomous response. This hypothesis is supported in that 69% of participant responses were dichotomous (either a 1- or 6-response), demonstrating that participants were not using the scale in the manner that it was intended. In addition, the current format of the 6-response scale could be confusing. The current format of the 6-point scale breaks down old/new responses into degrees of confidence. However, the degrees of confidence are presented in a reverse order for old judgments than for new judgments. Specifically, for old judgments, the response category furthest to the right represents the highest degree of confidence, but for new judgments, the opposite is true; the response category furthest to the right represents the lowest degree of confidence. Therefore, it could be that using the scale was not initially intuitive for participants and required several trials before participants are able to correctly use the scale. Other recognition studies have eluded this issue by having a large number of test trials (Heathcote, 2003; Pratte et al., 2010; Yonelinas, 1994) that could have allowed participants more time to practice breaking down their dichotomous response into six categories—perhaps due to implicit learning—allowing them to use the scale more appropriately.

Aim 3

The critical comparisons related to the hypotheses delineated in Aim 3 of this dissertation had to do with the extent to which familiarity influenced recognition of test pairs. Re-Paired test pairs—pairs that recombined a studied cue and a studied target—were created to test the limits of familiarity. According to the CLS model, test pairs that contain both a studied cue and studied target, even in a configuration different from their studied context, should produce a higher familiarity signal than either control or FTx test pairs. The DPSD model would also predict that Re-Paired test pairs would be significantly more likely to be falsely recognized than control and FTx test pairs due to an overall higher familiarity signal resulting from the previous encounter with both the cue and target. However, the CLS model would not predict that any 6-response given to Re-Paired test pairs would be due to recollection because of a lack of feature overlap between the Re-Paired item's representation and the studied item's stored representation.

Because both the CLS and DPSD models predict that Re-Paired test pairs should be more familiar than the FTx and control test pairs, then they would also predict that any high-confidence responses obtained would only be due to familiarity. Therefore, Hypothesis 9 predicted that Re-Paired test pairs would be falsely recognized at a higher rate than U-control test pairs, and Hypothesis 10 predicted that, because this higher rate of false recognition would be due to familiarity rather than recollection, the response distribution of Re-Paired test pairs would be more Gaussian than U-control test pairs. Both hypotheses were supported. The false recognition rate was higher for Re-Paired than U-control test pairs, and the response distribution was similar for the two test pairs, suggesting that the same cognitive process was used to recognize both.

A follow-up analysis compared the response distribution of Re-Paired test pairs with those of FTi test pairs to determine whether different retrieval processes were used for both pairs. We predicted that responses to Re-Paired test pairs would have a more Gaussian-shaped response distribution than those to FTi test pairs, which were predicted to have a dichotomous distribution due to recollection. If the implicit activation of the FTi false target led to recollection, as predicted by the CLS model, then the Re-Paired word pairs should have received significantly more middle-responses than FTi word pairs and their response distributions would have been different. Response distributions of FTx test pairs were also included to determine whether FTx and Re-Paired test pairs were retrieved via the same retrieval process. FTx pairs were predicted to have a similar Gaussian-shaped response distribution due to the influence of familiarity on false recognition because the test cue had been studied. However, the impact of familiarity should be greater for Re-Paired test pairs because both the test cue and test target were studied, which should have produced even more middle-responses on the scale than FTx pairs. However, the response distributions were similar for FTi, FTx, and Re-Paired test pairs. This finding suggests that the same cognitive processes were responsible for all three test pairs, a finding that is contrary to the predictions of the CLS model. This finding does provide support for the DPSD model, which states that any false recognition should be due to familiarity and not recollection.

The follow-up analysis regarding 6-responses across the three test pairs does not support this statement, however, because the three test pairs did receive a significantly different number of 6-responses. FTi and Re-Paired test pairs were the most likely to receive a 6-response, followed FTx test pairs. This result coupled with the response distribution comparison for FTi, FTx, and Re-Paired test pairs is puzzling in that it suggests that FTi and Re-Paired pairs had the highest familiarity signal. However, according to predictions from the CLS and DPSD models, only the Re-Paired pairs should have had high familiarity due to both the cue and target having been studied previously. One explanation is that the implicit activation of the FTi false-target also led to an increased level of familiarity. Because the false target was implicitly activated during study, when it was presented with the studied cue at test, the degree of familiarity for FTi test pairs could have been more similar to that of the Re-Paired test pairs; both test pairs could have a higher degree of familiarity resulting from re-presentation of both a studied cue and a studied target, albeit implicitly studied in the case of FTi test pairs. This supposition would be supported by the results from the follow up analysis that revealed that the two test pair types had similar response distributions. Therefore, it could be that the implicit activation of the false target for FTi pairs led to an increased level of activation in the MTLC that resulted in a stronger familiarity signal, increasing false recognition for those pairs.

Another explanation is that the false recognition of the Re-Paired test pairs was based on some heuristic other than familiarity. This explanation has potential because the manipulation of the Re-Paired test pairs was fundamentally flawed, allowing for another mechanism for identifying old versus new Re-Paired test pairs. Although a new, appropriate control test condition was created consisting of unrelated cue-target pairs to prevent participants from using the heuristic that any unrelated pairs were not studied, none of the unrelated pairs presented during study were presented again at test. Instead, all of the unrelated studied pairs were rearranged to create the test pairs leaving no unrelated studied pairs which participants could recognize as old. Therefore, the heuristic that could have been used was to call any unrelated pairs new. However, the idea that participants used relatedness as a heuristic to identify unrelated test pairs as new is not supported by our data. If participants used relatedness as a heuristic for identifying unrelated pairs new, then we would not have seen any false alarms for either Re-Paired and U-control pairs. However, 41% of the time, participants falsely identified unrelated test pairs (both Re-Paired and U-control) as old, demonstrating that they did not use relatedness as a heuristic for identifying unrelated pairs as new.

It could be the case that, despite our best efforts to select appropriate test pairs to examine the differential contribution of recollection versus familiarity to false recognition, a behavioral analysis is not sufficient to disentangle these two processes. It also could be that some FTi pairs were falsely recognized due to recollection, but that familiarity resulted in false recollection of more FTi test pairs than recollection. Given the current behavioral data, the contribution of each is only open to supposition. Significantly more false alarms and more 6-responses for FTi than FTx test pairs suggests the influence of recollection, but the response distributions and ROC analyses suggest only the influence of familiarity. Therefore, in an attempt to disentangle these two processes and determine whether recollection also contributed to the recognition of FTi test pairs, we conducted a modeling analysis. For the modeling analysis, we developed a series of models that included a false recollection parameter to determine whether the inclusion this parameter would provide a better fit of the behavioral data than the DPSD model. If the false recollection effect is being overshadowed by familiarity in the behavioral analysis, then the modeling analysis should be able to disentangle recollection from familiarity by allowing for the influence of false recollection, something not included in the DPSD model. By allowing for the influence of false recollection, the modeling analysis will be able to determine whether a model with a false recollection parameter or the DPSD model provides a better fit of the behavioral data. This analysis will determine whether the inclusion of a false recollection parameter is necessary to fit the data.

CHAPTER VII

MODELING RESULTS

Although a behavioral analysis provided some insight into the processes impacting recognition by allowing us to test *a priori* hypotheses and view the response distributions, this type of analysis is limited in terms of allowing us to directly test the influence of familiarity versus recollection. Therefore, to compensate for this limitation, we also performed a modeling analysis of the data. Modeling the data will allow us to test whether a model that only allows for familiarity nonstudied items or a model that allows for both familiarity and recollection provides a better fit of the data produced by the experiment. We did this by fitting both the DPSD model as well as two models that allowed for false recollection. To determine each model's ability to capture data patterns, while taking model complexity into account, a series of model fitting statistics including the AIC and BIC were performed for model comparison. In the following sections, we review the DPSD model and its subsequent amendments that included false recollection parameters.

The DPSD Model

According to the DPSD model, performance on a recognition task is the direct result of one of two retrieval processes: familiarity or recollection. In the model, when a studied item is presented for judgment, one of two processes occurs. Recollection is an all-or-none process that either occurs or not for the presented item. The model states that, during the recollection process, information associated with the presented studied item is attempted to be consciously retrieved. If the item is successfully retrieved, then the highest confidence old response is given. However, if recollection fails to occur, then responses are based on familiarity. Figure 26 is a complete graphical representation of the DPSD model. Because recollection is considered to be an all-or-none process, recollection occurs with some probability between zero and one, denoted as *R*, in the DPSD model. This probability is a free parameter that is estimated when the model is fit to behavioral data, providing a measure of how often recollection occurs. When recollection occurs, participants are assumed to respond "Sure Old", or the highest confidence old rating. Recollection will fail to occur with a probability of $1 - R$, and if this occurs, then responses are based on a signal-detection process, which is theorized to represent familiarity. Also, because familiarity is only relied on after recollection has failed, familiarity is represented as $1 - R$. It should be noted, that according to Yonelinas (1994, 2002), familiarity and recollection are not dependent. Familiarity always occurs with or without recollection. The model states that familiarity is only needed if recollection fails (Yonelinas, 1994, 2002).

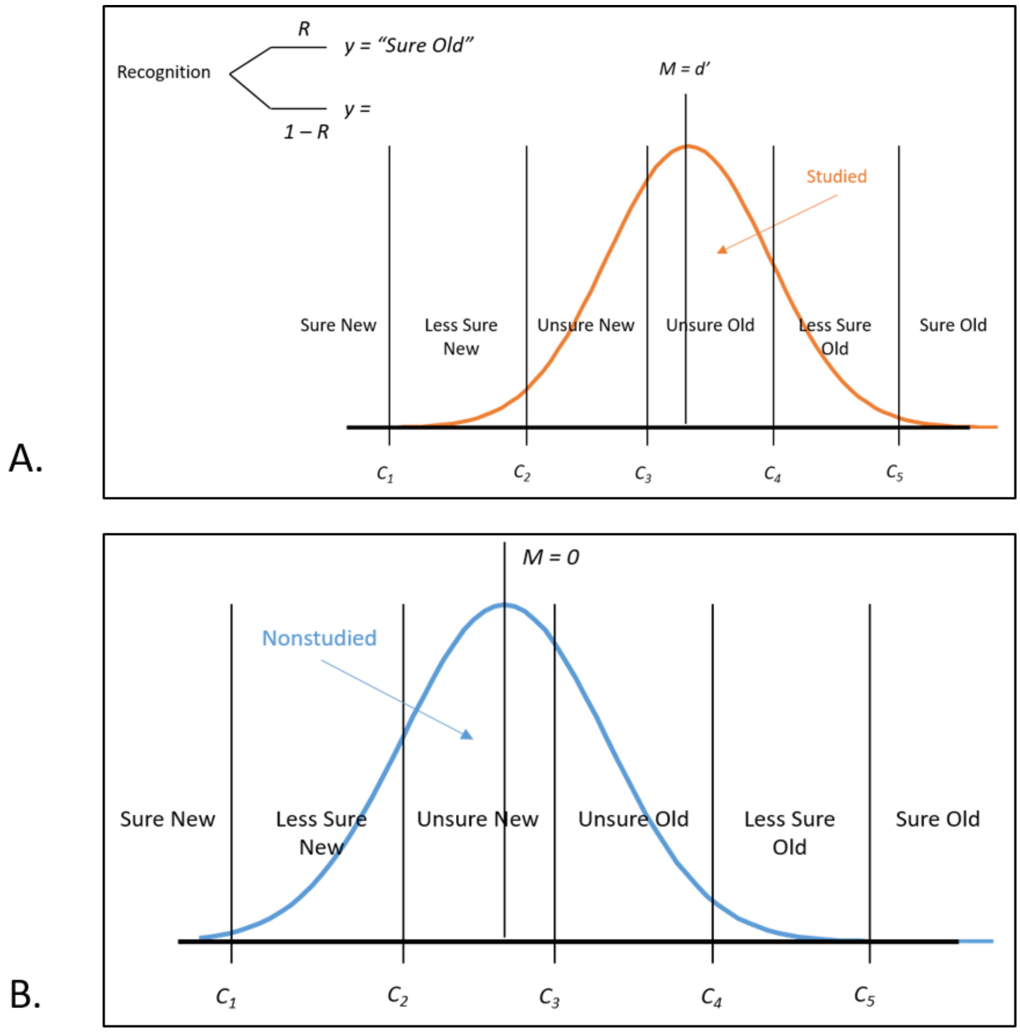


Figure 26. A) A graphical representation of the DPSD model for *studied* items. B) A graphical representation of the DPSD model for *nonstudied* items which includes a false recollection parameter.

In the DPSD model, familiarity is modeled as an equal variance signal-detection model, which assumes equal variance for both the studied and nonstudied item distributions. From the signal detection model, two types of parameters can be gathered that, together, determine how familiarity produces confidence recognition responses. These two parameters are discriminability and criterion threshold. Discriminability is denoted as *d'* and criterion threshold is annotated as C_k. Discriminability, or *d'*, represents the space between studied and nonstudied distributions, and theoretically represents the ability of a person to discriminate between studied and nonstudied items. The criterion, or C_k, represents the threshold of strength that an item's level of familiarity must surpass to be categorized as one of the 6 responses. In a typical old/new test, there is only one criterion threshold because there are only two possible categories that items can be grouped into. However, having only two categories is not sufficient to fit the DSPD model. Using a 6-point confidence recognition scale provides a more complete representation of both the studied and nonstudied item distributions by further dividing the distributions into six categories. The 6-point confidence recognition scale is favorable to the typical old/new recognition test because different theories, such as DPSD, make specific predictions regarding how these confidence ratings map on to process parameters such as recollection and familiarity. By dividing the distributions into six categories, five criteria values (shown in Figure 26) are used to map familiarity onto the possible six responses. According to the theory, these five criteria values correspond to the different levels of familiarity needed to provide a specific response. The item distribution area between each criterion value is the probability of making that response. Also, according to signal-detection theory, the criterion values are the same for both studied and nonstudied item distributions; given a level of familiarity for an item, either studied or nonstudied, the same response will be given. Therefore, the five criterion threshold values, discriminability, and recollection each serve as a parameter in the DPSD model for a total of seven parameters. It should be noted, though, that these parameters account for situations in which there is a single nonstudied-item condition. For each additional nonstudied-item condition, there is a separate nonstudied parameter (i.e., two nonstudied conditions equals two separate discriminability values). Because the Payne-Eakin paradigm has five, separate nonstudied-item conditions, a separate discriminability value was assigned to each nonstudied item condition to allow for varying rates of familiarity.

Additionally, for a modeling analysis, one parameter must always serve as the fixed parameter to provide the model a starting point for all other parameter values. Traditionally, in a signal-detection analysis, the mean of the nonstudied item distribution is fixed at zero instead of the middle criterion, as seen in Figure 26B. However, for this analysis, the middle criterion, C_3, served as the fixed parameter; therefore, the value of this parameter was always set to zero. Because this study had five nonstudied-item conditions, two of which were control conditions, it was decided that fixing the middle criterion would be more appropriate as it would allow for the mean of the nonstudied item distributions to vary. As a result, *dn* values for nonstudied items will be both positive and negative instead of just positive. Therefore, the DPSD model for this study has a total of 11 parameters. Table 8 shows a full list of DPSD parameters.

Table 8

A List of Parameters for the DPSD Model

Parameter:	Abbreviation:
Recollection	R
Studied Discriminability	d_S
FTi Discriminability	d_{FTi}
FTx Discriminability	d_{FTx}
Re-Paired Discriminability	d_{RP}
R-control Discriminability	d_{Rc}
U-control Discriminability	d_{Uc}
Criterion 1	C_1
Criterion 2	C_2
Criterion 3	*0*
Criterion 4	C_4
Criterion 4	C_5

Predictions of the DPSD model, and how different parameters affect these predictions, were explored using ROC curves. An ROC curve is a plot where hits, or studied items correctly identified as old, are plotted as a function of false alarms, or nonstudied items incorrectly identified as old. On an ROC curve, each response (e.g., 6-response) is plotted with the proportion of hits for that response as a function of the proportion of false alarms for the response (i.e., the proportion of studied items given a 6-response as a function of the proportion of nonstudied items given a 6-response). With each additional response plotted on the curve, the curve point reflects a cumulative point (i.e., 6-responses + 5-reponses) to produce the entirety of the curve. Given the parameters of the DPSD model, we can compute a predicted ROC curve:

$$P_S\ (C_k) = R_S + (1 - R_S) * (1 - \Phi\ (C_k - d_S))$$

and

$$P_N\ (C_k) = 1 - \Phi\ (C_k - d_n)$$

P_S (C_k) is the cumulative probability of making each response, beginning with 6-reponses; P_N (C_k) is the probability that a given nonstudied item is recognized as old given some level of criterion threshold; R_S is the probability that a studied item is recollected; d_S is the mean level of discriminability for studied items; and Φ is the cumulative density function of the standard normal distribution. The cumulative density function is a calculation of the area underneath a normal curve up to a specific value. The cumulative density function is used in the DPSD model to calculate the area underneath the normal curve for each criterion threshold. By calculating the area underneath the normal curve for each criterion, you can then calculate the area difference between two criteria which then gives you the probability of making a given response. For instance, if you calculate the area under the curve for C_2 and C_3, you can subtract the area for C_2 from C_3 to give you the area between C_2 and C_3 which represents the probability of giving a 3-response, as shown in Figure 26.

Changing either the recollection, *R*, or discriminability, *d'*, causes the ROC curve to change shape in different ways. If recollection is absent, or set to a null value, then the model reduces to a signal detection model which predicts a symmetrical curve as seen by the middle blue line in Figure 27. However, if recollection is high, then the model predicts a large increase in the number of 6-responses which increases the start of the curve on the y-axis as seen by the dark blue line in Figure 27. This increase then causes the curve to become asymmetrical by increasing the y-intercept. This asymmetry is often found in typical recognition studies (e.g., Egan, 1975; Murdock, 1982; Yonelinas 1994) and is the primary evidence Yonelinas (1994, 2002) has used to support the presence of recollection. Discriminability on the ROC curve is equal to the distance of the curve from the diagonal. Items that are easily discriminated from one another produce a larger distance from the diagonal than items that are not easily discriminated; whereas changing *R* increases the y-intercept of the ROC curve, either increasing or decreasing *d'* will move the curve either further away or closer to the diagonal (see light blue line in Figure 27 for a curve that has a reduced *d'* value). Finally, five criteria values are plotted and the summative area between each point is equal to the area of the curve.

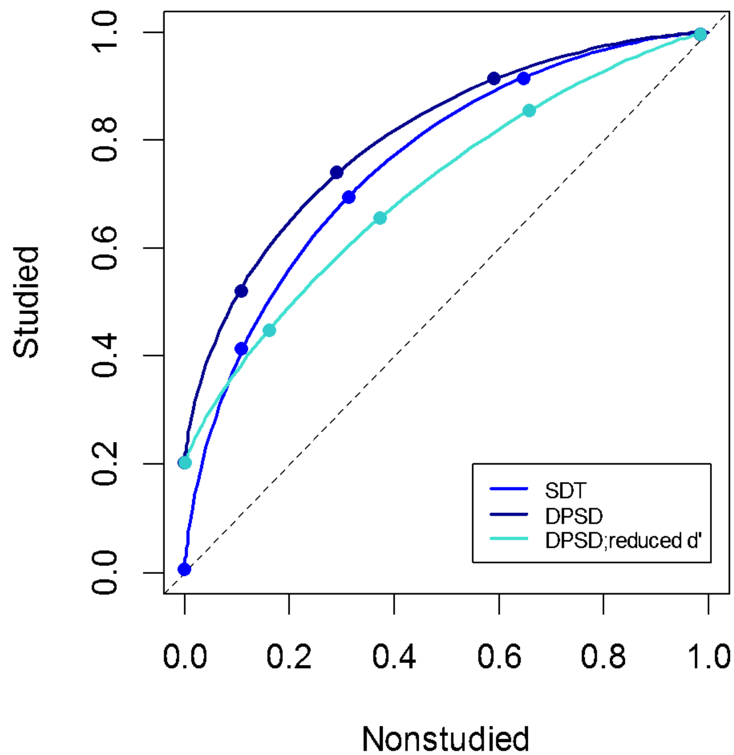


Figure 27. ROC curve for the SDT model with *d'* of 1; ROC curve of DPSD model with recollection at .2 and *d'* of 1; and ROC curve of DPSD model with *d'* reduced to .5.

Modifying the DPSD Model

Critically, according to the DPSD model, recollection can only occur for studied items. Therefore, the model states that judgments about nonstudied items can only be made based on familiarity (Yonelinas, 1994, 2002; Yonelinas et al., 2010). Because the DPSD model does not currently contain a mechanism to account for false recollection, the fourth aim of this dissertation was to develop and test a new model of recognition that contains a false recollection parameter. There is a large amount of behavioral evidence suggesting the occurrence of false recollection (e.g., Higham & Vokey, 2004; Jones, 2013; Lane & Zaragoza, 1995; Roediger & McDermott, 1995; Strange et al., 2011). Yonelinas et al. (2010) briefly acknowledges the possibility of recollection for nonstudied items; however, he argues that because false recollections occur at such a low rate, they are violations of the boundary conditions of the DPSD model and as such, including a mechanism to account for them is not necessary. However, we posited that increasing their incidence using the Payne-Eakin paradigm, would result in false recollection being able to be tested and evaluated a potential addition to the DPSD model. Therefore, we tested two models that include a false recollection parameter for nonstudied items to determine if they could provide a better account of the behavioral data produced by the Payne-Eakin paradigm than the DPSD model.

DPSD-FR

To amend the DPSD model, we developed a series of models that include a false recollection parameter. The first model that we developed is the dual process signal detection – false recollection model, or DPSD-FR model. The DPSD-FR model includes a separate false recollection parameter for each nonstudied item condition, for a total of five false recollection parameters. The inclusion of these five false recollection parameters increases the total number of parameters from 11 in the DPSD model to 16 in the DPSD-FR model. Table 9 provides a full list of the parameters.

Table 9

A List of Parameters for the DPSD-FR Model

Parameter:	Abbreviation:
Recollection	R
Studied Discriminability	d_S
FTi Discriminability	d_{FTi}
FTx Discriminability	d_{FTx}
Re-Paired Discriminability	d_{RP}
R-control Discriminability	d_{Rc}
U-control Discriminability	d_{Uc}
False Recollection - FTi	FR_{FTi}
False Recollection - FTx	FR_{FTx}
False Recollection - Re-Paired	FR_{RP}
False Recollection - R-control	FR_{Rc}
False Recollection - U-control	FR_{Uc}
Criterion 1	C_1
Criterion 2	C_2
Criterion 3	0
Criterion 4	C_4
Criterion 4	C_5

The purpose of the DPSD-FR model was to test whether false recollection occurs for all nonstudied test pair conditions at varying rates. As a result, the ROC curve predicted by the DPSD-FR model for both studied and nonstudied items is now:

$$P_S\ (C_k) = R_S + (1 - R_S) * (1 - \Phi\ (C_k - d_S))$$

and

$$P_N(C_k) = FR_N + (1 - FR_N) * (1 - \Phi(C_k - dn_k))$$

This model is almost identical to the DPSD model defined above, except for the inclusion of FR_N. FR_N is the probability that a nonstudied item is recollected. The inclusion of a false recollection parameter for all nonstudied items makes straightforward predictions about the ROC curve. Figure 28 shows a predicted ROC curve for the standard DPSD model in dark blue and a predicted ROC curve for nonstudied item conditions for the DPSD-FR model in red. Instead of the ROC curve having only an increased y-intercept, as seen with the DSPD model, if false recollection occurs, then the false recollection parameter will shift the x-intercept to the right for each of the nonstudied item conditions. This shift will occur because of an increased number of high-confidence responses for nonstudied test pairs.

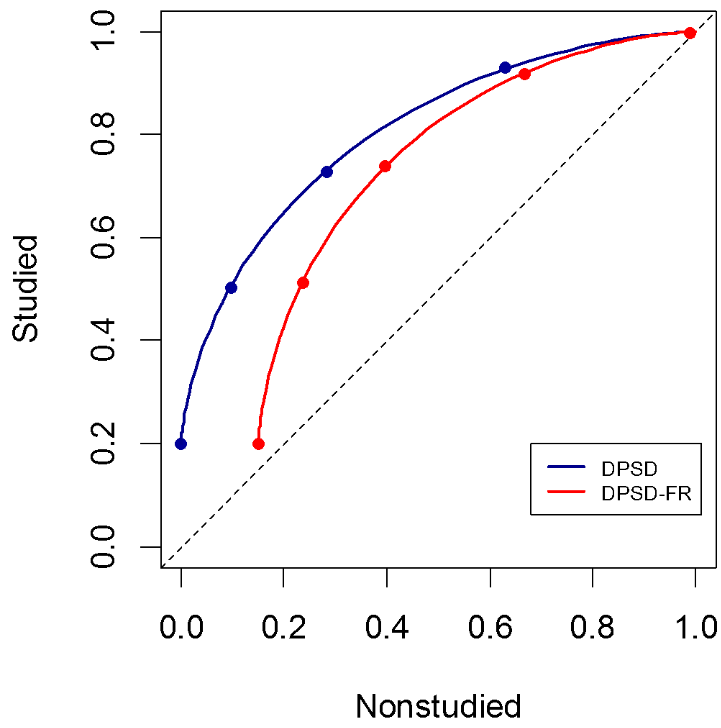


Figure 28. ROC curves for the DPSD model and DPSD-FR models. The dark blue line indicates a DPSD model with a recollection rate of .2, and the red line indicates a DPSD-FR model with a recollection rate of .2 and a false recollection rate of .15.

DPSD-FR1-FTi Model

The second model that we developed was the dual process signal detection–false recollection 1-FTi model, or DPSD-FR1-FTi model. The DPSD-FR1-FTi model includes a single false recollection parameter for FTi test pairs. The inclusion of this one false recollection parameter increases the total number of parameters from 11 in the DPSD model to 12 in the DPSD-FR1-FTi model. Table 10 provides a full list of each parameter.

Table 10

A List of Parameters for the DPSD-FR1-FTi Model

Parameter:	Abbreviation:
Recollection	R
Studied Discriminability	d_S
FTi Discriminability	d_{FTi}
FTx Discriminability	d_{FTx}
Re-Paired Discriminability	d_{RP}
R-control Discriminability	d_{Rc}
U-control Discriminability	d_{Uc}
False Recollection - FTi	FR_{FTi}
Criterion 1	C_1
Criterion 2	C_2
Criterion 3	0
Criterion 4	C_4
Criterion 4	C_5

The purpose of the DPSD-FR1-FTi model was to test whether false recollection occurs only for FTi test pairs. Because we originally hypothesized that FTi word pairs would be the only test pairs to facilitate false recollection, the DPSD-FR1-FTi model directly tests this hypothesis. The ROC curve predicted by the DPSD-FR1-FTi model for studied, nonstudied, and FTi items is now:

$$P_S\ (C_k) = R_S + (1 - \mathrm{R}_S) * (1 - \Phi\ (C_k - d_S)),$$

$$P_N\ (C_k) = FR_N + (1 - FR_N) * (1 - \Phi\ (C_k - d_S))$$

and

$$P_{FTi}\ (C_k) = FR_{FTi} + (1 - FR_{FTi}) * (1 - \Phi\ (C_k - d_{FTi})),$$

This model is almost identical to the DPSD-FR model defined above, except for the FR_{FTi} parameter. FR_{FTi} is the probability that a FTi test pair is recollected. The inclusion of a false recollection parameter for FTi test pairs makes straightforward predictions about the ROC curve. The same ROC predictions made for the DPSD-FR model can be made for the DPSD-FR1-FTi model. Instead of the ROC curve only having an increased y-intercept, as seen with the DSPD model, the false recollection parameter will shift the x-intercept to the right for FTi test pairs because of an increased number of high-confidence responses. The only difference between the two models' predictions is that the DPSD-FR model will predict a shift in the x-intercept for all nonstudied test pairs and the DPSD-FR1-FTi model will predict a shift in the x-intercept for FTi test pairs. Therefore, Figure 28 also shows a predicted ROC curve for the DPSD-FR1-FTi model in red.

Model Fitting and Comparison

Model fit was determined using a maximum likelihood estimation procedure (Myung, 2003). To calculate the maximum likelihood estimate, a likelihood function was computed for each model (Heathcote et al., 2006). The maximum likelihood estimate finds the model parameters that *maximizes* the likelihood function for each model. During this procedure, a search algorithm is used to search through the parameter space to find the maximum likelihood value given the best-fitting parameters, denoted $\hat{L}$. This likelihood value can then be used to determine which model provide the best account of the data using a variety of model comparison approaches.

More flexible models provide a better fit of the data due to added complexity, and not necessarily due to accuracy of the model's representation of the cognitive process.

Any method for comparing models must therefore consider both a model's fit to the data as well as the flexibility of the model (Pratte & Rouder, 2011). There is no one statistical test that can determine which of a set of models provides the best account of the data. Therefore, we used a method that also has been used by other recognition memory researchers to test DPSD and similar models (Heathcote et al., 2006). One way to consider both model fit and complexity is to compare models using the AIC and BIC model fit indices. These statistics characterize model fit using the maximum likelihood value, and penalize models for added complexity as characterized by the number of parameters in a model.

$$\text{AIC} = 2\,k - 2\ln(\hat{L})$$

and

$$\text{BIC} = \ln(n)\,k - 2\ln(\hat{L})$$

$\hat{L}$ denotes likelihood value of the given model, k denotes the number of parameters in each model, and n denotes the number of data points in the given data set. Each of these statistics quantifies model flexibility in terms of the number of parameters in the model, and attempts to mitigate the advantageous impact of model flexibility by providing a penalty for each parameter within the model. The main difference between the two statistics is that the BIC statistic has a stronger penalty for parameters and, as a result, tends to favor simpler models.

For the modeling analysis, we used both the AIC and BIC to determine which of our models provided the best account of the data. An AIC and BIC statistic was calculated for each model; smaller AIC and BIC values indicate better models. For all comparisons, the false recollection model (i.e., either DPSD-FR or DPSD-FR1-FTi) always served as the comparison model. For example, the criterion value for the false recollection model was always subtracted from the criterion value for the DPSD model. Therefore, negative difference values indicate an advantage of the DPSD model and positive values indicate an advantage of a false recollection model.

Although there are no strict cut off points for what differences in the AIC and BIC indicate a significantly better model, there are conventions that can be used to determine whether the difference between two models provides strong or weak evidence for one model over the other. According to Raftery (1995), a difference of 10 or larger indicates "very strong" evidence for one model over the other; a difference of 6 or more provides "strong" evidence for one model over the other; a difference of 2 or more provides "some" evidence for one model over the other; and a difference of 2 or less provides "weak" evidence for one model over the other. These conventions were used to determine the strength of evidence for each of our model comparisons.

Modeling Results

For the modeling analysis, we fit the behavioral data obtained from the dissertation experiment to the DPSD, DPSD-FR, and DPSD-FR1-FTi models at the aggregate level, or data averaged across individuals, as well as at the individual level. The modeling analysis sought to determine whether the inclusion of a false recollection parameter was necessary for fitting the false memory data produced by the Payne-Eakin paradigm. Hypothesis 11 predicted that either the DPSD-FR or DPSD-FR1-FTi model would provide a better fit of the behavioral data than the DPSD model.

Participants

A total of 64 participants participated in this study for course credit. Ten participants were removed from the behavioral analysis due to incomplete data, a chance level of identifying new items as old or vice versa, and a low rate of recognizing studied items as old, resulting in a total of 54 participants. These participants were also excluded from the modeling analysis. For the modeling analysis, an additional 16 participants were removed due to limited variability in responses. Specifically, participants whose responses consisted of 90% or more 1- and 6-responses were removed, resulting in a total of 38 participants in the modeling analysis. This criterion was determined necessary due to the difficulty of extracting meaningful information from fitting models to data missing responses from the 6-point confidence recognition scale.

Aggregate. One caveat with modeling the behavioral data was a possible lack of power due to participants not using the entire 6-point confidence recognition scale. To address this possible issue, we started with an analysis of the aggregated data to ensure adequate power. The proportion of responses for each condition was plotted as an ROC curve with responses to nonstudied items plotted as a function of responses to studied items. Those data are represented as points in Figure 29. We first fit the DPSD model to the aggregated data, as shown in Figure 29. Each nonstudied test condition was estimated to have a different *d'* value by the model with FTi word pairs being the least discriminable, evidenced by the ROC curve being closer to the diagonal. The reason for this is that the DPSD model can only account for varying rates of false alarm by allowing *dn* to vary. Table 11 below provides a list of the estimated parameter values for the

DPSD model. According to the DPSD model, recollection is occurring at a rate of approximately 10% for the aggregated data.

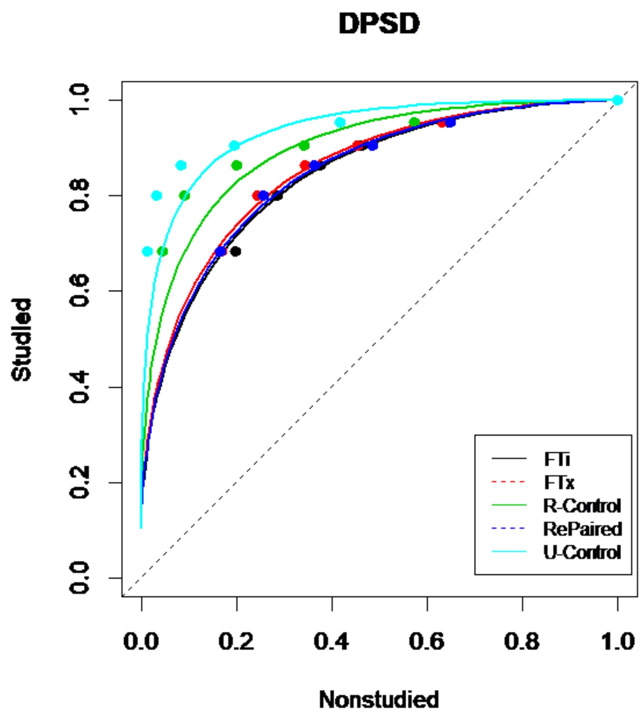


Figure 29. ROC curves for each condition (aggregated over participants). Behavioral data are represented by the points whereas predictions from the DPSD model are represented by the lines.

Table 11

Estimated Parameter Values for the Aggregated Data from the DPSD Model

Parameters:	DPSD
R	.10
d_S	.98
d_{FTi}	-.34
d_{FTx}	-.41
d_{RP}	-.37
d_{Rc}	-.73
d_{Uc}	-1.10
C_1	-.78
C_2	-.30
C_3	0
C_4	.30
C_5	.61

Although the DPSD model is providing a good fit for a portion of the data, it is failing to account for all nonstudied-item conditions. Specifically, the DPSD model appears to be failing to account for, or predict, a shift in the x-intercept for several of the nonstudied-item conditions particularly for the U-control and R-control conditions, as shown in Figure 29. A possible reason for this is due to the DPSD model's recollection constraint. Because the DPSD model has a single recollection parameter that can only occur for studied items, it cannot account for a large increase in the number of high-confidence false alarms. Therefore, to account for this increase in high-confidence false alarms, the first model that we tested was the DPSD-FR model.

DPSD-FR model. The DPSD-FR model includes a separate false recollection parameter for each of the nonstudied item conditions. By allowing for a separate false recollection parameter for each nonstudied condition, the model allows for varying rates of false recollection for each condition. The purpose of the DPSD-FR model was to determine whether false recollection occurs for all nonstudied item conditions. To do this, we fit the DPSD-FR model to the aggregated behavioral data. The DPSD-FR model has a total of 16 parameters: *R*, *ds*, five *FR* parameters, five *dn* parameters, and four criteria (Table 9). The behavioral data is again presented as points in Figure 30 with the predictions of the DPSD and DPSD-FR models represented as lines.

The model predicted each nonstudied item condition to have a different *dn* value with FTi word pairs being the least discriminable, evidenced by the ROC curve being closer to the diagonal. Additionally, the model allowed for a different *FR* value for each of the nonstudied-item conditions, evidenced by the different x-intercept values. According to the DPSD-FR model's predictions, false recollection is occurring at the highest rate for FTi word pairs (.20) followed by FTx word pairs (.17), Re-Paired word pairs (.17), R-control word pairs (.04), and U-control word pairs (.01). These estimates seem to support our hypothesis that the false memory effect obtained in the Payne-Eakin paradigm is due to false recollection, not familiarity, facilitated by the implicitly activated FTi false targets. Evidenced by the ROC curves in Figure 30, by allowing the model to account for false recollection, the model appears to be able to account for a substantially larger portion of the data.

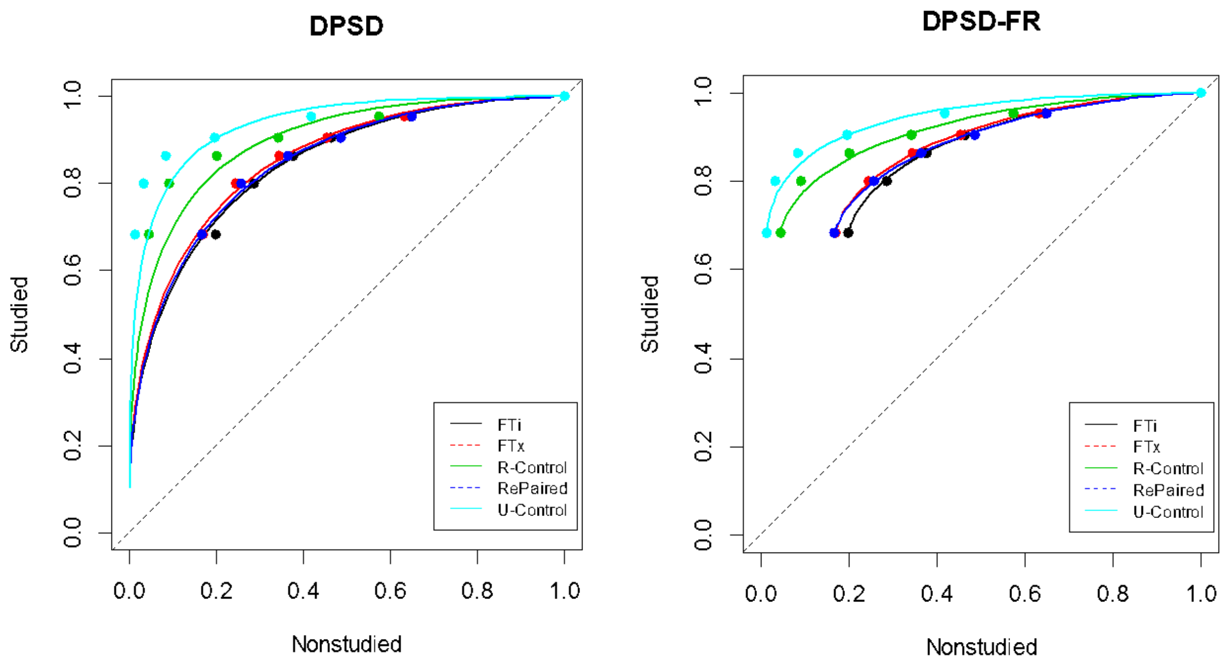


Figure 30. Comparison of DPSD and DPSD-FR model fits. The behavioral data is represented by the points, and the predicted model data is represented by lines.

Because the DPSD-FR model includes an additional five parameters which allow for greater flexibility, the DPSD and DPSD-FR models were compared using criteria that test the fit of the two models while accounting for model likelihood and complexity (i.e., number of parameters). We used the AIC and BIC statistics which provide a penalty for the number of parameters in a model to mitigate the advantage of model complexity. However, the BIC has a higher penalty for model complexity, often resulting in the test favoring the simpler model.

For all AIC and BIC comparisons, the false recollection model (i.e., DPSD-FR or DPSD-FR1-FTi) always served as the comparison model; the information criterion value for the false recollection model was subtracted from the information criterion value for the DPSD model. Therefore, negative test values indicate an advantage of the DPSD model and positive values indicate an advantage of the false recollection model. Table 12 shows the difference test values for the DPSD and DPSD-FR models. According to Raftery (1995), both the AIC and BIC comparisons reveal very strong evidence that the additional fit that the DPSD-FR model provides to the data justifies the inclusion five extra parameters in the model. This result demonstrates that the inclusion of a false recollection parameter for each nonstudied condition is necessary to fit the data produced by the Payne-Eakin paradigm, supporting our hypothesis.

Table 12

AIC and BIC Difference Scores for the Aggregated Data

Model Comparison	AIC'	BIC'
DPSD vs. **DPSD-FR**	93.48	77.71
DPSD vs. **DPSD-FR1-FTi**	19.58	17.94
DPSD-FR vs. DPSD-FR1-FTi	-73.90	-85.29

Although the DPSD-FR model supports our hypothesis by providing a better fit of the data than the DPSD model, it could be that only one false recollection parameter was necessary to fit the data. We originally predicted that the implicit activation of the false target during encoding for FTi word pairs would facilitate false recollection. However, we did not make specific predictions regarding false recollection for other nonstudied test pairs. Therefore, we also tested a model containing only one false recollection parameter for FTi test pairs (i.e., no false recollection for the other four nonstudied item conditions).

DPSD-FR1-FTi. The DPSD-FR1-FTi model includes only one false recollection parameter for FTi test pairs, resulting in a total of 12 parameters: *R*, ds, five *dn*, one

FR_{FTi}, and four criteria (see Table 10). By having only one false recollection parameter, false recollection for all other nonstudied item conditions was set to zero. The purpose of this model is to determine whether false recollection only occurs for FTi word pairs and not for other nonstudied pairs. To do this, we fit the DPSD-FR1-FTi model to the aggregated behavioral data. The behavioral data is presented as points in Figure 31 with the predictions of the DPSD and DPSD-FR1-FTi models represented as lines.

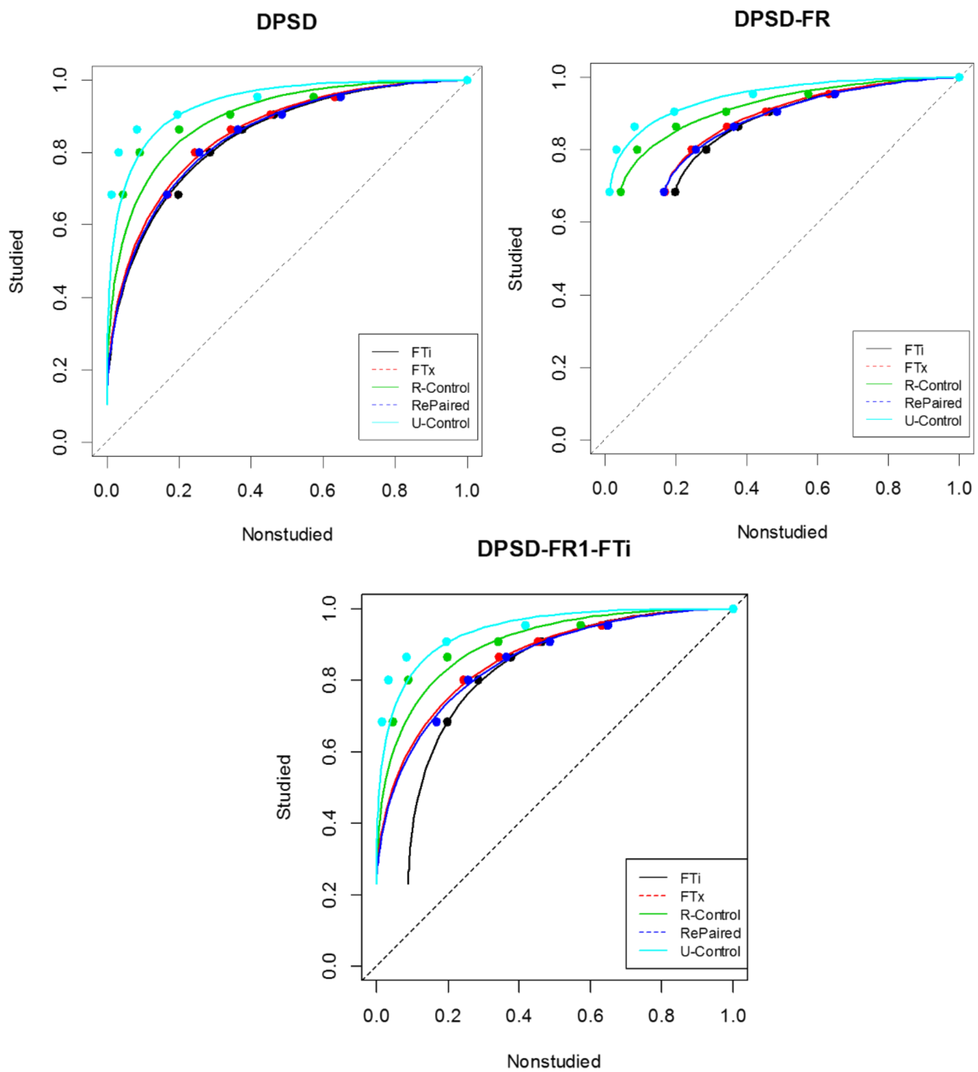


Figure 31. Comparison of DPSD and DPSD-FR1-FTi model fits. The behavioral data is represented by the points, and the predicted model data is represented by lines.

The DPSD-FR1-FTi model estimated each nonstudied item condition to have a different *dn* value with FTi word pairs being the least discriminable, evidenced by the ROC curve being closer to the diagonal. Additionally, the model predicted a different *FR* value for the FTi word pairs, evidenced by a different x-intercept value. According to the

DPSD-FR1-FTi model's predictions, false recollection is occurring approximately 9% of the time for FTi word pairs. Additionally, evidenced by the ROC curves in Figure 31, it appears that the DPSD-FR1-FTi model is accounting for a larger portion of the data than the DSPD model. However, it also seems that it is accounting for a smaller portion of the data than the DPSD-FR model. This finding that that the DPSD-FR1-FTi model is seemingly accounting for less of the data than the DPSD-FR model is likely to due to the higher degree of restriction in the DPSD-FR1-FTi model.

Because the DPSD-FR1-FTi model includes an additional parameter that allows for greater flexibility, the DPSD and DPSD-FR models were compared using the AIC and BIC fit indices. Table 12 shows the difference test values for the DPSD, DPSD-FR1-FTi, and DPSD-FR models. According to both the AIC and BIC comparison, the inclusion of a false recollection parameter in the DPSD-FR1-FTi model provided a better fit of the data than that DPSD model. According to Raftery (1995), both the AIC and BIC tests reveal strong evidence that the additional fit to the data justifies the one additional parameter in the DPSD-FR1-FTi model. This finding demonstrates that the inclusion of a false recollection parameter for FTi test pairs was necessary to fit the data. Therefore, the modeling results of the aggregated data supports Hypothesis 11, which predicted that a model containing a false recollection parameter would provide a better fit of the Payne-Eakin data than the DPSD model.[1]

[1] Several other models were also tested; however, the theoretical motivation for those models were not within the scope of this dissertation and not included. For a full list of models that were tested and their results, see Appendix F.

We also compared the fit of the DPSD-FR1-FTi model to the DPSD-FR model. The comparison revealed that the DPSD-FR model provided a better account of the data than the DPSD-FR1-FTi model, suggesting that false recollection occurs for more than one test-pair condition. This result is interesting in that it demonstrates that, although the DPSD-FR model contains five additional parameters and the DPSD-FR1-FTi model contains only one additional parameter, the added benefit of those five parameters outweighs the penalty of the AIC and BIC comparisons.

Individual. Fitting the model to aggregate data can be a very useful tool. However, it can also fail to account for process effects occurring at the individual level. Because responses are collapsed across participants, response patterns for individuals can be qualitatively different from the individual level (Pratte et al., 2010). Pratte et al. (2010) discussed the impact of qualitative differences between aggregate and individual level data on modeling data by demonstrating that averaging over participant responses can result in unaccounted participant variability, resulting in masking the underlying cognitive process. The implication of their findings can be seen in our data when comparing individual participant ROC curves to the aggregated data (see Figure 32). For instance, one difference that can be seen is that is that participants 4 and 29 have a higher recollection rate, demonstrated by the y-intercept, than participant 33 or the aggregate. These large differences in the rate of recollection can result in a large amount of unaccounted variability in the aggregate form. This unaccounted variability then impacts the ability of the model to accurately estimate recollection due to missing data, which results in a masking of the recollection process as stated by Pratte et al. (2010).

Therefore, the analysis on the aggregate data does not provide a final determination as to which model provides the most accurate account of the data.

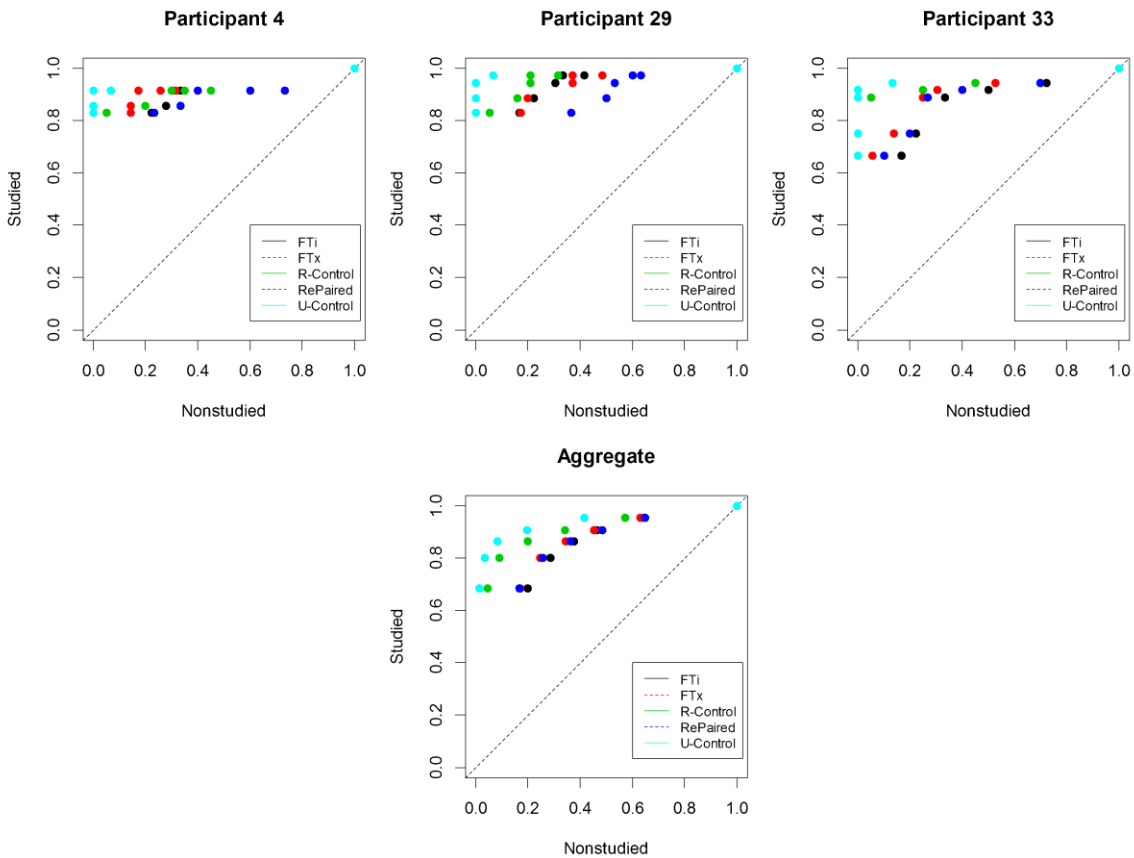


Figure 32. Comparison of response patterns between a group of individual participants and aggregate data.

DPSD. We fit the DPSD model to each individual participant. The parameter values and standard errors estimated by the DPSD model are reported below in Table 13. Contrary to the aggregate analysis, the estimated value for recollection increased from 10% for the aggregate to 34% for the individual. This increase shows a vast underestimation of recollection in the aggregate analysis and demonstrates that more people were relying on recollection. Consistent with the aggregate analysis, each nonstudied item condition was estimated to have a different *dn* value by the model with FTi word pairs having the lowest *dn* value. However, the DPSD model also estimated the *dn* for Re-Paired test pairs to be similar (not significantly different) to FTi word pairs, suggesting the same level of discriminability for both types of pairs. This finding suggests that within the DPSD model, both FTi and Re-Paired test pairs were inherently more familiar than the other nonstudied test pairs. However, in the next section, we will again introduce two false recollection models that will test whether the reduced discriminability for FTi pairs is partly due to the influence of recollection.

Table 13

Summary of DPSD Estimated Parameter Values for Individual Data

Parameter:	*M*	*SE*
R	.34	.05
d_S	.54	.20
d_{FTi}	-.38	.07
d_{FTx}	-.47	.08
d_{RP}	-.39	.06
d_{Rc}	-.82	.07
d_{Uc}	-1.48	.19
C_1	-1.09	.13
C_2	-.41	.05
C_3	0	0
C_4	.55	.05
C_5	2.86	.26

Evidence from the aggregate analysis suggested that false recollection can occur for nonstudied items, suggesting that the DPSD model should be amended with the inclusion of a false recollection parameter. Therefore, we next fit the DPSD-FR model to the individual data. This model contains a separate false recollection parameter for each nonstudied item condition, increasing the number of parameters from 11 to 16. The purpose of fitting this model was to see if false recollection occurs for all items, both studied and nonstudied.

DPSD-FR. The next model that we fit to the individual data was the DPSD-FR model. The DPSD-FR model has a total of 16 parameters: *R*, *ds*, five *dn*, five *FR*, and four C_k. A sample of participants was selected to demonstrate how the DPSD-FR model was fitting the data at the individual level compared with the DPSD model (see Figure 33). Overall, the DPSD model is accounting for a large portion of the data but is failing to predict a portion of the FTi word pairs, as seen with participant 4 and 29. It appears that the primary reason the DPSD model is unable to account for the FTi word pairs is due to its inability to move along the x-axis, representing false recollection. Therefore, it seems that the inclusion of a false recollection parameter would allow the model to capture more of those data points. This prediction can be seen in the fit of the DPSD-FR model. Overall, the DPSD-FR model appears to capture more of the data points, including FTi word pairs and Re-Paired word pairs. However, to determine which model provides the best account of the data, a series of fit tests are needed to determine whether the inclusion of a false recollection parameter was worth the added fit.

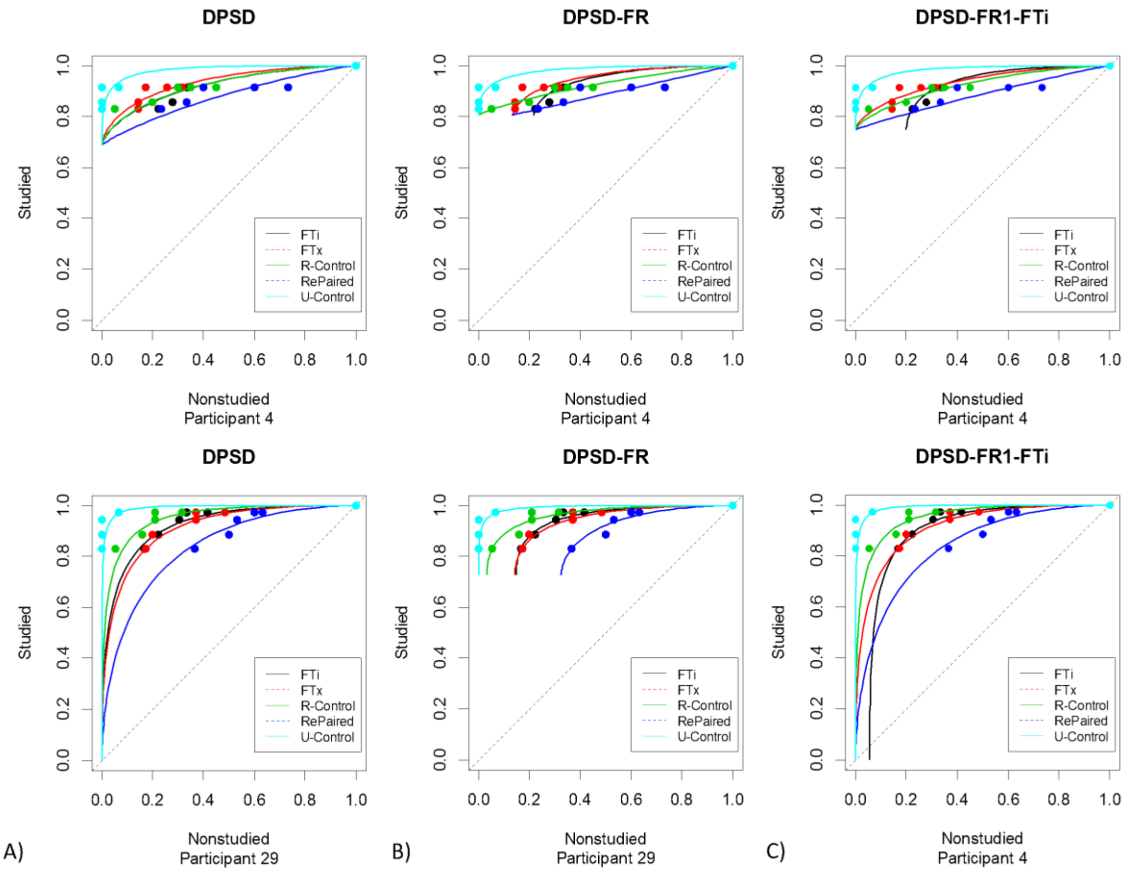


Figure 33. Comparison of model fits between the DPSD, DPSD-FR, and DPSD-FR1-FTi models for participant 4 and 29. A) The predicted model fit for the DPSD model. B) The predicted model fit for the DPSD-FR model. C) The predicted model fit for the DPSD-FR1-FTi model.

The fits of the DPSD and DPSD-FR models were compared using AIC and BIC tests. Table 14 shows the mean difference test values and standard error. The mean test value for both the AIC and BIC tests reveals that the added fit to the data does not justify the five extra parameters in the DPSD-FR model. We evaluated these test values for each participant to see how many participants' data was better fit by the DPSD model. At the individual level, we found that the DPSD model had the advantage (smaller AIC value) for 29 out of 38 participants, demonstrating that the addition of the five false recollection parameters was not necessary to fit their data. However, 6 of the 29 participants who favored the DPSD model had weak evidence to support the model (i.e., AIC difference value of two or less), suggesting that a definitive conclusion could not be drawn for those participants. For the BIC test, 36 out of 38 participants favored the DPSD model with the majority of them having strong evidence to favor the DPSD model. Therefore, contrary to the aggregate analysis, the individual analysis does not support our hypothesis that the inclusion of a false recollection parameter was necessary to fit the nonstudied conditions.

Table 14

Mean and Standard Error AIC and BIC Difference Scores for the Individual Data

Model Comparison	AIC’		BIC’	
	M	*SE*	*M*	*SE*
DPSD vs. DPSD-FR	-2.28	1.37	-18.00	1.37
DPSD vs. DPSD-FR1-FTi	-0.79	.82	-3.93	.82
DPSD-FR vs. **DPSD-FR1-FTi**	1.67	1.00	14.07	1.38

Although the model comparison tests do not support our hypothesis, interestingly, the DPSD-FR model did provide estimates of false recollection for each of the nonstudied item conditions. Table 15 provides a list of means and standard errors for each false recollection parameter. To test for differences between each of the false recollection estimates, a repeated measures ANOVA was conducted. False recollection estimates were significantly different across the five test pairs, $F(1, 37) = 57.84$, $p < .001$, $\eta_p^2 = .61$. FTi test pairs were estimated to have the highest rate of false recollection, followed by Re-Paired, FTx, R-control, and U-control pairs, suggesting that FTi test pairs were falsely recollected at a higher rate than other test pairs, consistent with our predictions. This finding, combined with the results from Hypotheses 4 and 5, provides support that the implicit activation of the FTi target during study did lead to false recollection of those pairs.

Table 15

Summary of DPSD-FR Estimated Parameter Values for Individual Data

Parameter:	*M*	*SE*
R	.34	.05
d_S	.54	.20
d_{FTi}	-.38	.07
d_{FTx}	-.47	.08
d_{RP}	-.39	.06
d_{Rc}	-.82	.07
d_{Uc}	-1.48	.19
FR_{FTi}	.17	.02
FR_{FTx}	.15	.02
FR_{RP}	.14	.02
FR_{Rc}	.03	.01
FR_{Uc}	.01	.00
C_1	-1.09	.13
C_2	-.41	.05
C_3	0	0
C_4	.55	.05
C_5	2.86	.26

Although the model comparison tests do not support the DPSD-FR model, it could be that the additional five false recollection parameters were too much added complexity for the individual data. Specifically, it could be that some nonstudied conditions had false recollection and some did not. This supposition is supported by the finding that false recollection for control test pairs was estimated to be less than .05. Therefore, the inclusion of a false recollection parameter for conditions that do not have false recollection could have led to an overly complex model that does not provide a good fit of the data. Both the AIC and BIC statistics attempt to prevent this possibility from occurring by providing a penalty for each parameter included in a model. It also could be that false recollection only occurs for FTi word pairs and not for other nonstudied test pairs. We originally predicted that false recollection would only occur for FTi test pairs and not for other nonstudied pairs. However, we tested the full model to determine whether false recollection occurs for all nonstudied items at a lower rate. Therefore, the next model that we tested was the DPSD-FR1-FTi model to determine whether the inclusion of a single false recollection parameter for FTi word pairs would provide a better fit of the data.

DPSD-FR1-FTi. The DPSD-FR1-FTi model has a total of 12 parameters: *R*, *ds*, FR_{FTi}, five *dn*, and four C_k. Although, the initial result of the DPSD-FR1-FTi analysis for the aggregated data showed no real added benefit of the single false recollection parameter over the DPSD and DPSD-FR models, we hypothesized that the inclusion of a single false recollection parameter might still be an option for the individual data due to the lack of influence from participant variability. We first compared the DPSD and DPSD-FR1-FTi models using an AIC and BIC test. Table 14 above shows the mean and standard error of the difference of the test values. The mean test value for the AIC and BIC tests reveals that the DPSD and DPSD-FR1-FTi model cannot be distinguished from one another using Raftery's (1995) standards. Because the mean AIC difference was -.79 and the mean BIC difference was -3.93, it cannot be concluded as to which model provides a better fit of the data. Although the BIC test provides some evidence to suggest the DPSD model over the DPSD-FR1-FTi model, it has not been determined in the literature as to whether the AIC or BIC provides a more definitive result (Heathcote et al., 2006). Instead, the convention is to choose the model that garners support from both tests, which our analysis does not. Therefore, we cannot conclusively determine whether the DPSD or DPSD-FR1-FTi model provides a better account of the behavioral data.

We broke down the AIC and BIC difference scores for each individual to see how many people adhered to this inconclusive finding. We found that the DPSD model had the slight advantage (smaller AIC value) for 24 out of 38 participants on the AIC test and the slight advantage for 31 out of 38 participants on the BIC test, suggesting that the addition of the false recollection parameter was not necessary to fit their data. However, the difference between these scores was negligible. For 22 out of 24 participants who supported the DPSD model on the AIC test, the evidence to support the DPSD model was weak evidence (i.e., difference value was 2 or less). For the 7 of the 31 participants who supported the DPSD model on the BIC test, the evidence to support the DPSD model was also weak evidence, further confirming the inconclusive result.

The second model comparison that we performed was between the DPSD-FR and DPSD-FR1-FTi models. Table 14 shows the mean and standard error of the difference of the test values. The mean difference AIC value reveals that the DPSD-FR and DPSD-

FR1-FTi model cannot be distinguished from one another using Raftery's (1995) standards. Because both tests did not support one model over the other due to the inconclusive evidence from the AIC values, the DSPD-FR and DPSD-FR1-FTi models cannot be distinguished from each other. We evaluated the individual participant AIC and BIC difference scores to determine how many were consistent with the mean. For the AIC statistic, a total of 29 out of 38 participants favored the DPSD-FR1-FTi model. Of these 29 participants, 24 had some evidence (value larger than 2) to support the DPSD-FR1-FTi model over the DPSD-FR model (*M* = 5.01, *SD* = 1.42), suggesting a greater advantage for that model. For the BIC statistic, 35 out of 38 participants had strong evidence to favor the DSPD-FR1-FTi model over the DPSD-FR model, indicating an advantage of the DSPD-FR1-FTi model. The individual AIC data coupled with the individual BIC data seem to support that the inclusion of a single false recollection parameter provided a better account of the behavioral data produced by the Payne-Eakin paradigm than a false recollection parameter for each nonstudied item condition, supporting our initial hypothesis.

Power Limitation in the Individual Analysis

A potential problem with the analysis of the individual participant data is that it might not have enough response variability to allow the model to adequately fit their data, resulting in inconclusive findings. One requirement of fitting models to behavioral data is that the data needs to have responses from the entire scale to extract meaningful information from the model fitting. Unfortunately, despite our attempt to mitigate this possibility by removing participants whose responses were 90% or more 1- or 6-responses, 56% of the remaining 38 participant responses were primarily 1- and 6-

responses. Therefore, it could be that the inconclusive results obtained for the DPSD and DPSD-FR1-FTi comparison for the individual analysis was due to a lack of response variability.

To determine whether the lack of variability in responses on the 6-point confidence recognition scale contributed to the inconclusive result, we performed a follow up analysis of the data. For this comparison, we concentrated on the comparison between the DPSD and the DPSD-FR1-FTi models as it was the most critical given our prediction that FTi test pairs would be falsely recollected due to the presence of an implicitly activated false target. To evaluate this possibility, I calculated the proportion of high-confidence responses (i.e., 1- or 6-response) that a participant gave on the 6-point confidence recognition scale. For example, a score of .80 indicates that 80% of that participant's responses were either 1- or 6-responses. Next, I calculated the AIC difference score for each participant between the DPSD and DPSD-FR1-FTi models. Each participant's AIC difference score was plotted as a function of the proportion of high-confidence responses, shown in Figure 34A. Because smaller AIC values indicate a better fit, positive AIC difference values indicate an advantage of the DPSD-FR1-FTi model and negative difference scores indicate an advantage of the DPSD model.

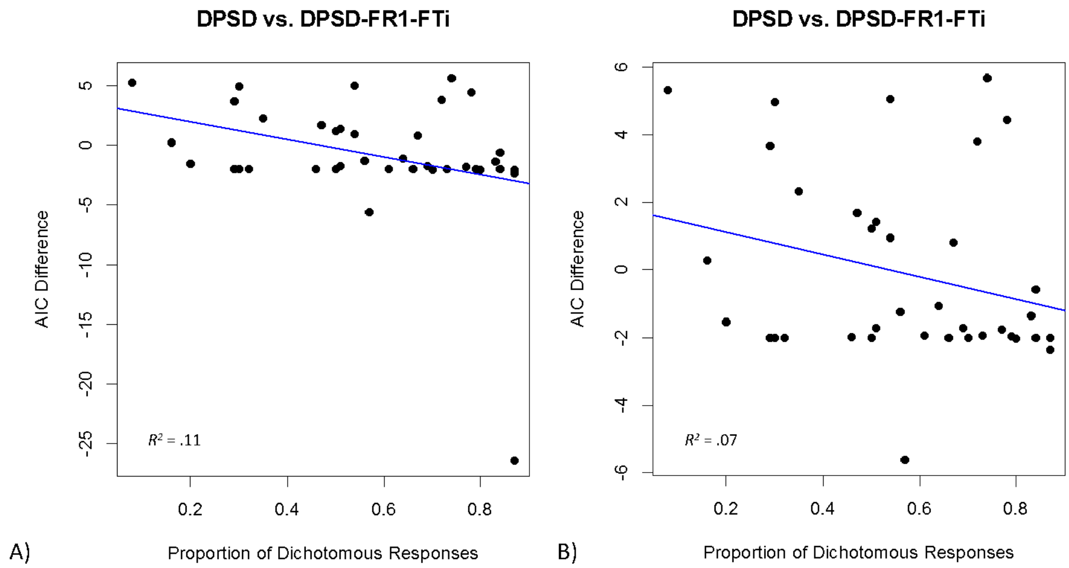


Figure 34. A) Comparison of AIC difference values between the DPSD and DPSD-FR1-FTi models plotted as a function of the proportion of dichotomous responses. B) Comparison of AIC difference values between the DPSD and DPSD-FR1-FTi models plotted as a function of the proportion of dichotomous responses without outliers.

A linear model was calculated to determine whether the probability that a participant gave a dichotomous response predicted the outcome of the DPSD or DPSD-FR1-FTi model providing a better fit. The initial analysis revealed a significant negative trend in the data, $F(1,36) = 4.22$, $p < .05$, $R^2 = .11$, demonstrating that as the proportion of dichotomous responses increased, so did the likelihood that the DPSD model provided a better fit of the data over the DPSD-FR1-FTi model. This finding indicates that the advantage of the DPSD model over the model with false recollection was driven by participants who responded primarily with high confidence ratings. However, this analysis included an outlier that could have skewed the results. Therefore, we ran the analysis again without the outlier, shown in Figure 34B. The analysis that excluded the outlier found the trend to be nonsignificant, $F(1,35) = 2.54$, $p = .12$, $R^2 = .07$, suggesting that the reduced response variability might not have contributed to the inability to distinguish between the two models.

Although the analysis that excluded the outlier did not support our hypothesis that dichotomous responses contributed to the advantageous fit of the DPSD model over the DPSD-FR1-FTi model, we suspect that the reduced response variability made it difficult to detect the presence of false recollection. Because participants gave mostly dichotomous responses (56%), their response data resulted in producing a straight line which can be difficult model. A consequence of linear data is that it becomes difficult to parse apart recollection from familiarity, resulting in the simpler model (i.e., the DPSD model) being favored. This supposition is illustrated by the results for Participant 4 and 29, shown in Figure 34, which demonstrate the impact of response variability on the ability of the false recollection models to fit the data. Specifically, Participant 29 had more response variability than Participant 4, and, as a result, the models were seemingly able to more adequately fit the data for Participant 29 than for Participant 4. Therefore, it is still likely that the reduced response variability contributed to the inconclusive results for the modeling analysis of the individual data.

Discussion

The purpose of the modeling analysis was to examine the role of recollection in the retrieval of FTi test pairs. It was predicted that FTi test pairs would be falsely recollected due to the presence of the implicitly activated false target during study. Although a behavioral analysis provided some insight into the processes impacting recognition of FTi test pairs, this type of analysis was limited in terms of allowing us to directly measure the influence of familiarity versus recollection. Therefore, we also performed a modeling analysis of the data by comparing the fit of the DPSD model to two amended models that allowed for false recollection. To amend the DPSD model, we developed two models that included a false recollection parameter. The first model that we developed was the DPSD-FR model. The DPSD-FR model included a separate false recollection parameter for each nonstudied item condition, for a total of five false recollection parameters. The purpose of this model was to determine whether false recollection occurs for all nonstudied items. The second amended model was the DPSD-FR1-FTi model which included one false recollection parameter for FTi test pairs and tested whether false recollection occurred for those pairs.

We first fit each of the models to the aggregated form of the data. One possible limitation with modeling the behavioral data was a possible lack of power due to participants not using the entire 6-point confidence recognition scale. We attempted to address this issue by removing participants who gave primarily high-confidence responses. However, to further avoid this issue, we also started with an analysis of the aggregated data to ensure adequate power. The results of the aggregate analysis revealed strong evidence that the DPSD-FR model provided the best account of the aggregated data over both the DPSD and DPSD-FR1-FTi models, suggesting that false recollection occurs for all nonstudied test pairs. This finding contradicts our predictions that false recollection would only occur for FTi test pairs and, instead, indicates that false recollection occurred for all test pairs. This finding is problematic for the CLS model, which posits that false recollection should only occur if there is a high degree of feature overlap (80%) between a test item's representation and a studied item's stored representation. Although we originally interpreted this condition of the CLS model to mean that both a studied cue and target would have to be present to match, it could be that only a small level of feature overlap based on representation of the cue is necessary to facilitate a match than was originally posited.

If the presentation of a portion of a studied pair provides enough feature overlap to allow for recollection, then we would predict that FTi test pairs would have the highest false recollection rate, followed by Re-Paired, and then the FTx test pairs. This prediction is based on the amount of studied information presented during test and is supported by the false recollection estimates provided by the DPSD-FR model. FTi test pairs would have had the highest false recollection rate due to the presence of both a studied cue and an implicitly activated target which would have had the highest level of feature overlap of the test pairs. Re-Paired test pairs would have the second highest false recollection rate due to the presence of a studied cue and target, but not overlapping features. Although presenting a studied cue and studied target from two different studied events would increase the probability that false recollection would occur, FTi test pairs would still have a higher likelihood of a match occurring due to more features overlapping for a single studied event. Finally, FTx test pairs would have the lowest false recollection rate of the experimental test pairs due to including only a studied cue. The false recollection parameter values estimated by the DPSD-FR model support each of these predictions, lending further credence to the possibility that false recollection could occur with a smaller rate of feature overlap than originally predicted by the CLS model.

Although the aggregate analysis revealed that the inclusion of a false recollection parameter for each test pair condition was necessary to fit the behavioral data, fitting the model to aggregated data can fail to account for effects due to cognitive processing at the individual level. Response patterns for individuals can occur in a wide variety of patterns that are qualitatively different from patterns detected in the aggregated data (Pratte et al., 2010). The impact of qualitative differences between aggregate and individual level data on modeling can result in masking the underlying cognitive process. This possibility was evaluated in Figure 32 which showed how different response patterns can be between individual and aggregate data. Therefore, we also performed the modeling analysis on the individual participant data.

The DPSD, DPSD-FR, and DPSD-FR1-FTi models were fit to each of the individual participant’s data. The results of the individual analysis revealed inconclusive evidence as to whether the DPSD or the DPSD-FR1-FTi model provided the best account of the data. This inconclusive result suggests that the behavioral data might not have had enough response variability to detect false recollection. This supposition was supported by the follow up analysis that investigated the relationship between the AIC difference score for the DSPD and DSPD-FR1-FTi models and the proportion of dichotomous responses in the data set. The follow up analysis revealed a trend that as the proportion of dichotomous responses increased in a particular data set, the likelihood that the DSPD model was favored also increased. This finding demonstrated that as the proportion of dichotomous responses increased, the likelihood that false recollection was detected was reduced, suggesting the deleterious impact of the reduced response variability at the individual level. Although this result does not provide a definitive determination that false recollection did not occur, it does demonstrate that individual data with more response variability is needed to determine whether the false recognition effect obtained by the Payne-Eakin paradigm was due to familiarity or recollection.

Because the analysis of the individual data yielded inconclusive results due to limited response variability, it could be that the aggregate form of the data provides a more accurate depiction of the recognition processes that occurred. According to Pratte et al. (2010), aggregated data should not be used to determine the effect of a cognitive process effects due to the possibility that participant variance can interfere with those effects. It could be that for our behavioral data, the aggregate data provides the most accurate depiction of the recognition processes that occurred because it provides enough power for these effects to be observed. Because the individual data had such low response variability, detecting the differential influence of familiarity and recollection proved impossible, but aggregating the data provided enough power that strong evidence to suggest that the DPSD-FR model provided the best account of the data over the DPSD and DPSD-FR1-FTi models. Therefore, it could be that false recollection did occur for each of the nonstudied test pairs, but the individual data did not have the power to detect it. This supposition is supported by the DPSD-FR model estimates from the individual data. According to the DPSD-FR model, each of the test pairs was estimated to have some level of familiarity and false recollection, with FTi pairs having the highest rate. Although the control test pairs were estimated to have significantly lower rates of false recollection than FTi, FTx, and Re-Paired pairs, it provides support that false recollection might have occurred for all test pair types but that our lack of response variability hindered the model from providing an adequate fit.

Another possibility is that the decision to exclude participants with 90% or more dichotomous responses contributed to our inconclusive findings. This decision was originally deemed necessary due to the difficulty of extracting meaningful information from fitting models to data missing responses from the 6-point confidence recognition scale. However, a key experimental condition, FTi test pairs, actually predicted dichotomous confidence judgements. In addition, with a high hit rate for studied test pairs and a low false alarm rate for control test pairs, these conditions would also be predicted to produce non-variable responses of all 6-responses or all 1-responses, respectively. Also, as previously discussed, with such a small number of test pairs, participants might not have had time to learn to use the relatively complex 6-point scale. Therefore, it could be that removing those participants actually eliminated our ability to detect false recollection, especially in the experimental conditions that predicted that response distribution. Although recollection was predicted to occur most frequently for studied and FTi test pairs, as the DPSD-FR model demonstrates for the aggregated data, false recollection could have occurred for all of the experimental conditions, including for FTx and Re-Paired test pairs. Therefore, because recollection is theorized to be an all-or-none process that impacts high-confidence responses, removing participants with a high proportion of those responses could have resulted in participants whose judgments were primarily influenced by recollection being removed. Despite this possibility, recollection was predicted only to occur for studied and FTi pairs which consisted of 43% of responses. As a result, we would have predicted that approximately 43% of responses should have been primarily dichotomous. Therefore, our decision to remove participants with 90% or more dichotomous responses would have only resulted in the elimination of people who would have relied on recollection for all word pair types.

CHAPTER VIII

GENERAL DISCUSSION

The main goal of this dissertation was to determine whether the false recognition effect obtained by the Payne-Eakin paradigm was due solely to familiarity, as predicted by the DPSD model, or a combination of familiarity and recollection, as predicted by the CLS model. False recollection has been found to occur under a variety of experimental conditions (e.g., Higham & Vokey, 2004; Jones, 2013; Lane & Zaragoza, 1995; Payne & Eakin, 2017; Roediger & McDermott, 1995). Although evidence for false recollection has been well-documented in the false memory literature, many of the current dominant accounts of recognition, such as Yonelinas's (1994, 2002) DPSD model, do not contain a mechanism for the retrieval of contextual information for nonstudied items. Instead, the DPSD model attributes any false recognition of nonstudied items to familiarity. Yonelinas et al. (2010) addressed the issue of false recollection by characterizing it as a violation of a boundary condition in the DPSD model. According to Yonelinas et al. (2010), boundary conditions are conditions that are typical for recognition and, as such, are within the scope of the model. Yonelinas et al. (2010) cites conditions such as false recollection as atypical, non-systematic, and occurring at rates as low as 0-5% (but see Higham & Vokey, 2004); therefore, false recollection is a violation of those boundary conditions. Although Yonelinas et al. (2010) acknowledges the possibility for false recollection, he states that amending the model to account for them is unwarranted due to it being such a small, random effect. Yonelinas et al. (2010) does, however, acknowledge that if false recollection effects could be increased so that the nature of the effects could be fully analyzed, then an amendment to the model should be considered. This dissertation sought to increase the rate of false recollection by utilizing the Payne-Eakin paradigm, which was based on manipulations theorized to impact recollection.

PIER2 and CLS Models

According to assumptions based on the PIER2 and CLS models, any associate of both a cue and target explicitly presented at study is also implicitly activated, resulting in the associate being part of the representation of the studied cue-target pair. The studied representation is then stored as a sparse representation of shared, overlapping features in the hippocampus that contains a limited number of unique features from the studied event, including features of the implicit associate. As a result, when the implicitly activated associate is presented as the false target in a test pair that also contains a studied cue, a match between the test item's representation and the studied representation in the hippocampus is detected. This match then results in the retrieval of context from the studied event, which is critical for recollection, and results in false recognition due to recollection.

The results from this dissertation provide some evidence to support the predictions of the PIER2 and CLS models. Test pairs containing an implicit associate as a false target (i.e., FTi test pairs) were significantly more likely to be falsely recognized, as well as more likely to be falsely recognized with high confidence, than test pairs without an implicitly associated target (i.e., FTx test pairs). Although these findings support the predictions of both models, identifying the underlying recognition process responsible for them proved challenging. To identify the recognition process responsible for the false recognition for each test pair type, we analyzed the response distributions for each of the word pairs by examining histograms and ROC curves. According to Elfman et al. (2008), response distributions can provide some insight as to the underlying process, with familiarity producing Gaussian-shaped distributions and recollection producing dichotomous, linear distributions. Although, we predicted that FTi test pairs would have a dichotomous distribution because their false recognition was predicted to be due to recollection, the analysis revealed no obvious differences in the response distributions between the FTi and FTx test pairs. This finding suggests that the same cognitive processes were responsible for false recognition of both test pairs.

One explanation as to why the response distributions for FTi and FTx test pairs did appear different, which would be supported by the CLS model, is that most of the FTi test pairs were recognized via familiarity and only a small number of them were recognized via recollection. For FTi test pairs that were rejected by the hippocampus due to a mismatch between the test item's representation and the stored representation, because FTi pairs also included a studied cue, those pairs could have been falsely recognized due to familiarity stemming from activation in the MTLC. This possibility was not originally predicted because the explanation as to how familiarity and recollection work together in the CLS model is not clearly defined. Therefore, our understanding was that FTi test pairs that were not recollected in the hippocampus would be rejected and recognized as new. However, it is possible that, because FTi test pairs were identical in every other way to the FTx pairs (i.e., both test pairs had a repeated, studied cue and a semantically associated target), familiarity also could have produced false recognition for some of those pairs. This explanation would require an adjustment to the CLS model to allow for familiarity processes to occur when the hippocampus rejects recollection. False recognition could have been higher for FTi than FTx pairs, despite their similar familiarity, because of the additional potential for recollection of some of the FTi test pairs due to the prior implicit activation of the false target for FTi, but not FTx test pairs. This combination of recollection plus familiarity could have produced the higher proportion of false recognition for the FTi over the FTx test pairs.

Alternatively, as the lack of difference between the response distributions for the FTi versus the FTx test pairs suggests, a single recognition process—familiarity—could have led to the higher rate of false recognition for FTi pairs over FTx pairs. It could be that the presentation of the implicitly activated false target strengthened the familiarity signal for those pairs over the FTx test pairs. Although we predicted that the re-presentation of an implicitly activated false target during study would increase the probability of a match occurring in the hippocampus between the test and studied representations, allowing for false recollection to occur, it could be that the implicitly activated target increased familiarity with those test pairs instead. This alternative explanation is supported by the finding that Re-Paired test pairs, predicted to be most familiar because both the cue and target for those pairs had been studied, garnered the same proportion of false recognition and 6-responses as the FTi test pairs. A reasonable explanation would be that FTi and Re-Paired test pairs were equally familiar; the FTi test pairs because the implicit activation of the test target resulted in the FTi test pairs effectively being a repetition of both a studied cue and (implicitly) studied target and the

Re-Paired test pairs being an explicit repetition of both a studied cue and (explicitly) studied target.

The CLS model states that the impact of recognizing a studied item due to familiarity results from a process known as sharpening. According to the model, sharpening occurs as items are encountered over time. During an item's first encounter, a large number of nodes are activated; some of these nodes are relevant and some are not. Over time, this sharpening process constricts activation to only items that were studied, resulting in only items that have been sharpened being recognized. Therefore, we predicted that the model would argue against the idea that the implicitly activated target initiating the sharpening process required for recognition due to activation in the MTLC. However, it could be that the CLS model did not consider the role of implicitly activated information in the sharpening process and that the implicitly activated material is included in the sharpened nodes. This finding requires further exploration to perhaps inform and expand the assumptions of the CLS model to allow for the influence of implicitly activated information at study on familiarity and be added as part of the constrained activation in the MTLC.

Finally, another explanation for why the response distributions did not differ for FTi, FTx, and Re-Paired test pairs is that a combination of recollection and familiarity occurred for each of them. Although we predicted that false recollection would only occur for FTi test pairs based on assumptions from the CLS model, it could be that recollection also occurred for the other test pairs. This hypothesis was supported by the findings in the modeling analysis of the aggregated data which demonstrated strong evidence for the false recollection of all nonstudied items. Further, the analysis of the individual data also revealed estimates of false recollection for each of the nonstudied pairs. According to the CLS model, the purpose of the sparse representation is to create a higher degree of specificity in the level of feature overlap to allow for a match and can only be fulfilled by the presentation of the studied cue and target. Based on this assumption, we predicted that only test pairs that contained a studied cue and target – either implicitly or explicitly – could accomplish this level of feature specificity. However, despite the CLS model's assumption that both a studied cue and target would be necessary to facilitate a match between the sparse study representation and the test representation in the hippocampus, it could be that the degree of feature overlap necessary for a match could be accomplished with the presentation of a single studied item. It could be that the presentation of a studied cue provided enough feature overlap in the hippocampus to allow for a match between the test item representation and the sparse representation.

This explanation is supported by several findings in this dissertation. The first finding to provide support is that FTx test pairs received significantly more 6-responses than control test pairs but significantly fewer 6-responses than FTi test pairs. If the presentation of a studied cue provided enough feature overlap for a match, then those items would have a higher likelihood of receiving a high-confidence old response. However, because the FTx pairs do not share the same level of feature overlap as a test pair that contains a studied cue and an implicitly activated false target, they would not have been recognized at the same rate as FTi test pairs. In the same vein, the finding that Re-Paired test pairs received the same proportion of false recognition and 6-responses as the FTi test pairs suggests that those pairs were equally likely to be recognized. Because the Re-Paired test pairs presented a studied cue and a studied target, the presentation of either one could have allowed for a large enough feature overlap for a match to occur in the hippocampus, resulting in an increased likelihood of some of those pairs being recollected. Finally, the last finding to support this explanation is that the DPSD-FR model provided the best account of the aggregate data and estimated that false recollection occurred for approximately 20% of the data. In the DPSD-FR model, each nonstudied test-pair condition was allowed to have a unique rate of false recollection. The results of the modeling analysis of the aggregated data demonstrated that Re-Paired and FTi test pairs were the most likely to be falsely recollected, followed by FTx test pairs. This account of the data by the DPSD-FR model was compared to the DPSD model which only allowed for familiarity and was found to provide a better fit of the data. Although this explanation requires further exploration, it could inform the CLS model to allow for a portion of re-presented studied event to match, resulting in recollection.

DPSD Model

Although the findings that tested the assumptions of the CLS model leave some questions, the manipulations for each of the test pair types did serve one goal: to increase the number of false recognitions that received 6-responses on the 6-point recognition confidence scale. This goal was set to address Yonelinas et al.'s (2010) call for a paradigm that produced more than a minimal, error-based number of false recollections to determine whether model corrections should be made to account for false recollection. According to Yonelinas et al. (2010), false recollection was deemed an inconsequential effect in recognition data due to the low rate at which it occurs, typically less than 5%. As a result, Yonelinas et al. (2010) goes on to state that because it is such a small effect, amending the DPSD model to include a parameter to account for them is unnecessary, unless the effect can be increased to a level that can be evaluated. The results of this dissertation demonstrated that a false memory effect, based on theoretical manipulations assumed to impact false recollection, could occur at a large enough rate to be evaluated, evidenced by the modeling analysis.

Because the false memory effect in this dissertation was large enough to be evaluated, several of the findings of that investigation provided support for the DPSD model. One finding that supported the DPSD model was that the nature of the response distributions for each of the test pair types were not significantly different from one another. According to Elfman et al. (2008), response distributions resulting from familiarity versus recollection should produce two distinguishable shapes; Gaussian for familiarity and dichotomous for recollection. Therefore, we predicted that test pairs containing an implicit associate as a false target (i.e., FTi test pairs) would be more dichotomous than test pairs without an implicitly associated target (i.e., FTx and Re-Paired test pairs), demonstrating the impact of recollection. This prediction was not supported; all three types of test pairs appeared to have similar response distributions, suggesting either a single cognitive process was responsible for the false recognition or the same cognitive processes were responsible for the recognition of each test pair type. This finding is consistent with predictions from the DPSD model (Yonelinas, 1994, 2002) which assumes that familiarity is the only process that can contribute to the recognition of nonstudied items. Although the response distributions for the types of word pairs did not differ significantly and provides support for Yonelinas’s (2002) explanation that all false recognition occurs due to familiarity, it does not dissuade the limitation of a familiarity explanation that it is difficult to pre-experimentally predict. Because Yonelinas (2002) does not provide concrete characteristics that impact familiarity, predicting what will and will not be recognized due to familiarity is difficult. Instead, only a post-hoc explanation can determine which items were familiar enough to be recognized and which ones were not.

An alternative explanation that was offered for the lack of difference in response distributions for the different word pair types was that despite our manipulations to increase the false recollection rate, it might have occurred at too small a rate to be detected by a behavioral analysis. Although the Payne-Eakin paradigm was successful at generating a larger number of false memory items than other false memory paradigms (e.g., Higham & Vokey, 2004; Jones, 2013; Lane & Zaragoza, 1995; Roediger & McDermott, 1995), it could be that the paradigm still did not produce enough items necessary for a full evaluation of false recollection. It could be that because false recollection is such a small effect, it could not cause a large enough difference to overall response distributions to be detected by standard behavioral analyses. Although this explanation is inconsistent with the predictions of the DPSD model, which would predict that familiarity is only process responsible for false memories, it is consistent with Yonelinas et al.'s (2010) explanation that false recollection is too small to warrant evaluation and do not warrant an amendment to the model. The findings from our modeling analysis also support his argument.

The results of the individual modeling analysis could not provide strong evidence to suggest the benefit of a false recollection parameter in the DSPD model, suggesting that familiarity alone might be enough to account for the false memory effects obtained in the Payne-Eakin paradigm. We predicted that the inclusion of a false recollection parameter would be necessary to fit the behavioral data due to the influence of false recollection on FTi test pairs. However, the modeling analysis of the individual data, which is argued to be the most accurate evaluation of a cognitive process, could not determine whether a false recollection parameter was necessary to account for the data for the majority of participants. However, when a false recollection parameter was allowed for each test pair type, the model did provide a robust estimate of false recollection, suggesting that false recollection did occur but could not be properly measured with the given data set. It was suspected that low variability in responses caused this inconclusive result and as supported by a follow up analysis that revealed as participant response variability decreased, the likelihood that false recollection was detected also decreased. Although we predict that if response variability is increased, the inclusion of a false recollection parameter will be necessary to fit the data; however, it could be that an analysis of data with increased response variability will favor the DPSD model. We do not consider this to be a strong possibility due to the result of the modeling analysis of the aggregate data which revealed strong evidence that false recollection was necessary to fit the data. However, it could be that the aggregate analysis does not accurately reflect the recollection effect in the data due to the influence of unaccounted for participant variability (Pratte et al., 2010).

Although the findings of this dissertation provide some support for the DPSD model and Yonelinas et al.'s (2010) argument that false recollection is too small to warrant an amendment to the model, these findings also add to the evidence that false recollection can occur, even if at low rates (e.g., Higham & Vokey, 2004; Jones, 2013;

Lane & Zaragoza, 1995; Roediger & McDermott, 1995). Despite false recollection occurring at low rates, the finding that it occurs at all warrants an examination of our current models of recognition. Because current models such as the DPSD model do not allow for the contribution of false recollection in recognizing nonstudied items, our current understanding of how recognition works is incomplete. The impact of this incompleteness can hinder our ability to make accurate predictions regarding recognition. Specifically, despite false recognition occurring less than 5% of the time, because recognition models do not contain a mechanism to allow for it, we cannot predict those occurrences at all. As a result, the models could be discredited due to their limitation in predicting even the smallest aspects of recognition.

Limitations

There are some limitations of this dissertation that should be kept in mind. One design flaw is that we failed to include an unrelated studied test pair condition. Because recombining studied cues and targets for the Re-Paired test pair condition produced unrelated word pairs, a new unrelated control condition was included to provide a baseline measure of the degree to which nonstudied unrelated pairs could be recognized. However, due to an oversight, we failed to include a studied test condition to measure the degree to which studied unrelated test pairs could be recognized. As a result, the degree to which we can make conclusions about the nature of the Re-Paired test pairs is limited. All comparisons of the false alarm rates for Re-Paired tests pairs were made to the hit rate for related studied test pairs. It could also be that unrelated test pairs are inherently more difficult to discriminate from studied unrelated pairs which would have resulted in a reduced hit rate for those pairs. This supposition is supported by the finding that R-control (*M* = .17, *SE* = .02) test pairs were more likely to be falsely identified as old than U-control (*M* = .07, *SE* = .01) test pairs, $t(54) = 5.32$, $p < .001$, demonstrating that they were more difficult to discriminate. A reduced hit rate could have then resulted in the Re-Paired test pairs having a reduced *d'* value from what is estimated for the hit rate for related studied test pairs. A consequence of being able to accurately calculate *d'* could be that differences between the Re-Paired and FTi test pairs could have been detected. Specifically, if *d'* for Re-Paired pairs was calculated based on an unrelated hit rate, rather than a related hit rate, then the value could have been reduced. This reduced value could then have been significantly different from the *d'* value for FTi test pairs which would have provided a more accurate depiction of the relationship between the two types of test pairs. It also could be that different criterion threshold values (i.e., more conservative or more liberal) were used for unrelated than related test pairs. This difference in criterion could have skewed the degree to which Re-Paired word pairs were recognized, affecting both the hit and false alarm rates; however, criterion cannot be calculated without the unrelated studied test condition. Therefore, this hypothesis cannot be fully tested using the current set of materials.

A second limitation of this dissertation experiment was that the lack of variability in responses. In the pilot data for this dissertation, we attempted to increase the variability of responses by increasing the time and complexity of the distractor task. We further attempted to increase the variability by increasing the overall number of stimuli for the dissertation experiment over the pilot studies. Although the behavioral data did show an increase in the variability of responses for some participants, others continued to only use the outside responses. One reason for this result could be that the task was too easy.

Overall, we obtained a high hit rate, demonstrating that people were good at identifying studied test pairs. Typically, recognition studies use hundreds of stimuli as a means of reducing ceiling effects and increasing response variability. However, this was not a viable option for this study because of the experimental constraints on test-pair construction. Loosening these constraints could have blurred the lines between the different types of test pairs, reducing measurement sensitivity.

Related to the fewer number of stimuli is another potential problem that could have contributed to the low variability in the response distributions, that is the use of the 6-point confidence recognition scale. Overall, 69% of participants' responses were dichotomous (either a 1- or 6-response), suggesting that they were not using the scale as was intended. Using this scale is not intuitive or seemingly associated with the recognition process itself. Typically, recognition is characterized as a dichotomous result; an item is either recognized as old or rejected as new. However, the 6-point confidence recognition scale requires participants to break down this dichotomous decision into degrees of confidence, which could result in the task being too difficult or confusing. In addition, the scale itself has unusual anchors for a typical Likert-type scale which measures varying degrees of a single factor (e.g., confidence on a scale of 1 (very low confidence) – 7 (very high confidence). The 6-point confidence recognition scale measured two factors—newness and oldness—using the same scale with 1 (very high confidence new) – 3 (very low confidence new) and 6 (very high confidence old) – 4 (very low confidence old). Certainly, other studies have shown that participants can learn to use this scale, but most of those studies use huge numbers of test trials (e.g., Pratte & Rouder, 2011; Pratte et al., 2010; Yonelinas, 1994). Having so much practice with the scale could result in implicit learning to convert their dichotomous decisions into degrees of confidence and splitting the range of the scale to report their confidence. With only 168 test trials that were further reduced into 5 test-pair conditions, there was little time for implicit learning to take place. The result was low variability in responses, with people maintaining the dichotomous use of the scale on most of the trials.

This low variability in participant responses contributed to issues in the modeling analysis. The modeling analysis of the individual participant data returned inconclusive evidence to support the DPSD over the DPSD-FR1-FTi model. The cause of this inconclusive finding was determined to be the result of low variability in participant responses on the 6-point confidence recognition scale through a follow up analysis. The follow up analysis investigated the relationship between the proportion of dichotomous responses to the likelihood that one model was favored over the other. The analysis revealed that as the proportion of dichotomous responses increased, the likelihood that false recollection was detected was decreased, demonstrating the deleterious impact of response variability on the analysis. The hypothesis that the inconclusive results of the individual data is supported by the results of aggregate data which revealed strong evidence of false recollection. Although we predict that increased response variability at the individual level will result in the detection of false recollection, the possibility that a model containing a false recollection parameter will provide a better account of the data than the DSPD model (Yonelinas, 1994, 2002) still needs further investigation.

Future Directions

Beyond correcting some of the methodological problems stated previously, such as including a studied unrelated test pair condition and finding a way to create more test items for all five types of test pairs to create more response variability, one future direction for this line of research is to investigate the role of context in familiarity and recollection. One future study that we would like to explore is to manipulate font using the Payne-Eakin paradigm. Specifically, we would like to manipulate the color of the font as well as the placement on the screen to push the limits of recollection.

Another future direction for this line of research is to investigate the role of implicit associations on recognition. In this dissertation, we found evidence to support the hypothesis that the implicit activation of the false-target in FTi test pairs increased recognition. However, it is unclear as to whether this increased recognition is due to an increased familiarity signal, recollection, or a combination of both. Therefore, we would like to investigate the role of implicit associations by manipulating the strength of the associate and further evaluating the subsequent and differential impact of familiarity and recollection.

Conclusion

The findings of this dissertation constitute a step toward a more solid understanding of recognition of nonstudied items. This dissertation demonstrates that modeling false recollection is possible. Previously, it had been suggested that modeling false recollection was not possible due to the low rates at which it typically occurs in recognition studies (Yonelinas et al., 2010). However, the current dissertation demonstrates that, using the Payne-Eakin paradigm produced a large enough number of false recognitions to examine using modeling methods. As a result, this dissertation provides an examination of current models of recognition and the impact of false recollection on those models. The results of this dissertation suggest that, because current models of recognition do not provide a mechanism to account for false recollection, our understanding of recognition is incorrect. Specifically, this dissertation highlights that the current understanding of how false recollection contributes to recognition performance is poorly understood and is an area in need of further development.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723.

Atkinson, R. C., & Juola, J. F. (1973). Factors influencing speed and accuracy of word recognition. *Attention and Performance IV*, 583-612.

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*(1), 158-173.

Barba, G. D. (1993). Confabulation: Knowledge and recollective experience. *Cognitive Neuropsychology*, *10*(1), 1-20.

Brainerd, C. J., Aydin, C., & Reyna, V. F. (2012). Development of dual-retrieval processes in recall: Learning, forgetting, and reminiscence. *Journal of Memory and Language*, *66*(4), 763-788.

Brainerd, C. J., Wright, R., Reyna, V. F., & Mojardin, A. H. (2001). Conjoint recognition and phantom recollection. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(2), 307-327.

Broadbent, N. J., Gaskin, S., Squire, L. R., & Clark, R. E. (2010). Object recognition memory and the rodent hippocampus. *Learning & Memory*, *17*(1), 5-11.

Broadbent, N. J., Squire, L. R., & Clark, R. E. (2004). Spatial memory, recognition memory, and the hippocampus. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(40), 14515-14520.

Clark, R. E., Zola, S. M., & Squire, L. R. (2000). Impaired recognition memory in rats after damage to the hippocampus. *Journal of Neuroscience*, *20*(23), 8853-8860.

Dennis, N. A., Bowman, C. R., & Peterson, K. M. (2014). Age-related differences in the neural correlates mediating false recollection. *Neurobiology of Aging*, *35*(2), 395-407.

Dewhurst, S. A. (2001). Category repetition and false recognition: Effects of instance frequency and category size. *Journal of Memory and Language*, *44*(1), 153-167.

Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2008). The effects of unitization on familiarity-based source memory: testing a behavioral prediction derived from neuroimaging data. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 730-740.

Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2007). Imaging recollection and familiarity in the medial temporal lobe: a three-component model. *Trends in Cognitive Sciences*, *11*(9), 379-386.

Eakin, D. K. (2010). ListChecker Pro 1.2: A program designed to facilitate creating word lists using the University of South Florida word association norms. *Behavior Research Methods*, *42*(4), 1012-1021.

Eakin, D. K., Schreiber, T. A., & Sergent-Marshall, S. (2003). Misinformation effects in eyewitness memory: The presence and absence of memory impairment as a function of warning and misinformation accessibility. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(5), 813-825.

Egan, J. P. (1975). *Signal detection theory and {ROC} analysis SE - Series in Cognition and Perception*. New York, NY: Academic Press. Retrieved from citeulike-article-id:931676

Elfman, K. W., Parks, C. M., & Yonelinas, A. P. (2008). Testing a neurocomputational model of recollection, familiarity, and source recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 752-768.

Elfman, K. W., & Yonelinas, A. P. (2015). Recollection and familiarity exhibit dissociable similarity gradients: a test of the complementary learning systems model. *Journal of Cognitive Neuroscience*, *27*(5), 876-892.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*(2), 211-245.

Fandakova, Y., Shing, Y. L., & Lindenberger, U. (2013). Differences in binding and monitoring mechanisms contribute to lifespan age differences in false memory. *Developmental Psychology*, *49*(10), 1822-1832.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175-191.

Gallo, D. A., Foster, K. T., & Johnson, E. L. (2009). Elevated false recollection of emotional pictures in young and older adults. *Psychology and Aging*, *24*(4), 981-988.

Gallo, D. A., & Roediger, H. L. (2003). The effects of associations and aging on illusory recollection. *Memory & Cognition*, *31*(7), 1036-1044.

Gardiner, J. M., Ramponi, C., & Richardson-Klavehn, A. (1998). Experiences of remembering, knowing, and guessing. *Consciousness and Cognition*, *7*(1), 1-26.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *91*(1), 1-67.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Heathcote, A. (2003). Item recognition memory and the receiver operating characteristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(6), 1210-1230.

Heathcote, A., Raymond, F., & Dunn, J. (2006). Recollection and familiarity in recognition memory: Evidence from ROC curves. *Journal of Memory and Language*, *55*(4), 495-514.

Hicks, J. L., & Marsh, R. L. (1999). Remember-know judgments can depend on how memory is tested. *Psychonomic Bulletin & Review*, *6*(1), 117-122.

Higham, P. A. (1998). Believing details known to have been suggested. *British Journal of Psychology*, *89*(2), 265-283.

Higham, P. A., & Vokey, J. R. (2004). Illusory recollection and dual-process models of recognition memory. *Quarterly Journal of Experimental Psychology Section A*, *57*(4), 714-744.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*(5), 513-541.

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

Jones, A. E. (2013). Do photographs produce false memories? (Unpublished master's thesis). Mississippi State University, Mississippi State, MS.

Kim, J., & Yassa, M. A. (2013). Assessing recollection and familiarity of similar lures in a behavioral pattern separation task. *Hippocampus*, *23*(4), 287-294.

Kintsch, W. (1970). Models for free recall and recognition. *Models of Human Memory*, 331-373.

Lampinen, J. M., Watkins, K. N., & Odegard, T. N. (2006). Phantom ROC: Recollection rejection in a hybrid conjoint recognition signal detection model. *Memory*, *14*(6), 655-671.

Lane, S. M., & Zaragoza, M. S. (1995). The recollective experience of cross-modality confusion errors. *Memory & Cognition*, *23*(5), 607-610.

Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. Psychology press.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, *87*(3), 252-271.

Mandler, G., & Boeck, W. J. (1974). Retrieval processes in recognition. *Memory & Cognition*, *2*(4), 613-615.

Mather, M., Henkel, L. A., & Johnson, M. K. (1997). Evaluating characteristics of false memories: Remember/know judgments and memory characteristics questionnaire compared. *Memory & Cognition*, *25*(6), 826-837.

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.

Medina, J. J. (2008). The biology of recognition memory. *Psychiatric Times*, *25*(7), 13-15.

Mickes, L., Wais, P. E., & Wixted, J. T. (2009). Recollection is a continuous process: Implications for dual-process theories of recognition memory. *Psychological Science*, *20*(4), 509-515.

Mickes, L., Wixted, J. T., & Wais, P. E. (2007). A direct test of the unequal-variance signal detection model of recognition memory. *Psychonomic Bulletin & Review*, *14*(5), 858-865.

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, *89*(6), 609.

Murdock, B. B., & Dufty, P. O. (1972). Strength theory and recognition memory. *Journal of Experimental Psychology*, *94*(3), 284–290. https://doi.org/10.1037/h0032795

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, *47*(1), 90-100.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 402-407.

Nelson, D. L., McKinney, V. M., Gee, N. R., & Janczura, G. A. (1998). Interpreting the influence of implicitly activated memories on recall and recognition. *Psychological Review*, *105*(2), 299-324.

Nelson, D. L., Zhang, N., & McKinney, V. M. (2001). The ties that bind what is known to the recognition of what is new. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(5), 1147-1159.

Norman, K. A., & O'reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychological Review*, *110*(4), 611-646.

Norman, K. A. (2010). How hippocampus and cortex contribute to recognition memory: revisiting the complementary learning systems model. *Hippocampus*, *20*(11), 1217-1227.

Norman, K. A., & Schacter, D. L. (1996). Implicit memory, explicit memory, and false recollection: A cognitive neuroscience perspective. In L. M. Reder (Ed.), *Implicit memory and metacognition*. Hillsdale, NJ: Erlbaum Associates.

O'reilly, R. C., & Rudy, J. W. (2001). Conjunctive representations in learning and memory: principles of cortical and hippocampal function. *Psychological Review*, *108*(2), 311.

Payne, A. E., & Eakin, D. K. (2014). *Measuring the role of metamemory during the production of false memories.* Unpublished Manuscript, Department of Psychology, Mississippi State University, Mississippi State, MS.

Payne, A. E., & Eakin, D. K. (2017). *The impact of implicit associations on false recognition: A new paradigm for measuring false memories.* Unpublished Manuscript, Department of Psychology, Mississippi State University, Mississippi State, MS.

Payne, D. G., Elie, C. J., Blackwell, J. M., & Neuschatz, J. S. (1996). Memory illusions: Recalling, recognizing, and recollecting events that never occurred. *Journal of Memory and Language*, *35*(2), 261-285.

Peterson, H. W., & Birdsall, T. G. (1953). *The theory of signal detectability. University of Michigan: Electronic Defense Group* (No. 13). Technical Report.

Peterson, W. W. T. G., Birdsall, T., & Fox, W. (1954). The theory of signal detectability. *Transactions of the IRE Professional Group on Information Theory*, *4*(4), 171-212.

Pratte, M. S., & Rouder, J. N. (2011). Hierarchical single-and dual-process models of recognition memory. *Journal of Mathematical Psychology*, *55*(1), 36-46.

Pratte, M. S., Rouder, J. N., & Morey, R. D. (2010). Separating mnemonic process from participant and item effects in the assessment of ROC asymmetries. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(1), 224-232.

Raftery, A. E. (1995). Bayesian Model Selection in Social Research. *Sociological Methodology*, *25*, 111–163. https://doi.org/10.2307/271063

Roediger III, H. L. (1996). Memory illusions. *Journal of Memory and Language*, *35*(2), 76-100.

Roediger III, H. L., McDermott, K. B., & Robinson, K. J. (1998). The role of associative processes in creating false memories. *Theories of Memory II*, 187-245.

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*(4), 803.

Rotello, C. M., & Heit, E. (2000). Associative recognition: A case of recall-to-reject processing. *Memory & Cognition*, *28*(6), 907-922.

Schacter, D. L., Koutstaal, W., Johnson, M. K., Gross, M. S., & Angell, K. E. (1997). False recollection induced by photographs: a comparison of older and younger adults. *Psychology and Aging*, *12*(2), 203-215.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461-464.

Slotnick, S. D., & Dodson, C. S. (2005). Support for a continuous (single-process) model of recognition memory and source memory. *Memory & Cognition*, *33*(1), 151-170.

Stark, C. E., Bayley, P. J., & Squire, L. R. (2002). Recognition memory for single items and for associations is similarly impaired following damage to the hippocampal region. *Learning & Memory*, *9*(5), 238-242.

Strange, D., Garry, M., Bernstein, D. M., & Lindsay, D. S. (2011). Photographs cause false memories for the news. *Acta Psychologica*, *136*(1), 90-94.

Tulving, E. (1983). Ecphoric processes in episodic memory. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *302*(1110), 361–371. Retrieved from http://rstb.royalsocietypublishing.org/content/302/1110/361.abstract

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498-505.

Wais, P. E., Mickes, L., & Wixted, J. T. (2008). Remember/know judgments probe degrees of recollection. *Journal of Cognitive Neuroscience, 20*(3), 400-405.

Whittlesea, B. W. (2002). False memory and the discrepancy-attribution hypothesis: The prototype-familiarity illusion. *Journal of Experimental Psychology: General*, *131*(1), 96-115.

Wickelgren, W. A., & Norman, D. A. (1966). Strength models and serial position in short-term recognition memory. *Journal of Mathematical Psychology*, *3*(2), 316-347.

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, *114*(1), 152-176.

Wixted, J. T., & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review*, *117*(4), 10251054.

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(6), 1341.

Yonelinas, A. P. (1999). The contribution of recollection and familiarity to recognition and source-memory judgments: A formal dual-process model and an analysis of receiver operating characteristics. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(6), 1415.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*(3), 441-517.

Yonelinas, A. P., Aly, M., Wang, W. C., & Koen, J. D. (2010). Recollection and familiarity: Examining controversial assumptions and new directions. *Hippocampus*, *20*(11), 1178-1194.

Yonelinas, A. P., & Jacoby, L. L. (1996). Noncriterial recollection: Familiarity as automatic, irrelevant recollection. *Consciousness and Cognition*, *5*(1-2), 131-141.

APPENDIX A

WORD PAIR INFORMATION

Mean Word Lists Characteristics

Word Pair Type:	CSS:	Cue Freq:	Cue Concreteness:	TSS:	Target Freq:	Target Concreteness:	FAS:	BAS:	Shared Associates:	Mediators:
Studied	14.97	53.74	4.40	14.67	58.45	4.35	0.08	0.06	3.02	1.54
FTi	14.97	53.74	4.40	14.89	81.66	4.54	0.09	0.05	2.47	2.29
FTx – T1	14.56	112.78	4.09	12.73	81.81	4.05	0.08	0.01	.72	.92
FTx – T2	14.56	112.78	4.09	15.19	85.71	4.18	.13	.04	.77	1.03
Re-Paired	11.27	36.73	4.49	-	-	-	-	-	-	-
R-control	16.15	32.05	3.75	16.95	53.55	3.57	0.08	0.03	1.4	2.8
U-control	11.27	36.73	4.49	-	-	-	-	-	-	-

Figure A1. Summary of mean word list characteristics.

APPENDIX B

COUNTERBALANCE SCHEME

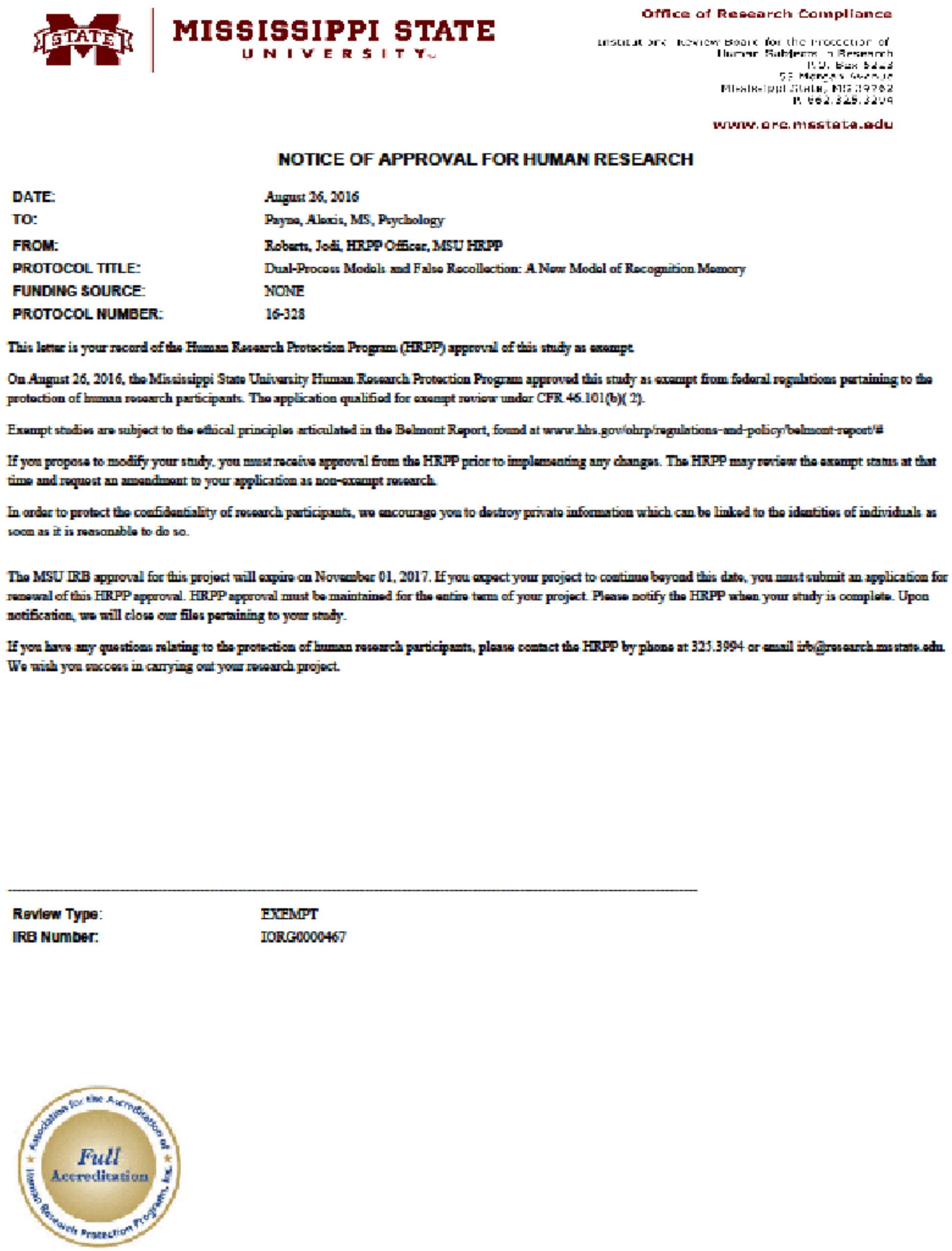
Table B1

Summary Scheme of Counterbalanced Lists

List 1		*List 3*	
Study:	Test:	Study:	Test:
Studied – a	Studied – a	Studied – a	FTi – a
Studied – b	FTi – b	Studied – b	Studied – b
FTx – a	FTx – b	FTx – b	FTx – a
Unrelated	Re-Paired	Unrelated	Re-Paired
	R-Control		R-Control
	U-Control		U-Control
List 2		*List 4*	
Study:	Test:	Study:	Test:
Studied – a	FTi – a	Studied – a	Studied – a
Studied – b	Studied – b	Studied – b	FTi – b
FTx – a	FTx – b	FTx – b	FTx – a
Unrelated	Re-Paired	Unrelated	Re-Paired
	R-Control		R-Control
	U-Control		U-Control

APPENDIX C

IRB APPROVAL FORM



MISSISSIPPI STATE UNIVERSITY

Office of Research Compliance

Institutional Review Board for the Protection of Human Subjects in Research
P.O. Box 5223
53 Morgan Avenue
Mississippi State, MS 39762
P. 662.325.3294

www.orc.msstate.edu

NOTICE OF APPROVAL FOR HUMAN RESEARCH

DATE:	August 26, 2016
TO:	Payne, Alexis, MS, Psychology
FROM:	Roberts, Jodi, HRPP Officer, MSU HRPP
PROTOCOL TITLE:	Dual-Process Models and False Recollection: A New Model of Recognition Memory
FUNDING SOURCE:	NONE
PROTOCOL NUMBER:	16-328

This letter is your record of the Human Research Protection Program (HRPP) approval of this study as exempt.

On August 26, 2016, the Mississippi State University Human Research Protection Program approved this study as exempt from federal regulations pertaining to the protection of human research participants. The application qualified for exempt review under CFR 46.101(b)(2).

Exempt studies are subject to the ethical principles articulated in the Belmont Report, found at www.hhs.gov/ohrp/regulations-and-policy/belmont-report/#

If you propose to modify your study, you must receive approval from the HRPP prior to implementing any changes. The HRPP may review the exempt status at that time and request an amendment to your application as non-exempt research.

In order to protect the confidentiality of research participants, we encourage you to destroy private information which can be linked to the identities of individuals as soon as it is reasonable to do so.

The MSU IRB approval for this project will expire on November 01, 2017. If you expect your project to continue beyond this date, you must submit an application for renewal of this HRPP approval. HRPP approval must be maintained for the entire term of your project. Please notify the HRPP when your study is complete. Upon notification, we will close our files pertaining to your study.

If you have any questions relating to the protection of human research participants, please contact the HRPP by phone at 325.3994 or email irb@research.msstate.edu. We wish you success in carrying out your research project.

Review Type:	EXEMPT
IRB Number:	IORG0000467

Figure C1. A copy of approval from Mississippi State University's IRB.

APPENDIX D

EXPERIMENT CONSENT FORM

Mississippi State University
Informed Consent Form for Participation in Research

Title of Research Study: Examining Relatedness (ER)

Study Site: Magruder Hall, Room 231

Researchers: Alexis E. Payne, M.S., Mississippi State University aej85@msstate.edu
Deborah K. Eakin, Ph. D., Mississippi State University de115@msstate.edu

You are being asked to be a volunteer in a research study. The purpose of this form is to tell you about the study you will be participating in today and to inform you about your rights as a research volunteer. Before you participate, you should read this consent form carefully and completely. You will be given a copy of this consent form to keep and you do not waive any of your legal rights by signing this consent form.

Thank you for volunteering to participate in this study. Our work could not be done without your help and willingness to give of your time and yourself.

Purpose
Our research focuses generally on how people process information, and specifically on what factors help or hinder processing of that information. This study is designed to examine how people process verbal materials, such as words or pairs of words (for example, "DOCTOR – FRIEND").

Procedures
If you decide to participate in this study, you will be asked to study a list of related word pairs. For each word pair, you will also be asked to rate each pair for relatedness. At each stage, you will be given instructions to make sure you know what you are supposed to do throughout the study. The word pairs will be presented on a computer screen one at a time and you will make all of your responses on the keyboard.

Risks or Discomforts
There are no major physical discomforts involved in this study. Risks are minimal and do not exceed those of normal office work. Please tell us if you are having trouble with any task or if you need additional rest and the researcher will be happy to accommodate you in any way possible. If you feel any discomfort, please tell the person assisting you immediately.

Benefits
This study will provide knowledge about how people learn to associate verbal information. Approximately 50 people will participate in various aspects of the study. You are not likely to benefit personally in any way from joining this study, but thanks to the willingness of people like you, we will continue to learn about the cognitive system, as well as how to improve quality of life.

Incentive to participate
The study will take 1 hour, and you will receive course credit for your participation. Student participants will receive ½ research credit for every half-hour of participation. We want you to know, however, that you are free to change your mind and withdraw from this research at any time. There will be no penalty for doing so. You will receive compensation equal to the time involved in the study. However, students will receive no less than ½ of a research credit.

Confidentiality
All of your responses will be kept strictly confidential. To protect the confidentiality of this information, we will assign your data a code number that will only be known to the members of the research project. All of the information which you provide us today will be marked with the code number, not your name. All information will be stored in a computer for analysis using only your code number for identification. No indication of your individual answers to questions will be given to anyone. We want you to be completely confident that you may feel free to answer all questions without concern that it may affect you in any way.

Figure D1. A copy of the consent form used in the dissertation experiment.

APPENDIX E

EXPERIMENT DEBRIEFING FORM

DEBRIEFING STATEMENT
Examining Relatedness (ER)

Thank you for participating in our study. Please read this statement carefully and ask the person assisting you about any questions that you may have.

The study in which you just participated was designed to examine how the relationship between two words affects your ability to recognize it at a later time. On the memory test, some of the word pairs were old and some were new. We wanted to determine how well you could recognize the old versus new pairs.

Again, thank you for participating in our study. Our research could not be done without your help. If you should have any questions, please do not hesitate to contact:

Alexis E. Payne, M.S.
Department of Psychology
Mississippi State University
662-325-5804
aej85@msstate.edu

Deborah K. Eakin, Ph.D.
Department of Psychology
Mississippi State University
662-325-7949
deakin@psychology.msstate.edu

Figure E1. A copy of the debriefing form used in the dissertation experiment.

APPENDIX F

ADDITIONAL MODELS

DPSD-FR2-FTi & Re-Paired Model

Another model that we developed was the dual process signal detection–false recollection 2-FTi and Re-Paired model, or DPSD-FR2-FTi & RP model. The DPSD-FR2-FTi & RP model includes two false recollection parameters for FTi and Re-Paired test pairs. The inclusion of this one false recollection parameter increases the total number of parameters from 11 in the DPSD model to 13 in the DPSD-FR2-FTi & RP model. Table F1 provides a full list of each parameter.

Table F1

A List of Parameters for the DPSD-FR2-FTi & RP Model.

Parameter:	Abbreviation:
Recollection	R
Studied Discriminability	d_S
FTi Discriminability	d_{FTi}
FTx Discriminability	d_{FTx}
Re-Paired Discriminability	d_{RP}
R-control Discriminability	d_{Rc}
U-control Discriminability	d_{Uc}
False Recollection - FTi	FR_{FTi}
False Recollection – Re-Paired	FR_{RP}
Criterion 1	C_1
Criterion 2	C_2
Criterion 3	0
Criterion 4	C_4
Criterion 4	C_5

Results

Fits to the aggregated data are shown in Figure F1. The DPSD-FR2-FTi & RP model served as the comparison model; the information criterion value for this was subtracted from the information criterion value for the DPSD model. Therefore, negative test values indicate an advantage of the DPSD model and positive values indicate an advantage of the DPSD-FR2-FTi & RP model. Table F2 shows the difference test values for the DPSD and DPSD-FR2-FTi & RP models. According to Raftery (1995), both the AIC and BIC comparisons reveal very strong evidence that the additional fit that the DPSD-FR2-FTi & RP model provides a better account of the data than the DPSD model for the aggregated data. However, for the analysis of the individual data supports the DPSD model.

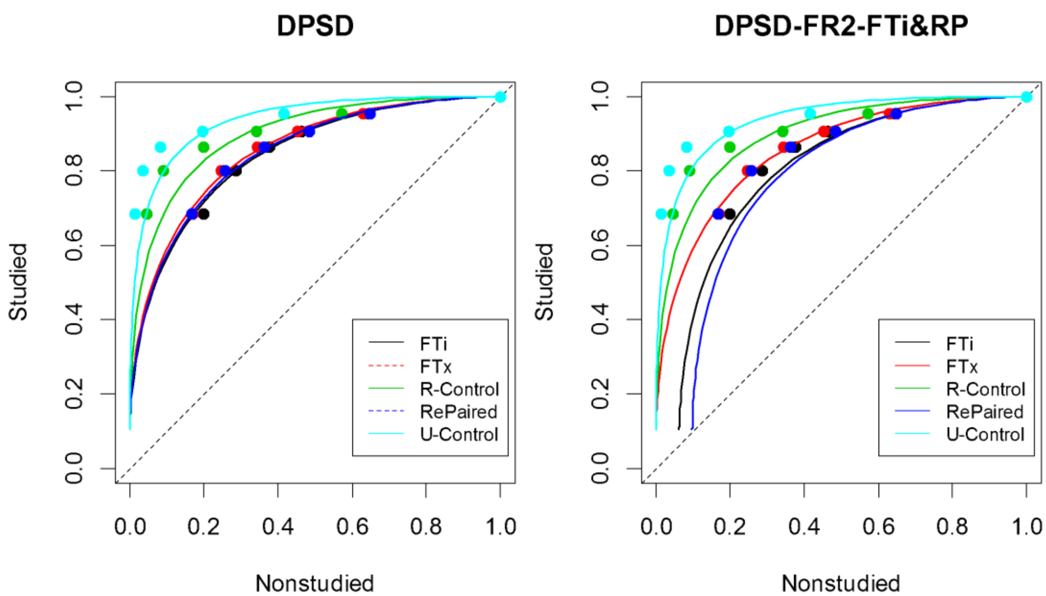


Figure F1. Comparison of DPSD and DPSD-FR2-FTi & RP model fits to the aggregated data. The behavioral data is represented by the points, and the predicted model data is represented by lines.

Table F2

Mean and Standard Error AIC and BIC Difference Scores

	Aggregate		Individual			
Model Comparison	AIC	BIC	AIC		BIC	
			M	*SE*	*M*	*SE*
DPSD vs. DPSD-FR2-FTi & RP	22.30	19.13	-1.33	.63	-4.50	.63
DPSD vs. DPSD-FR2-FTi & fFR	9.84	6.67	-2.14	1.10	-5.31	1.10
DPSD vs. DPSD-FR-ff	14.36	12.78	-3.76	1.63	-5.35	1.63

DPSD-FR2-FTi & Fixed Recollection Model

Another model that we developed was the dual process signal detection–false recollection 2-FTi and fixed false recollection model, or DPSD-FR2-FTi & fFR model. The DPSD-FR2-FTi & fFR model includes two false recollection parameters; one parameter is for FTi pairs and the other parameter is a fixed parameter for all other test pairs. The inclusion of this one false recollection parameter increases the total number of parameters from 11 in the DPSD model to 13 in the DPSD-FR2-FTi & fFR model. Table F3 provides a full list of each parameter.

Table F3

A List of Parameters for the DPSD-FR2-FTi & fFR Model

Parameter:	Abbreviation:
Recollection	R
Studied Discriminability	d_S
FTi Discriminability	d_{FTi}
FTx Discriminability	d_{FTx}
Re-Paired Discriminability	d_{RP}
R-control Discriminability	d_{Rc}
U-control Discriminability	d_{Uc}
False Recollection - FTi	FR_{FTi}
False Recollection – Fixed	FR_F
Criterion 1	C_1
Criterion 2	C_2
Criterion 3	0
Criterion 4	C_4
Criterion 4	C_5

Results

Fits to the aggregated data are shown in Figure F2. The DPSD-FR2-FTi & fFR model served as the comparison model; the information criterion value for this was subtracted from the information criterion value for the DPSD model. Therefore, negative test values indicate an advantage of the DPSD model and positive values indicate an advantage of the DPSD-FR2-FTi & fFR model. Table F2 shows the difference test values for the DPSD and DPSD-FR2-FTi & fFR models. Both the AIC and BIC comparisons reveal strong evidence that the additional fit that the DPSD-FR2-FTi & fFR model provides a better account of the data than the DPSD model for the aggregated data. However, for the analysis of the individual data supports the DPSD model.

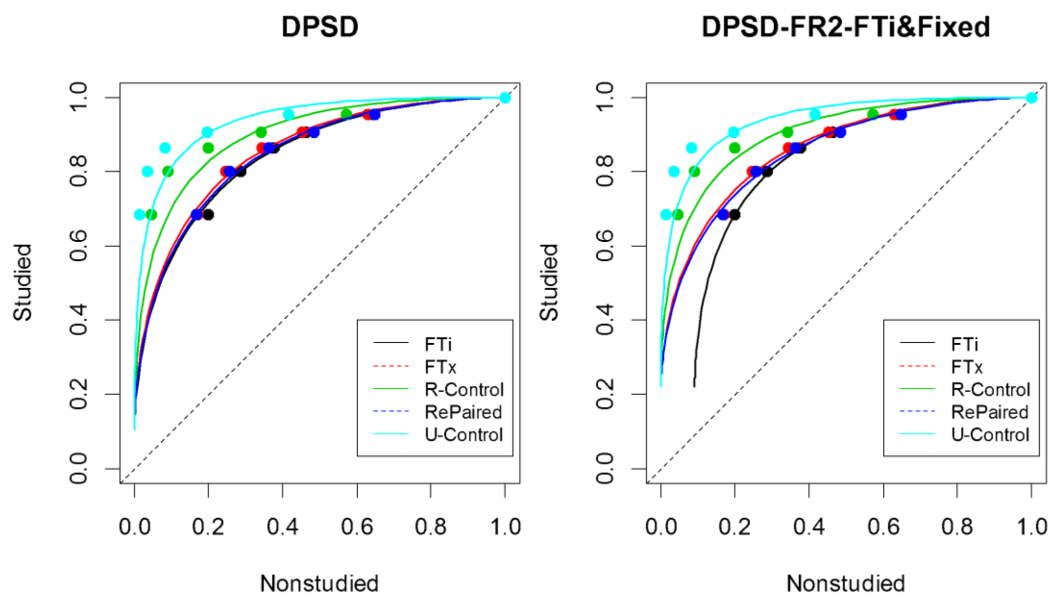


Figure F2. Comparison of DPSD and DPSD-FR2-FTi & Fixed False Recollection model fits to the aggregated data. The behavioral data is represented by the points, and the predicted model data is represented by lines.

DPSD-FR-ff Model

Another model that we developed was the dual process signal detection–false recollection –fixed familiarity model, or DPSD-FR-ff model. The DPSD-FR-ff model includes five false recollection parameters – one parameter for each of the test pairs – and a single familiarity parameter. Therefore, instead of five familiarity parameters like the DPSD model, this model has one d_n parameter. The inclusion of the five false recollection parameter and one nonstudied familiarity parameter changes the total number of parameters from 11 in the DPSD model to 1 in the DPSD-FR-ff model. Table F4 provides a full list of each parameter.

Table F4

A List of Parameters for the DPSD-FR-ff Model

Parameter:	Abbreviation:
Recollection	R
Studied Discriminability	d_S
Fixed Discriminability	d_F
False Recollection	FR_{ff}
Criterion 1	C_1
Criterion 2	C_2
Criterion 3	*0*
Criterion 4	C_4
Criterion 4	C_5

Results

Fits to the aggregated data are shown in Figure F3. The DPSD-FR-ff model served as the comparison model; the information criterion value for this was subtracted from the information criterion value for the DPSD model. Therefore, negative test values indicate an advantage of the DPSD model and positive values indicate an advantage of the DPSD-FR-ff model. Table F2 shows the difference test values for the DPSD and DPSD-FR-ff models. Both the AIC and BIC comparisons reveal very strong evidence that the additional fit that the DPSD-FR-ff model provides a better account of the data than the DPSD model for the aggregated data. However, for the analysis of the individual data supports the DPSD model.

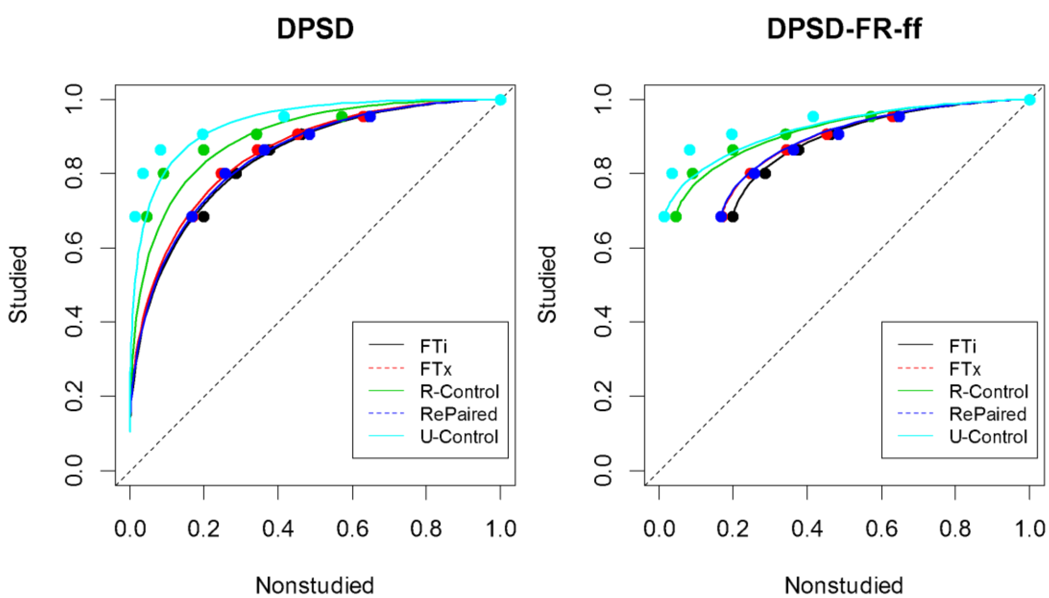


Figure F3. Comparison of DPSD and DPSD-FR-ff model fits to the aggregated data. The behavioral data is represented by the points, and the predicted model data is represented by lines.