Title
A signal-detection-based investigation into the nature of recognition memory

Permalink
https://escholarship.org/uc/item/9gv148q4

Author
Mickes, Laura Beth

Publication Date
2010

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

A Signal-Detection-Based Investigation into the Nature of Recognition Memory

A dissertation submitted in partial satisfaction of the requirements for the degree
Doctor of Philosophy

in

Psychology and Cognitive Science

by

Laura Beth Mickes

Committee in charge:

Professor John T. Wixted, Chair
Professor Stephan Anagnostaras
Professor Andrea Chiba
Professor Christine Fennema-Notestine
Professor Mark Jacobson
Professor Guerry Peavy
Professor Timothy Rickard
Professor David Salmon

2010

The Dissertation of Laura Beth Mickes is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2010

EPIGRAPH

Most of my memories consist of information about the topic of memory.

Daniel Bajic

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

First I would like to express my deepest gratitude for my inspiring advisor, John T. Wixted. Among many things, John taught me how to think experimentally, and communicate my ideas more clearly. In one word, he is superb!

I would be remiss if I neglected to recognize my committee members, especially Christine Fennema-Notestine, Mark Jacobson, Guerry Peavy, and David Salmon, all of whom mentored me and taught me how to think like an experimental clinician, and who all remained on my committee despite the drastic change in topic.

My graduate compadre and fellow instigator, Peter Wais, also deserves my acknowledgement. Graduate school was a more unique experience because of him. In the insightful words of an anonymous student on a course evaluation, "Laura and Peter are weird together" sums it up.

Next, my work would not be what it is without the help of the small army of research assistants who became dear friends, including: Philip Tseng, Janice Lee, Miki Ogasawara, Kris Wisner, Emily Johnson (Bumpy Brains co-founder!), Sarah Cullado, and Drew Hoffman.

Finally, I deeply thank Mom and Howard Field, Dad and Carla Mickes, and Stephen Dye who supported and encouraged me throughout my never-ending quest for higher education. I love you!

Chapter 1, in full, is a reprint of the material as it appears in A Direct Test of the Unequal Variance Signal Detection Model of Recognition Memory in *Psychonomic Bulletin & Review, 14,* 858-865. Mickes, L., Wixted, J. T., & Wais, P. (2007). The dissertation author was the primary investigator and author of this paper.

Chapter 2, in full, is a reprint of the material as it appears in Recollection is a Continuous Process: Implications for Dual Process Theories of Recognition Memory in *Psychological Science, 20,* 509-515. Mickes, L., Wais, P. E., & Wixted, J. T. (2009). The dissertation author was the primary investigator and author of this paper.

Chapter 3, in full, has been submitted for publication of the material as it may appear in Continuous Recollection vs. Unitized Familiarity in Associative Recognition in *Journal of Experimental Psychology: Learning, Memory & Cognition.* Mickes, L., Johnson, E. M., & Wixted, J. T. (2010). The dissertation author was the primary investigator and author of this paper.

CURRICULUM VITAE

2002	Bachelor of Arts, California State University, Los Angeles
2003-2009	Graduate Student, Department of Psychology, University of California, San Diego
2000-2003	Laboratory Assistant, University of California, Los Angeles, Laboratory of Brain, Behavior, and Pharmacology, University of California, Los Angeles, Semel Institute for Neuroscience and Human Behavior
1999-2002	Research Assistant, California State University Los Angeles, Department of Psychology, Memory and Psychopathology Laboratory
2004	Master of Arts, University of California, San Diego

PUBLICATIONS

Mickes, L., Johnson, E. M., & Wixted, J. T. (in press). Continuous recollection vs. unitized familiarity in associative recognition. *Journal of Experimental Psychology: Learning, Memory & Cognition.*

Wixted, J. T., & Mickes, L. (in submission). Memory strength vs. memory content: A signal-detection-based, dual process interpretation of remember/know judgments.

Wixted, J. T., & Mickes, L. (in submission). Useful scientific theories are useful: A reply to Rouder, Pratte and Morey (2010).

Mickes, L., Wais, P. E., & Wixted, J. T. (in submission). Strong memories are not scalable.

Mickes, L., Jacobson, M., Peavy, G., Wixted, J. T., Lessig, S., Goldstein, J. L., & Corey-Bloom, J. (in press). A comparison of the utility of two brief screening measures for identifying cognitive impairment in Huntington's disease. *Movement Disorders.*

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,* 509-515.

Mickes, L., Wixted, J. T., Shapiro, A., & Scarff, M. (2008). The effects of pregnancy on memory: Recall is worse but recognition is not. *Journal of Clinical and Experimental Neuropsychology, 31,* 754-761.

Eisen, M. L., Oustinayaka, M., Kistorian, R., Morgan, D. Y., & Mickes, L. (2008). The effect of question format on resistance to misleading post event information and self-report of events occurring during hypnosis. *International Journal of Clinical and Experimental Hypnosis, 56.*

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

Mickes, L., Wixted, J. T., Fennema-Notestine, C., Thal, L., Galasko, D., Bondi, M., & Salmon, D. P. (2007). Progressive impairment on neuropsychological tasks in a longitudinal study of preclinical Alzheimer's disease. *Neuropsychology, 21,* 696-705.

Wais, P., Mickes, L., & Wixted, J. T. (2007). Remember/know judgments probe degrees of recognition. *Journal of Cognitive Neuroscience, 20,* 400-405.

Cook, I.A., Bookheimer, S. Y., Mickes, L., Leuchter, A. F., & Kumar, A. (2007). Aging and brain activation with working memory tasks: An fMRI study of connectivity. *International Journal of Geriatric Psychiatry, 22,* 332-342.

Cook, I. A., Leuchter, A. F., Morgan, M. L., Dunkin, J. J., Witte, E., David, S., Mickes, L., O'Hara, R., Simon, S., Lufkin, R., Abrams, M., & Rosenberg, S. (2004). Longitudinal Progression of Subclinical Structural Brain Disease in Normal Aging. *American Journal of Geriatric Psychiatry, 12,* 190-200.

Cook, I. A., Leuchter, A. F., Morgan, M. L., Witte, E., David, S., Lufkin, R., Babaie, A., Dunkin, J. J., O'Hara, R., Simon, S., Lightner, A., Thomas, S., Broumandi, D., Badjatia, N., Mickes, L., Mody, R., Arroyo, S., Zheng, Z., Abrams, M. &Rosenberg-Thompson, S. (2002). Cognitive and Physiological Correlates of Subclinical Structural Brain Disease in Elderly Normal Controls. *Archives of Neurology, 59,* 1612-1620.

Eisen, M. L., Morgan, D. Y., & Mickes, L. (2002). Individual differences in eyewitness memory and suggestibility: Examining relations between acquiescence, attention and resistance to misleading information. *Personality and Individual Differences, 33*, 553-571.

FIELDS OF STUDY

Major Field: Psychology
Sub-fields: Cognitive Neuropsychology and Cognitive Science

ABSTRACT OF THE DISSERTATION

A Signal-Detection-Based Investigation into the Nature of Recognition Memory

by

Laura Beth Mickes

Doctor of Philosophy in Psychology and Cognitive Science

University of California, San Diego, 2010

Professor John T. Wixted, Chair

Recognition memory is the ability to consciously appreciate that an item or event was previously presented or experienced. Signal detection theory has long provided one influential interpretation of recognition memory, and numerous investigations conducted over the last 50 years have sought to clarify the particulars of this account. Analyzing receiver operating characteristic (ROC) data can distinguish between two versions of signal detection theory, specifically, the equal and unequal variance models. The equal variance signal detection model is intuitively appealing, but the unequal variance signal detection model usually provides a better fit of the ROC data. Chapter 1 describes two experiments that provide a novel test of the unequal variance assumption. This new method of analysis required subjects to directly rate their memory strength on a fine-grained scale, and then the mean and standard deviations of the target and lure ratings were directly computed. Results from the new method support the unequal variance signal detection model. Though the unequal variance signal detection theory of recognition memory provides a useful way to conceptualize recognition, there is another long-standing theory of recognition known as dual process theory that seems to contradict it. This theory holds that two processes (familiarity and recollection) contribute to recognition decisions. A critical point of contention between standard dual process models and signal detection theory concerns the nature of the recollection process, specifically, whether it is continuous or categorical. Dual-process theories generally assume that recollection is categorical, but signal detection theory requires it to be continuous. Chapters 2 and 3 provide direct evidence that recollection is a continuous process. In Chapter 2, two versions of a source memory experiment were conducted. The continuous view of recollection was supported because the relationship between confidence and accuracy on this recollection-based task was graded. The results detailed in Chapter 3 further validate recollection as a continuous process. The method involved an associative recognition test, which (like a source memory procedure) purportedly tests recollection in the absence of familiarity. In this task, word pairs were studied and then at test, the pairs were either intact or rearranged. When the word pairs were strengthened, we observed the typical result of an increasingly curvilinear ROC. Evidence from various procedures (i.e., remember/know procedure, old/new decision, cued recall test) converged to suggest that recollection is a continuous process. The three chapters support the unequal variance signal detection theory of recognition memory and the idea that two continuous processes aggregate to yield a continuous

memory strength variable.

INTRODUCTION

A Signal-Detection-Based Investigation into the Nature of Recognition Memory

The ability to consciously appreciate that an item or event was previously presented or experienced is referred to as *recognition* memory. Understanding how recognition memory works has been a central focus for the field of experimental psychology for more than 100 years, and, in recent years, psychological models of recognition have exerted great influence over investigations in to the brain basis of human memory.

The study of recognition memory in the laboratory generally follows a standard sequence of events. There are two phases of a typical recognition memory experiment: study and test. During study, subjects are presented with a list of items (e.g., words) to memorize. Next, they are tested on their ability to distinguish between items that were presented during study, which are referred to as targets, from items that were not presented during study, known as lures. In the most common test format, the targets and lures are randomly intermixed and presented one at a time for an old or new decision. That is, for each item, subjects respond yes, or “old”, to an item they believe appeared during study, and no, or “new”, to an item they believe was not presented during study. Each individual subject’s proportion of correct old responses for targets (hit rate) and proportion of incorrect responses for lures (false alarm rate) are calculated. For example, a subject might correctly identify 80 out of 100 words as old, but incorrectly declare 30 of the 100 lures old. In this scenario, the hit rate would be 80% and false alarm rate would be 30%.

Signal detection theory is a longstanding and influential account that provides an interpretation of the hit and false alarm rate. According to this theory, the test items vary in memory strength, with the mean strength of the targets being higher than that of the lures. The distributions of memory strengths for targets and lures are generally assumed to be Gaussian in form, though this is not a necessary assumption. Somewhere along the memory strength axis, subjects place a decision criterion. Any test item that generates a memory signal that is greater than the decision criterion receives an old response, and any item that generates a memory signal that is weaker than the criterion receives a new response. A subject with a hit rate of 84% and a false alarm rate of 16% would be interpreted by signal detection theory as 84% of the targets and 16% of the lures exceeded the criterion.

One way to get more information about a subject's memory is to obtain their confidence for each recognition memory decision. That is, instead of making a binary old/new decision, they rate their confidence on (for example) a 6-point scale. On this scale, a rating of 1 indicates that the subject is *sure* that the item was new, a rating of 2 indicates that the item was *probably* new, and a rating of 3 indicates that the item is *maybe* new. On the other side of the scale, ratings of 4, 5 or 6 indicate that the item is maybe old, probably old or definitely old, respectively. From these ratings, hit and false alarm rates can be computed for each level of confidence. For example, it might be observed that 31% of the targets and 1% of the lures received a confidence rating of 6. That would correspond to a hit rate of 31% and a false alarm rate of 1% for ratings of 6 for that individual subject. Next, another hit and false alarm rate pair would be obtained by computing the proportion of targets and lures that received ratings of 5 and 6. The ratings are cumulated in this manner until there are five separate hit and false alarm rate pairs. A plot of the hit rate vs. the false alarm is known as the Receiver Operating Characteristic (ROC).

Signal detection theory interprets the confidence rating data in a way that is similar to the simpler case in which the subject makes only an old/new decision yielding only one hit and false alarm rate pair. To make confidence ratings, the model assumes that additional decision criteria placed on the memory signal strength axis. If a 6-point confidence scale is used, the model assumes that five distinct decision criteria are set. A high confidence rating of 6 is made when the memory strength of a test item exceeds the criterion set for 6 (i.e., the highest criterion). In this example, the hypothetical subject gave a rating of 6 to 31% of targets and 1% of lures, which means that the memory strength of 31% of the targets and 1% of the lures exceeded the criterion placed for a very high confidence rating. If the memory strength of an item is strong enough to exceed the criterion for a rating of 5 but is not high enough to exceed the criterion for a rating of 6, then the subject would rate it 5. To compute a hit and false alarm rate associated with the criterion for making ratings of at least 5, the proportion of the target and lure distributions that fall to the right of that criterion would be estimated by computing the proportion of targets and the proportion of lures that receive ratings of 5 or 6. Thus, each of the hit and false alarm rate pairs computed from confidence rating data to plot an ROC are conceptualized as estimates of the proportion of the target and lure distributions that fall to the right of a confidence-specific decision criterion.

The ROC provides theoretically relevant information because it helps determine the details of the underlying signal detection model.

Signal Detection Theory: Equal Variance Model or Unequal Variance Model

Analyzing ROC data can distinguish between two different signal detection models – the equal and unequal variance models. The equal variance model has a lure distribution and a target distribution that have different means but are equal in variance. The unequal variance model has a target distribution that is greater in variance than the lure distribution. Though the equal variance model is an intuitively appealing model for describing and predicting how recognition memory operates, the unequal variance model actually seems more plausible. Consider that target items are effectively lures that have had strength added to them by the virtue of having been presented on a list during the study session. Thus, the target distribution is shifted to the right on the memory strength axis. To maintain the precise shape of the original lure distribution (i.e., for the variance to remain unchanged as the mean increases), precisely the same amount of strength would have to be added to every studied item. This seems unlikely. Rather, it seems more plausible that during study, the strength added to individual items varies across items (e.g., a subject may pay a lot of attention to some items but not much to others). If so, there would then be a shift rightward of the target distribution, and there would be the additional increase in its variance. Therefore, because it is plausible to assume that items differ in the amount of strength added during study, the variance of the target distribution should be greater than that of the lure distribution.

If the equal variance account were true, the ROC data would reveal two related properties (Macmilan & Creelman, 2005). First, when the hits and false alarm rate pairs for each confidence rating are plotted, it would trace out a curvilinear *symmetrical* path. That is, symmetry of the distributions translates to symmetry of the ROC curve. Another way to determine if the equal variance account is accurate is to measure the slope of the z-ROC. A z-ROC is a plot of the z-transformed hit rates vs. the z-transformed false alarm rates. If the underlying distributions are Gaussian in form, the z-ROC should be linear (Macmillan & Creelman, 2005). In addition, an equal variance model would yield a slope of 1 because the slope is theoretically equal to the standard deviation of the lure distribution divided by the standard deviation of the target distribution. In summary, if the two distributions are equal in variance, then the ROC will be symmetrically curvilinear and the z-ROC will be linear with a slope equal to one.

In practice, the ROC data typically support the unequal variance account because when the hit and false alarm rate pairs are plotted in ROC space, the path is usually *asymmetrically* curvilinear.. The asymmetry here corresponds to the asymmetry of the distributions. Moreover, when plotted in z-ROC space, the slope is typically found to be approximately .80, which indicates that the standard deviation of the lure distribution is .80 times that of the target distribution. ROC data almost invariably support the unequal variance account, and this is the main reason why the unequal-variance signal-detection model has been a prominent way to conceptualize decision-making in recognition memory since 1958 (Egan, 1958).

Chapter 1 describes two experiments that provide a novel test of the signal detection theory unequal variance assumption. Here, direct ratings (on a 20-point scale, experiment 1; and on a 99-point scale, experiment 2) were compared to the estimates provided by ROC analysis. Unlike ROC analysis, the direct rating method does not rely on any assumptions about the form of the underlying target and lure distributions (i.e., it does not assume that they are Gaussian). The findings support the unequal variance account of recognition memory.

Recollection is Categorical or Continuous

Though the unequal variance signal detection model provides a useful way to conceptualize recognition memory, there is another long-standing and seemingly contradictory account of recognition known as dual process theory (Atkinson & Juola, 1974; Mandler, 1980). Dual process theory holds that two processes contribute to recognition decisions, namely, familiarity and recollection. Familiarity is a context-free sense of prior occurrence, whereas recollection is the ability to recount details, or retrieve contextual or source information, associated with the item. The dual process model seems to contradict the signal detection model in that the unequal variance model has only one memory signal strength axis. It thus seems inherently incompatible with Mandler's idea that there are two processes. However, this is not actually the main source of incompatibility between extant dual-process models and the unequal-variance signal detection account. Wixted (2007) proposed that the two processes may be continuous variables that are combined to yield a composite memory strength signal. Thus, the mere fact that recollection and familiarity support recognition decisions is not fundamentally incompatible with a signal detection model that involves only one memory strength axis. Instead, the signal detection account is incompatible with particular versions of dual process theory that have had great influence over the past 15 to 20 years.

Perhaps the most prominent dual process theory is the dual process signal detection (DPSD) theory originally advanced by Yonelinas (1994). Despite its name, this theory is not a pure signal detection account. The theory holds that there are two independent processes that make up recognition memory (in agreement with earlier dual process theories, e.g., Mandler, 1980): familiarity and recollection. According to the DPSD model, the familiarity process is governed by an *equal* variance signal detection model. By contrast, recollection is construed a categorical process. According to this idea, recollection either occurs for a particular test item or it does not occur (with no degrees of recollection in between). Thus, recollection is not construed as a continuous process. If a subject recollects an item, the model assumes that a high-confidence "old" decision is made (i.e., they supply a rating of 6 if a 6-point confidence scale is used). If a subject does not recollect an item, the model assumes that they then determine whether that item is familiar enough to declare it to be old. Purely familiarity-based responding would yield a symmetrical ROC, but this model can account for asymmetrical ROC data by assuming that categorical recollection occurs for some of the test items. That is, in the DPSD model, ROC asymmetry is the signature of recollection. Thus, both the DPSD and the unequal variance signal detection theory can explain typical ROC results.

The main difference between these theories is how the process of recollection is viewed. The DPSD theory posits that recollection is categorical, whereas a dual process version of the unequal variance signal detection theory holds that recollection is continuous (Wixted, 2007). The process of recollection is therefore the crux of the debate. Chapter 2 describes two experiments that test whether recollection is a continuous or a categorical process. To do this, we tested the recollection process by using a source memory paradigm. In this procedure, memory for items that are associated with certain attributes, or sources (i.e., words at various locations on the screen) is tested. Recollection is widely thought to be the process that underlies the ability to remember source details, whereas familiarity is thought to play little or no role. If the categorical view of recollection were correct, then accuracy for items given the highest confidence rating would be high, but accuracy should fall off as a step function to chance levels for all lower confidence ratings. However, if the continuous view of recollection were supported, then the relationship between accuracy and confidence ratings would be graded (i.e, the highest ratings would have the highest accuracy, and the next highest ratings would have the next highest accuracy, and so on). Two source memory experiments are described in detail, and both experiments support the idea that recollection is a continuous process.

Chapter 3 describes another set of experiments in which the nature of the recollection process is further examined. Like source memory, associative recognition is thought to rely on recollection with little or no contribution from the familiarity process. In this task, word pairs are studied and then at test, the pairs are left intact or they are rearranged. Because all of the individual words are familiar for both intact and rearranged pairs, recollection is usually thought to be the only process that can be used to solve the task.

Initially, associative recognition ROC data were found to be linear and the z-

ROC data were found to be curvilinear (Yonelinas, 1997), which is the opposite of the pattern obtained for old/new ROC data. If recollection is a categorical process, and if associative recognition is based only on recollection, then a linear ROC is predicted. The fact that a linear ROC was in fact obtained therefore offered strong support for the idea that recollection is a categorical process.

Later, it was shown that the associative recognition ROC becomes increasingly curvilinear as the memory strength is increased (e.g., as subjects are given more time to study the pairs). This was hard for the DPSD model to explain because increased study time should increase recollection, which would not be expected to change the shape of the linear ROC. However, the DPSD model explains the increased curvilinearity with a new idea known as "unitized familiarity" (Yonelinas, 1997). Unitized familiarity theoretically results from encoding the pair of words as a unit. If intact pairs have greater unitized familiarity than rearranged pairs, then familiarity could be used to solve the associative recognition task after all (and the ROC would be curvilinear). According to this account, extra study time results in increased unitized familiarity, which is why the ROC becomes curvilinear.

Conversely, signal detection theory, which holds that recollection is a continuous process, has trouble explaining why the associative recognition ROC tends to be linear when memory is weak. A continuous process should yield a curvilinear ROC whether memory is weak or strong. By contrast, the theory easily accounts for the curvilinear ROC that emerges when memory is strong.

In this chapter, the mixture signal detection (MSD) model is introduced, which includes a third "noise" distribution that represents pairs for which no associative information was encoded at study. The existence of such pairs would also cause an ROC to be linear, and there should be more such pairs when memory is weak than when it is strong. Chapter 3 describes two experiments in which these theories (DSPD vs. MSD) are put to the test, and the MSD model is supported in the weak condition. Additionally, the overall pattern of results supports the idea that recollection is a continuous process, not a categorical process.

The three chapters presented here contribute to a growing body of evidence that the signal detection theory of recognition memory is viable for interpreting recognition memory data, whether recollection is based on recollection, familiarity or both.

CHAPTER 1: A Direct Test of the Unequal-Variance Signal-Detection Model of Recognition Memory

Laura Mickes, John T. Wixted & Peter Wais (2007). *Psychonomic Bulletin & Review, 14(5),* 858-865.

Summary

Analyses of the Receiver Operating Characteristic (ROC) almost invariably suggest that, on a recognition memory test, the standard deviation of memory strengths associated with the lures (σ_{lure}) is smaller that of the targets (σ_{target}). Often, $\sigma_{\text{lure}}/\sigma_{\text{target}} \approx 0.80$. However, that conclusion is based on a model that assumes that the memory strength distributions are Gaussian in form. In two experiments, we investigated this issue in a more direct way by asking subjects to simply rate the memory strengths of targets and lures using a 20-point or a 99-point strength scale. The results showed that the standard deviation of the ratings made to the targets (s_{target}) was, indeed, larger than the standard deviation of the ratings made to the lures (s_{lure}). Moreover, across subjects, the ratio $s_{\text{lure}}/s_{\text{target}}$ correlated highly with the estimate of $\sigma_{\text{lure}}/\sigma_{\text{target}}$ obtained from ROC analysis, and both estimates were, on average, approximately equal to 0.80.

Dual-Process Theory and Signal-Detection Theory of Recognition Memory

Signal-detection theory has long been a prominent theoretical framework for understanding how subjects make decisions on recognition memory tasks. The textbook version of the theory involves two equal-variance Gaussian distributions and a decision criterion placed somewhere along the memory strength axis. One distribution represents the memory strengths of the lures, and it has a low average value. The other distribution represents the memory strengths of the targets, and it has a higher average value. Any test item that generates a memory strength exceeding the criterion is declared to be Old, otherwise it is declared to be New (as illustrated in the upper panel of Figure 1). Although the aesthetically appealing equal-variance version of the model is often used to illustrate signal-detection theory, analyses of the empirical Receiver Operating Characteristic (ROC) almost always imply an unequal-variance model in which the standard deviation of the target distribution exceeds that of the lure distribution (Egan, 1958, 1975; Ratcliff, Shue & Gronlund 1992), as illustrated in the lower panel of Figure 1.

INSERT FIGURE 1 ABOUT HERE

An ROC is simply a plot of the hit rate (HR) vs. the false alarm rate (FAR) for different levels of bias. A typical ROC is obtained by asking subjects to supply confidence ratings for their recognition memory decisions, often on a 6-point scale. Signal-detection theory predicts that the ROC will be curvilinear in probability space (HR vs. FAR) and linear in z-space (z-HR vs. z-FAR), and it holds that the slope of the z-ROC provides an estimate of the ratio of the standard deviation of the lure distribution to the standard deviation of the target distribution ($\sigma_{lure}/\sigma_{target}$). If an equal-variance model applies (as in the upper panel of Figure 1), then the slope should be 1.0. But if the standard deviation of the target distribution exceeds that of the lure distribution (as in the lower panel of Figure 1), then the slope of the z-ROC should be less than 1.0.

Previous reviews of the ROC literature indicate that z-ROCs are well-characterized by a straight line and that the slope of the best-fitting line is, on average, approximately 0.80 (Glanzer, Kim, Hilford, & Adams, 1999; Ratcliff et al., 1992). Thus, according to the signal detection account, the standard deviation of the target distribution is often about 1.25 (i.e., 1/0.80) times that of the lure distribution. Findings like these explain why the unequal variance model shown in the lower panel of Figure 1 is regarded by some as the standard model of decision-making on a recognition memory task. Others, however, find the model to be less compelling. For example, the majority of investigations into the neuroanatomical basis of recognition memory either implicitly or explicitly reject this way of thinking (Wixted, in press). If signal-detection theory does provide an accurate model of decision-making, then those investigations could be led astray by the alternative decision-making models they embrace.

One issue that bears on the validity of the detection account is its suggestion that the standard deviation of the target distribution is greater than that of the lure distribution. That conclusion is based on an analysis that assumes that the underlying distributions of memory strength are Gaussian in form. Although ROC data are well fit by a Gaussian model, it has long been known that other distributions -- ones that are quite unlike the Gaussian -- also fit ROC data well. Instead of relying on ROC analysis, a more direct test of the unequal-variance idea would be to simply ask subjects to rate the memory strengths of targets and lures using a fine-grained scale (e.g., 1 to 99). The mean and standard deviation of the ratings for the targets (m_{target} and s_{target}, respectively) could then be directly computed and then compared to the mean and standard deviation of the ratings for the lures (m_{lure} and s_{lure}, respectively). Although the mean rating for the targets would undoubtedly be greater than the mean rating for the lures, would the standard deviation of the target ratings be greater as well? And, if so, would the ratio of the standard deviation of the lure ratings to the standard deviation of the target ratings be approximately 0.80, as suggested by ROC analysis? These are the questions we set out to address.

Experiment 1

In the first experiment, subjects were presented with a list of 150 words to memorize, after which they completed a recognition memory test that involved those 150 targets randomly intermixed with 150 lures. For each test item, the subject was asked to rate the strength of their memory for that item on a 1-to-20 scale.

Method

Participants

Fourteen undergraduates from University of California, San Diego participated for lower division psychology course credit.

Materials and design

The word pool used consisted of 705 three-to-seven letter words taken from the MRC Psycholinguistic Database (Coltheart, 1981), of which 300 words were randomly selected for testing (150 of which were randomly selected to be targets, while the remainder were lures). Instructions and stimuli were displayed for each participant on a NEC MultiSync LCD1760NX monitor, and powered by a Dell Dimension 4550. Stimuli were presented using an E-prime program (www.pstnet.com; Psychology Software Tools).

Procedure

Participants signed a consent form, were read instructions, studied the 150 targets, and completed a recognition test in which the 150 targets were randomly intermixed with the 150 lures. Each word was presented for 2 seconds during study. During testing, participants indicated whether or not the word was on the presented list by pressing a key; then they indicated the strength of their memory for that word by entering a number on the keypad ranging from 1 to 20, with 1 meaning the word was definitely not on the list and with 20 meaning the word was definitely on the list. These instructions were given verbally prior to list presentation and appeared again on the screen after the list was presented. In addition, the verbal instructions asked participants to be cautious about using the endpoints of 1 and 20. They were instructed to use those values only when they were 100% certain, as they might be if their own name was used as a test item.

Results

Subjects generally distributed their responses over the full range of the scale. The upper panel of Figure 2 shows the frequency distribution for targets and lures pooled over subjects. The lure distribution appears to be somewhat truncated on the left (as if some lures would have received lower ratings, if possible), and the target distribution appears even more truncated on the right (as if some targets would have received higher ratings), but the figure illustrates the central assumption of signal-detection theory: the distribution of memory strengths for the targets and lures overlap, with the mean of the target distribution being higher than that of the lure distribution. The lower panel of Figure 2 shows decision accuracy for each rating. A rating in the range of 1 through 10 was scored as a correct response to lures (and an incorrect response to targets), whereas the reverse was true for ratings in the range of 11 through 20. In accordance with the predictions of signal detection theory, accuracy varies continuously as the distance from the indifference point increases.

INSERT FIGURE 2 ABOUT HERE

As shown in Table 1, most subjects were relatively unbiased in their use of the rating scale such that their ratings for all items averaged together (m_{overall}) were close to the mid-point of the scale (10.5), with the overall mean being 10.77. However, Subject 11 was an exception. That subject's average rating across targets and lures was 14.2, which is 2.40 standard deviations above the mean. Indeed, even for lures, this subject's mean rating exceeded 10. This is an important consideration because if a subject's ratings are biased towards one end of the scale (as this subject's ratings clearly are), then the ratings for one class of items will be more compressed than the ratings for the other class. Except where noted, this subject was excluded from the main analysis.

Table 1 also shows the mean and standard deviations for the ratings made to the targets (m_{target} and s_{target}, respectively) and to the lures (m_{lure} and s_{lure}) for each subject. Across all subjects (excluding Subject 11), the mean rating for the targets was 12.98, and the mean rating for the lures was 8.04. The corresponding standard deviations -- which are the main measures of interest -- were 4.62 and 3.83, respectively. Table 1 also shows, for each subject, the ratio of the standard deviation of the lure ratings to the standard deviation of the target rating (s_{lure}/s_{target}). Excluding the outlier, the mean ratio is 0.83, which is significantly less than 1.0, t(12) = 3.52. With the outlier included, the mean ratio is 0.87, which is still significantly less than 1.0, t(13) = 2.42.

We next conducted an ROC analysis on these data by counting the number of responses to targets and number of responses to lures that exceeded the following cutoffs on the rating scale: 17, 14, 11, 8, 5, and 1. That is, we treated the rating scale as if it were a 6-point confidence scale, with a rating of 17 to 20 being regarded as a high-confident Old response, a rating of 14 to 16 as a medium-confident Old response, and so on down to ratings of 1 to 4, which were treated as high-confident New responses. The confidence scale is assumed to provide only an ordinal scale of measurement. That is, the high-confident Old criterion is assumed to be higher on the memory strength scale than the medium-confident Old criterion, but the distance between those two criteria need not be the same as the distance between the medium-confident criterion and the low-confident criterion. Because the direct rating method and the ROC method entail quite different assumptions, they need not agree in their conclusions (as illustrated in detail later).

The ROC analysis was performed by fitting the Gaussian detection model to the ROC data of each individual subject using maximum likelihood estimation. One of the parameters of the model is the ratio of the standard deviation of the lure distribution divided by the standard deviation of the target distribution ($\sigma_{\text{lure}}/\sigma_{\text{target}}$). The estimated value of that ratio for each subject is shown in Table 1. The mean value was 0.79, which is typical and is quite close to the value obtained from direct ratings (0.83). Figure 3 shows the scatterplot of ratio measures derived from the two procedures for each subject. It is clear for the figure that the estimates are in good agreement ($r = .61$, $p < .05$).

INSERT FIGURE 3 ABOUT HERE

The ROC analysis also yielded a discriminability measure for each subject, and the corresponding version of that measure was also computed directly from the ratings. The typical detection-based discriminability measure is d', which is the distance between the means of the target and lure distributions in standard deviation units. That is, $d' = (\mu_{\text{Target}} - \mu_{\text{Lure}}) / \sigma$, where σ is the standard deviation of both the target and lure distributions. When an unequal variance model applies, a related (and better) measure is d_a, which is the distance between the means relative to the root mean square of the target and lure standard deviations:

$$d_a = \frac{\mu_{target} - \mu_{lure}}{\sqrt{(\sigma^2_{target} + \sigma^2_{lure})/2}}$$

For each subject, d_a was estimated from ROC analysis. A value analogous to d_a, denoted d_r, was then computed for each subject directly from the ratings according to the following formula:

$$d_r = \frac{m_{target} - m_{lure}}{\sqrt{(s^2_{target} + s^2_{lure})/2}}$$

Remarkably, the discriminability estimates obtained from ROC analysis and from the ratings were nearly identical (r = .99, mean d_a = 1.15, mean d_r = 1.17).

Discussion

ROC analyses of recognition memory almost invariably suggest that the memory strengths of the targets are more variable than the memory strengths of the lures. Using direct ratings of memory strength for targets and lures in which the means and standard deviations could be computed directly, we found that the results were in good agreement with ROC analysis. Both methods suggested that the standard deviation of the lure distribution is about .80 times that of the target distribution, on average, and the ratio estimates for individual subjects derived from the two methods correlated significantly.

The ratings method and the ROC method are not constrained to agree on this issue. To illustrate this, we conducted two simulations based on the signal-detection models shown in the upper and lower panels of Figure 4. For both simulations, memory strengths for targets and lures were drawn from an equal-variance signal-detection model with d' equal to 1.5. The two simulations differed only in how the 20-point rating scale was related to the underlying memory strength scale. In the first simulation, which is illustrated in the upper panel, the ratings were spread out on the weak end of the scale and compressed together on the strong end. As such, the difference in memory strength between ratings of 19 and 20 was small compared to the difference between ratings of 1 and 2. Thus, the rating scale did not have interval scale properties with respect to the psychological variable of interest (memory strength). The simulation involved drawing 150 memory strength values from the lure distribution and assigning a rating to each. Another 150 memory strength values were drawn from the target distribution, and ratings were assigned to them in the same way. The resulting rating data were then analyzed exactly as we analyzed the data presented above. That is, means and standard deviations were computed directly from the ratings and from ROC analysis. The ROC analysis yielded a standard deviation ratio close to the true value of 1 (namely, 1.05), which simply shows that ROC analysis is not dependent on the assumption of a linear measurement scale associated with the confidence ratings. By contrast, the direct ratings method yielded an answer that was far off the mark (0.63 in this case) because it *is* dependent on the assumption of a linear measurement scale associated with the ratings (and that scale is intentionally non-linear in this simulation).

The simulation was repeated using the equal-variance model shown in the lower panel of Figure 4. This time, the nonlinear relationship between scale ratings and memory strength was reversed such that the difference in memory strength between a rating of 19 and a rating of 20 was very large compared to the difference in memory strength between a rating of 1 and a rating of 2. Once again, the ROC analysis returned a ratio estimate close to the true value of 1 (0.97), but the ratings returned an answer that was far off the mark (1.50), this time in the other direction. While the slope estimate is very sensitive to the nature of the measurement scale, the estimates of d_r are much less affected.

INSERT FIGURE 4 ABOUT HERE

Experiment 2

Although the results of Experiment 1 showed a correspondence between the ratio estimates derived from ROC analysis and from direct ratings of memory strength, the strength of the relationship may have been reduced by some rating anomalies that we attempted to eliminate in a second experiment. The ratings for one subject, for example, were clearly influenced by the Old/New question that preceded the rating. This subject tended to avoid using the mid-range of the scale and gave fairly high ratings to all items declared to be Old and fairly low ratings to all items declared to be New. Thus, in Experiment 2, we eliminated the Old/New question and asked for ratings only, this time using a 1-to-99 rating scale.

Method

Participants

Sixteen undergraduates from University of California, San Diego participated for lower division psychology course credit.

Materials and design

The words, list length, duration of presentation were the same as those presented in Experiment 1.

Procedure

The procedure was the same as in Experiment 1 except that participants did not make an initial Old/New decision, they were informed that half of the words on the test were on the list presented and half were not, and they indicated the strength of their memory on a 1 to 99 scale.

Results

Scale biases were more apparent using the 1-99 scale. For example, subjects often supplied ratings at intervals of 5 on the scale, which means that, for them, this was effectively a 20-point scale, and there was a noticeable bias to choose the mid-point rating of 50 for both targets and lures. In addition, as in Experiment 1, the target distribution showed evidence of a ceiling effect, with 11.6% of the targets (and virtually none of the lures) receiving a rating of 99. Otherwise, the distribution and accuracy data were similar to the results of Experiment 1. The subjects were largely unbiased in the use of the scale such that their ratings for all items averaged together (targets and lures) was close to the mid-point of the scale, with the mean value being 50.99. All of the scores were symmetrically distributed about 50 (ranging from 38.63 to 60.60), with no apparent outliers.

Table 2 shows the mean and standard deviations for the ratings made to the targets and to the lures for each subject. Across all 16 subjects, the mean rating for the targets was 64.78 on the 99-point scale, and the mean rating for the lures was 37.20. Also shown for each subject is the ratio of the standard deviation of the lure ratings to the standard deviation of the target rating (i.e., $s_{\mathrm{lure}}/s_{\mathrm{target}}$). The mean ratio is 0.77, which is significantly less than 1.0, t(15) = 3.76. We next conducted an ROC analysis on these data by tabulating the number of responses to targets and number of responses to lures that exceeded the following cutoffs on the rating scale: 83, 67, 51, 33, 17, and 1. The estimated $\sigma_{\mathrm{lure}}/\sigma_{\mathrm{target}}$ ratio values for each subject are also shown in Table 2. The mean value of that ratio was also 0.77. Figure 5 shows the scatterplot of ratio measures derived from the two procedures, and the level of agreement is even higher than it was in Experiment 1 ($r = .83$, $p < .001$).

INSERT FIGURE 5 ABOUT HERE

Finally, as in Experiment 1, the values of d_a estimated from the ROC analysis were remarkably similar to the d_r values estimated directly from the ratings ($r > .99$, mean $d_a = 1.12$, mean $d_r = 1.18$). It has been argued that, in the unequal-variance situation, d_a is the single best estimate of discriminability (Macmillan & Creelman, 2005), but it has never been widely used in the recognition literature because an ROC analysis was needed to obtain an estimate of it. It seems that a simpler way to obtain that estimate is to compute it directly from ratings of memory strength -- ratings that are as easy to obtain as Old/New decisions are.

General Discussion

The two experiments reported here support a conclusion that is commonly drawn from ROC analysis, namely, that the memory strengths of the targets are more variable than the memory strengths of the lures. Using direct ratings of memory strength on a 1-to-20 scale or a 1-to-99 scale, we found that the standard deviation of the lure ratings was about .80 times the standard deviation of the target ratings. This is the predicted result given that the slope of the z-ROC is often approximately .80. Also, across individual subjects, ratio estimates derived from the direct ratings method correlated highly with ratio estimates derived from ROC analysis. These two methods are not constrained to agree, and they rely on different assumptions. The ROC analysis relies on the assumption that the memory strength distributions are

Gaussian in form. That assumption makes it possible to avoid the assumption that confidence ratings are made on a linear scale. The direct rating method, by contrast, assumes a linear scale and so avoids having to make any assumption about the mathematical form of the distribution of memory strengths. Even so, the level of agreement between the two methods is remarkably high.

The close agreement between the model-based ROC analysis and the model-free ratings method supports not only an unequal variance model but also the idea that the memory strengths are distributed in such a way that fitting a specifically Gaussian model to the data yields accurate conclusions (even if the true underlying distributions are not strictly Gaussian). However, as indicated earlier, the subjects tended to choose the highest rating for about 10% of the targets, which might indicate that the target distribution has a long tail that extends well beyond the highest rating. If so, the estimated difference in variance between the targets and lures based on the ratings (but not the ROC analysis) would have been even greater. In that case, both methods would still support an unequal-variance model, but they would not agree on the degree of inequality.

The fact that quite a few targets but almost no lures received the highest rating in both experiments is consistent with the idea that only recollection gives rise to the highest memory strengths (in that recollection is likely to be associated with targets, but not lures). On the surface, this pattern might appear to suggest that recollection is an all-or-none phenomenon, but evidence weighing against this idea can be found in source memory studies showing that lower degrees of confidence are associated with lower degrees of recollective accuracy, not the absence of recollection (e.g., Slotnick

& Dodson, 2005). Although we did not use a source memory procedure to test this idea here, it seems likely that varying degrees of recollective success were associated with different ratings of memory strength. As such, recollection is probably not represented solely in the highest rating even though especially strong recollection may be responsible for the fact that only targets tend to receive that rating.

As noted by Wixted (in press), although it might seem that an unequal variance model is inherently less plausible than the more aesthetically appealing equal-variance model, the opposite is actually true. The targets can be thought of as lures that have had memory strength added to them by virtue of their appearance on the study list. An equal-variance model would result if each item on the list had the exact same amount of strength added during study. However, if the amount of strength that is added differs across items, as it must, then both strength and variability would be added, and an unequal-variance model would apply. It is, of course, possible to imagine forces that would work against the increased variance (e.g., if the amount of strength added during study is inversely proportional to baseline strength). However, because few would dispute the notion that varying degrees of strength are added at study, it is actually the equal-variance model that is, a priori, the less plausible account. The ratings data reported here suggest that the more plausible unequal-variance account, which has long been supported by ROC analysis, is substantiated by direct ratings of memory strength.

Table 1.

Mean rating made by each subject to all test items ($m_{overall}$) in Experiment 1. Also shown are means and standard deviations of the ratings made to targets (m_{target} and s_{target}, respectively) and lures (m_{lure} and s_{lure}, respectively). The last two columns show the ratios of the lure standard deviation to the target standard deviation obtained directly from the ratings (s_{lure}/ s_{target}) and from a separate ROC analysis ($\sigma_{lure}/\sigma_{target}$) of the same data.

Subject	$m_{overall}$	m_{target}	m_{lure}	s_{target}	s_{lure}	s_{lure}/s_{target}	$\sigma_{lure}/\sigma_{target}$
1	12.09	15.34	8.83	5.55	6.19	1.12	0.91
2	9.83	13.09	6.57	6.35	4.30	0.68	0.76
3	10.10	13.11	7.08	4.87	3.75	0.77	0.64
4	11.87	14.68	9.05	3.78	2.40	0.63	0.56
5	10.32	12.28	8.35	4.34	2.81	0.65	0.55
6	10.71	14.45	6.97	5.47	5.61	1.03	0.87
7	8.89	10.47	7.31	5.04	3.65	0.72	0.76
8	8.93	11.90	5.95	5.35	3.60	0.67	0.75
9	10.29	10.73	9.85	2.31	2.39	1.03	1.06
10	10.89	12.89	8.89	4.26	3.71	0.87	0.83
11	14.17	17.61	10.73	3.51	4.54	1.30	0.73
12	11.79	15.80	7.78	4.59	4.85	1.06	0.74
13	9.55	10.47	8.62	4.43	3.61	0.81	0.88
14	11.37	13.51	9.23	3.77	2.96	0.79	1.01
Mean	**10.77**	**12.98**	**8.04**	**4.62**	**3.83**	**0.83**	**0.79**

Note: Except for the first column, Subject 11's scores were excluded from the mean values

Table 2.

Mean rating made by each subject to all test items (m_{overall}) *in Experiment 2. Also shown are means and standard deviations of the ratings made to targets* (m_{target} *and* s_{target}, *respectively) and lures* (m_{lure} *and* s_{lure}, *respectively). The last two columns show the ratios of the lure standard deviation to the target standard deviation obtained directly from the ratings* ($s_{\text{lure}}/ s_{\text{target}}$) *and from a separate ROC analysis* ($\sigma_{\text{lure}}/\sigma_{\text{target}}$) *of the same data.*

Subject	m_{overall}	m_{target}	m_{lure}	s_{target}	s_{lure}	$s_{\text{lure}}/s_{\text{target}}$	$\sigma_{\text{lure}}/\sigma_{\text{target}}$
1	56.62	64.63	48.60	18.36	12.63	0.69	0.81
2	51.92	67.62	36.21	24.79	17.49	0.71	0.66
3	43.91	70.27	17.54	36.26	23.42	0.65	0.56
4	53.47	66.23	40.71	32.74	33.76	1.03	0.95
5	44.88	54.07	35.68	23.84	19.16	0.80	0.78
6	56.34	68.79	43.88	28.00	27.91	1.00	0.81
7	58.09	65.27	50.91	20.59	18.39	0.89	0.82
8	44.32	73.07	15.57	30.56	14.14	0.46	0.63
9	60.60	79.31	41.89	22.55	23.29	1.03	0.79
10	43.49	62.11	24.87	33.48	21.00	0.63	0.62
11	51.40	55.14	47.66	40.70	41.24	1.01	1.18
12	56.99	65.33	48.65	18.29	6.60	0.36	0.55
13	38.63	51.33	25.92	28.89	12.89	0.45	0.70
14	60.35	74.58	46.11	20.53	25.22	1.23	1.11
15	49.15	58.50	39.80	19.05	14.31	0.75	0.75
16	45.71	60.22	31.19	31.11	18.40	0.59	0.64
Mean	**50.99**	**64.78**	**37.20**	**26.86**	**20.61**	**0.77**	**0.77**

Figure Captions

Figure 1. Equal-variance (upper panel) and unequal-variance (lower panel) signal-detection models of recognition memory.

Figure 2. Upper panel. Frequency distribution showing the number of responses made to targets and lures pooled over participants. Lower panel. Accuracy associated with each rating based on the pooled data in the upper panel.

Figure 3. Scatterplot (and regression line) of lure-to-target standard deviation ratio estimates (ROC estimate vs. direct ratings estimate) from Experiment 1.

Figure 4. Hypothetical signal-detection models illustrating two non-linear relationships between a 20-point rating scale and the memory strength scale.

Figure 5. Scatterplot (and regression line) of lure-to-target standard deviation ratio estimates (ROC estimate vs. direct ratings estimate) from Experiment 2.

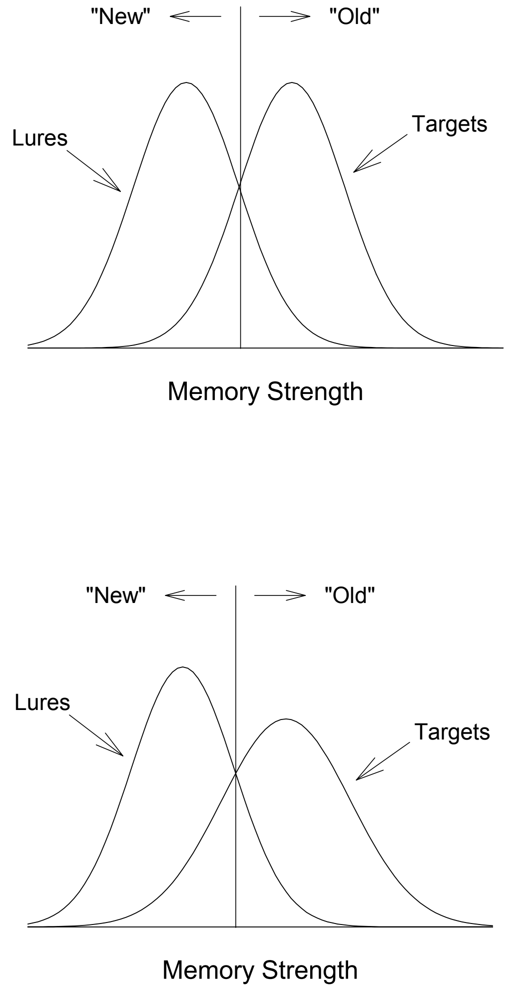


Figure 1. Equal-variance (upper panel) and unequal-variance (lower panel) signal-detection models of recognition memory.

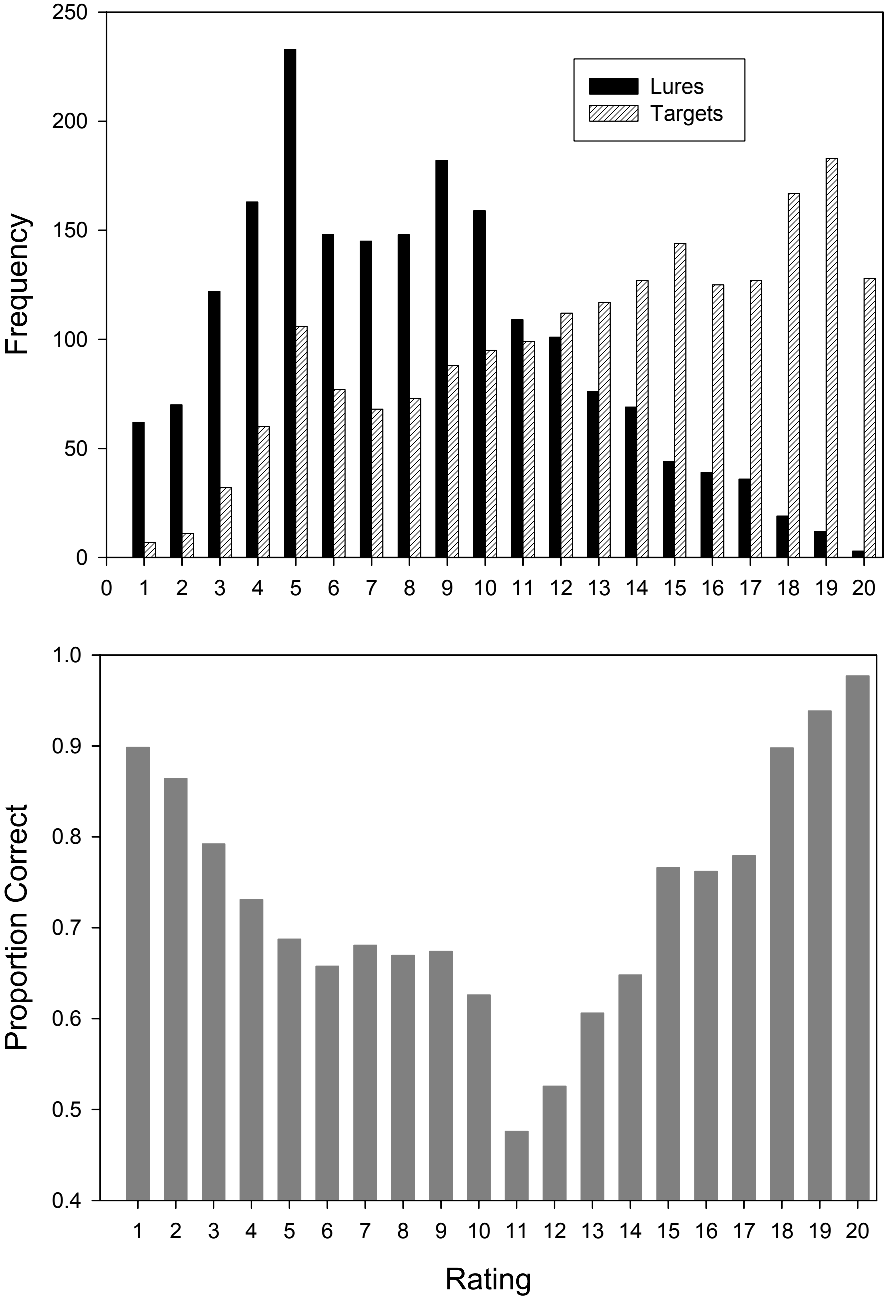


Figure 2. Upper panel. Frequency distribution showing the number of responses made to targets and lures pooled over participants. Lower panel. Accuracy associated with each rating based on the pooled data in the upper panel.

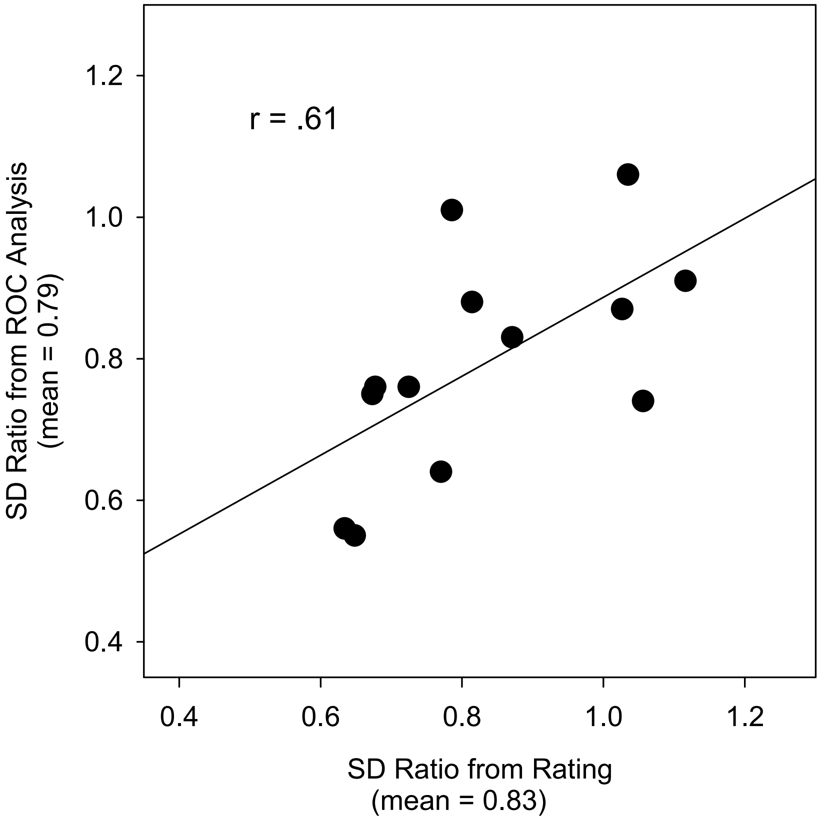


Figure 3. Scatterplot (and regression line) of lure-to-target standard deviation ratio estimates (ROC estimate vs. direct ratings estimate) from Experiment 1.

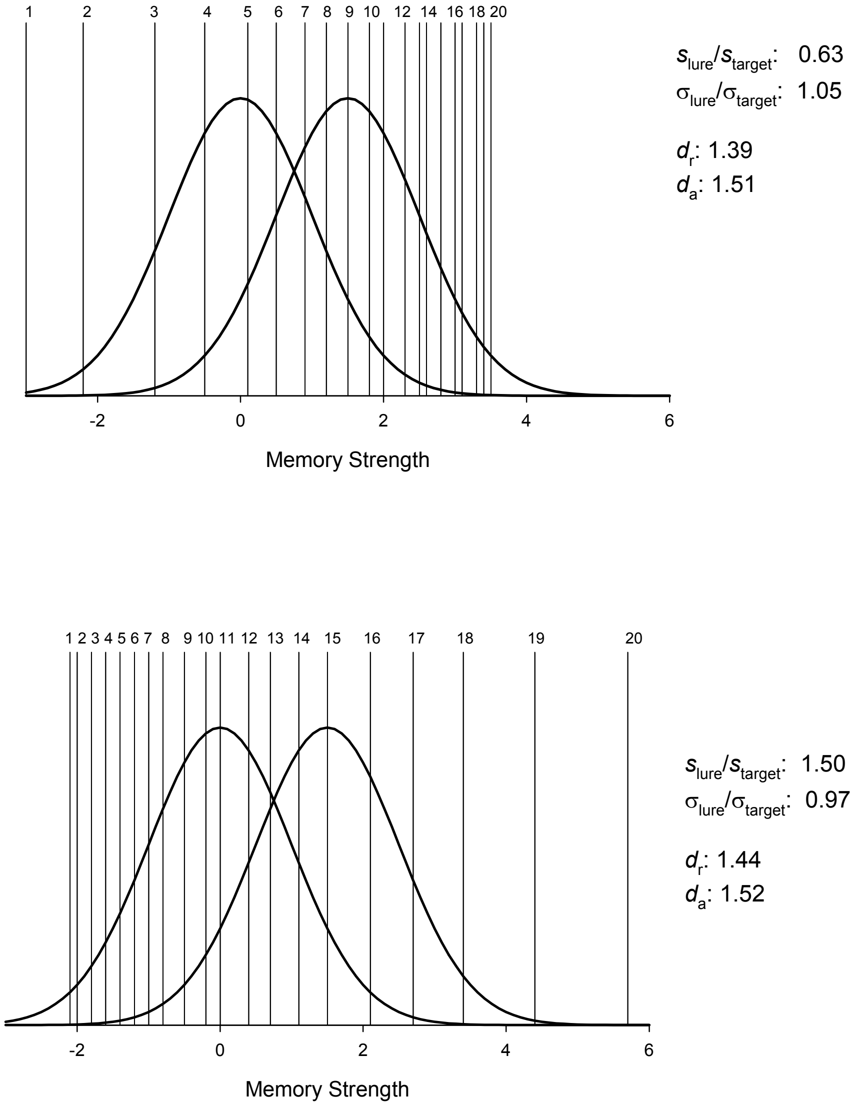


Figure 4. Hypothetical signal-detection models illustrating two non-linear relationships between a 20-point rating scale and the memory strength scale.

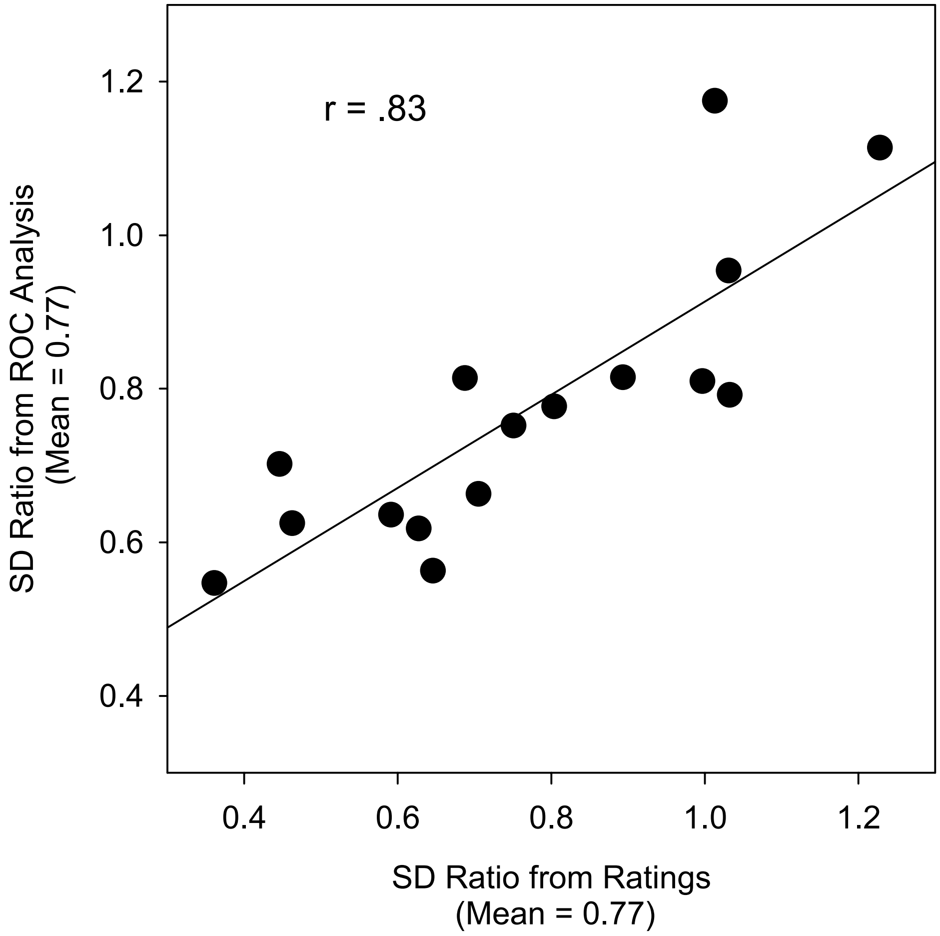


Figure 5. Scatterplot (and regression line) of lure-to-target standard deviation ratio estimates (ROC estimate vs. direct ratings estimate) from Experiment 2.

CHAPTER 2. Recollection is a Continuous Process: Implications for Dual Process Theories of Recognition Memory

Mickes, L., Wais, P. E., & Wixted, J. T. (2009). *Psychological Science, 20,* 509-515.

Abstract

Dual process theories hold that recognition decisions can be based on recollection or familiarity. Such theories have long seemed incompatible with signal-detection theory, which holds that recognition decisions are based on a singular, continuous memory strength variable. Although dual-process theories regard familiarity as a continuous process (i.e., familiarity comes in degrees), they construe recollection as a categorical process (i.e., recollection either occurs or does not occur). A continuous process is characterized by a graded relationship between confidence and accuracy, whereas a categorical process is characterized by a binary relationship such that high confidence is associated with high accuracy but all lower degrees of confidence are associated with chance accuracy. Using a source memory procedure, we found that the relationship between confidence and source recollection accuracy was graded. Because recollection, like familiarity, is a continuous process, dual-process theory is more compatible with signal-detection theory than previously thought.

Recollection is a Continuous Process:

Implications for Dual Process Theories of Recognition Memory

A longstanding theory holds that recognition memory decisions are supported by two processes, namely, recollection and familiarity. The following anecdote, offered by Mandler (1980), describes a common experience that illustrates how these two processes sometimes unfold in real time:

> Consider seeing a man on a bus whom you are sure that you have seen before; you "know" him in that sense. Such a recognition is usually followed by a search process asking, in effect, Where could I know him from? Who is he? The search process generates likely contexts (Do I know him from work; is he a movie star, a TV commentator, the milkman?). Eventually the search may end with the insight, That's the butcher from the supermarket! (pp. 252-253)

The initial sense of familiarity refers to a memory signal pertaining to the item itself (based, perhaps, on its perceptual features), whereas the subsequent awareness of recollection refers to the retrieval of source information that is associated with that item. Familiarity is widely assumed to be a continuous process in the sense that it is experienced in *degrees*. Low degrees of familiarity are associated with low confidence and low accuracy, whereas high degrees of familiarity are associated with high confidence and high accuracy. By contrast, the recollection process is almost always thought to be categorical in nature in that, theoretically, it either occurs (yielding high confidence and high accuracy) or does not occur.

For continuous processes, the notion of a *decision criterion* almost inescapably comes into play. Thus, for example, on a typical Old/New recognition memory test, the participant's task is to distinguish between targets (i.e., the items that were presented on the list) and lures (i.e., the items that were not presented on a prior list). Although the targets are likely to be relatively familiar because of their recent appearance on a list, the lures will be associated with some degree of familiarity as well. Thus, for a decision that is based on familiarity, a participant must decide *how much* familiarity is enough to decide that the item is Old. In other words, the participant must set a criterion familiarity value.

An early dual-process model proposed by Atkinson and colleagues envisioned two criteria for familiarity-based decisions (Atkinson & Juola, 1973, 1974). According to this model, if the degree of familiarity associated with a test item was strong enough to fall above a high criterion or weak enough to fall below a low criterion, then a familiarity-based decision would be made (Old or New, respectively). If the degree of familiarity instead fell between the two criteria (i.e., if familiarity was of intermediate strength), then a retrieval search would be initiated. That search was assumed to either succeed (in which case the item was declared to be old) or fail (in which case it was declared to be new). Thus, in this model, recollection was construed as a categorical process – one that does not involve a decision criterion. Mandler (1980) also pointed out that a decision criterion plays a role in familiarity-based decisions, but no such considerations were brought to bear on recollection, which, again, was treated as a categorical process that either succeeded or did not succeed. The same considerations apply to the way in which recollection and familiarity are construed in studies that use the process dissociation procedure to obtain quantitative estimates of recollection and familiarity (Jacoby, 1991). In computing those estimates, recollection is again considered to be a categorical process, whereas familiarity is assumed to be a continuous process that involves a decision criterion (Jacoby, Toth & Yonelinas, 1993). Finally, Yonelinas (1994) proposed a model in which recollection was assumed to be a categorical process that always yields high confidence and does not involve a decision criterion, whereas familiarity was regarded as a continuous signal-detection process that does involve a decision criterion.

A common feature of all of these dual-process models, in addition to the fact that they regard recollection as a categorical process, is that they assume that individual recognition decisions are based either on one process or on the other. That is, according to all of these models, old/new decisions about items that elicit recollection are based solely on recollection, whereas old/new decisions about items that do not elicit recollection are based solely on familiarity. This is a natural way to think if one begins with the assumption that recollection is categorical in nature. That is, in the categorical view, the occurrence of recollection would yield high confidence that an item was previously encountered, thereby rendering unnecessary any consideration of familiarity. But when recollection fails completely, the only recourse would be to rely on familiarity.

An alternative view is that recollection and familiarity are both continuous processes that are aggregated into a memory strength signal (Wixted, 2007). According to this account, both processes play a role in Old/New decisions about individual test items. The core difference between this model and all earlier dual-process models is its assumption that recollection is a continuous process (i.e., that recollection comes in degrees), not a categorical process. If recollection were a continuous process (e.g., Dodson, Holland, & Shimamura, 1998; Johnson, Hashtroudi, & Lindsay, 1993), then high degrees of recollection would result in high confidence and high accuracy, but low degrees of recollection would result in low confidence and low accuracy. In that respect, recollection would be like familiarity. In other respects, however, the two processes would remain distinct. That is, according to this view, familiarity is a fast process that involves the retrieval of information about the item per se, whereas recollection is a slower process that involves the retrieval of associated contextual information. But because recollection is assumed to occur in graded fashion, any degree of recollection that happens to occur would add to the extant familiarity-based memory signal instead of usurping it.

The present research is concerned with differentiating between the categorical and continuous views of recollection. To investigate this issue, participants were exposed to a source memory procedure (Johnson et al., 1993), which is commonly used to study the recollection process. In this procedure, some items on a list are associated with one source attribute (e.g., the color red) and others are associated with a different source attribute (e.g., the color blue). On a later recognition test, the participants are presented with test items in source-neutral fashion (e.g., in black) and asked to recollect the original source attribute. Instead of asking for a binary Source A vs. Source B decision, participants in this experiment were asked to rate their confidence in the item's source using a 20-point scale, with 1 representing highest confidence in Source A (e.g., blue) and 20 representing highest confidence in Source

B (e.g., red). In a test like this, the familiarity of the test item is not diagnostic of its source because the items from both sources recently appeared on the same study list.

The categorical and continuous views of recollection make contrasting predictions about the relationship between the confidence in a source decision and the accuracy of that decision. The categorical view of recollection predicts that the relationship will be a step function. For example, an all-or-none recollection version of the categorical model predicts that accuracy will be high for ratings made with the highest confidence (i.e., for ratings of 1 or 20) and will be no better than chance for all other ratings. By contrast, the continuous view of recollection predicts that the relationship will fall off in graded fashion (i.e., accuracy will be highest for ratings of 1 or 20, next highest for ratings of 2 and 19, and so on).

In prior investigations into this issue, participants were first asked for an old/new decision using a 6-point confidence scale and then asked for a source decision. Yonelinas (2001) reported that source accuracy was above chance only for old/new decisions that were made with the highest level of confidence (consistent with the categorical view of recollection), but Wixted (2007) reviewed results from several other studies showing that source recollection was above chance even for old/new decisions that were made with low and medium levels of confidence (consistent with the continuous view of recollection). In the present experiment, we tested the relationship between confidence and accuracy for the source decision itself to directly test the categorical vs. continuous accounts, and we used a 20-point scale to examine the relationship over a wide range of confidence.

Method

Participants

The participants were 91 college undergraduates, who were recruited from the university experimental participants pool, gave their informed consent according to the university IRB protocol, and received class credit for completing our experiment. All participants were fluent in English.

Stimuli

The word pool used consisted of 705 three-to-seven letter words extracted from the MRC Psycholinguistic Database (Coltheart, 1981), of which 300 words were randomly selected for testing. Instructions and stimuli were presented using E-Prime 1.1.4.1 (www.pstnet.com; © Psychology Software Tools, Inc.) scripts on a Dell Dimension 4550 desktop computer and 17-inch monitor. A second word pool of 1018 words drawn from Gilhooly and Logie (1980) was used for testing 6 participants over an especially large number of trials.

Procedure

Two similar versions of the experiment were run. In one, the relevant source attribute was font color, and in the other it was screen location. Two versions were run because some neuroimaging evidence suggests that recollecting a feature of the item itself, such as its color, may differ from recollecting an extra-item detail, such as its location (Staresina & Davachi, 2006). Participants were informed that a list of 100 words would be presented in red or blue (Version 1, n = 49) or at the top or bottom of the screen (Version 2, n = 36) for 2 seconds each, and they were advised that their memory for color (or location) would be tested after list presentation was complete.

Participants first completed a brief practice session to ensure they understood the task. On the subsequent recognition test, items were presented one at a time in black (on a white background) at the center of the screen for a source decision (red or blue in Version 1, top or bottom in Version 2) using a 20-point rating scale. On this scale, 1 indicated 100% certainty that the item was presented in blue (or at the bottom of the screen), and 20 indicated 100% certainty that the item was presented in red (or at the top of the screen). Lesser degrees of certainty were indicated using less extreme numbers, with ratings of 10 and 11 indicating choices of blue or red (or bottom or top), respectively, made with complete uncertainty. An additional 6 participants were tested in the color memory version of the experiment using lists of 200 words in each of 5 sessions so that their individual confidence-accuracy functions could be examined.

Results

In the color version of the task, overall source recollection accuracy (67.1%) was significantly above chance, $t(48) = 13.1$, $p_{rep} > .99$. The question of most interest concerns the relationship between confidence and accuracy. The 20-point rating scale provided 10 levels of confidence in "blue" decisions (where 1 = highest confidence that the item had been presented in blue, and 10 = lowest confidence) and 10 levels of confidence in "red" decisions (where 20 = highest confidence that the item had been presented in red, and 11 = lowest confidence). Thus, for purposes of analysis, the 20-point rating scale was converted to a 10-point confidence scale in which a value of 1 corresponds to ratings of 10 or 11 (lowest confidence in blue and red decisions, respectively), a value of 2 corresponds to the next highest ratings (9 and 12), and so on up to a value of 10, which corresponds to the highest ratings (1 and 20). For each participant, recollection accuracy was computed for each level of confidence. As shown in Figure 1, accuracy was no better than chance for confidence ratings of 1 through 4, but it was marginally greater than chance for ratings of 5 and 6. Accuracy was higher still (and was significantly greater than chance) for ratings of 7, 8 and 9, and it was far above chance for ratings of 10. For the highest level of confidence (10), accuracy significantly exceeded that for the next lowest level (9), $t(38) = 3.73$, $p_{rep} = .99$.

In the location version of the task, overall source recollection accuracy (77.1%) was also significantly above chance, $t(35) = 11.4$, $p_{rep} > .99$. The accuracy scores for each level of confidence were more variable than on the color version of the task, so the 10-point confidence scale was reduced to a 5-point confidence scale by averaging together adjacent confidence levels. As shown in Figure 2, recollection again increased in continuous fashion as confidence increased. Performance was no better than chance for confidence levels of 1 and 2, but it was clearly greater than chance (falling at approximately 75% correct) for confidence levels of 3 and 4. For the highest level of confidence (5), accuracy exceeded 90% correct and significantly exceeded accuracy for the next lowest level (4), $t(32) = 2.94$, $p_{rep} = .96$.

The results shown in Figures 1 and 2 are inconsistent with a categorical view according to which recollection always yields the highest level of confidence. However, a different view of recollection might hold that although recollection is categorical, it does not always yield the highest level of confidence. That is, when recollection for Source A occurs, a participant might provide ratings between, say, 1 and 3, whereas when recollection for Source B occurs, the participant might provide ratings between 18 and 20. When converted to a 10-point scale, a plot of the relationship between confidence and accuracy for a participant like this would reveal a step-function, with accuracy for ratings of 8 through 10 being very high and all other ratings being associated with chance performance. If data from different participants yielded different categorical breakpoints along the confidence rating scale, then averaging over participants would create the false impression of a continuous relationship.

To investigate this possibility, we also collected enough data from six individual participants to assess the confidence-accuracy relationship at the individual level (Figure 3). If the confidence-accuracy plot is characterized by a step function, then the slope of a line fit through the possible below-threshold accuracy values should, on average, be zero, and accuracy for these values should be close to chance. Similarly, the slope of a line fit through the possible above-threshold values should also be zero, but at a level much greater than chance. Figure 3 shows candidate threshold step-functions for each participant, with possible below-threshold values indicated by light gray bars and possible above-threshold values indicated by dark gray bars. Also shown are straight lines that were separately fit to the presumptive below- and above-threshold data. No line could be fit to the single above-threshold value for Participant 2, whose overall pattern is clearly consistent with an all-or-none recollection model. However, of the remaining 11 sets of above- and below-threshold fits, all but two have positive slopes. For each participant, the above-threshold and below-threshold slopes were averaged together (except for Participant 2, whose below threshold slope was used) to obtain the most reliable slope estimate. The average slope was positive for 5 of the 6 participants (all except Participant 2) and was significantly greater than zero, t(5) = 2.66, p_{rep} = .92. These conclusions do not change if alternative breakpoints are used (e.g., with the first above-threshold value set to confidence level 7 for Participant 1, to 6 for Participant 3, or to 8 for Participant 6). Thus, although the data do not rule out the possibility that a subset of subjects experience categorical recollection, the average plot of the relationship between confidence and accuracy shown in Figures 1 and 2 would appear to be representative of the majority of individual participants.

Evidence for threshold recollection on source memory tasks is often based on an analysis of Receiver Operating Characteristic (ROC) data. A continuous signal-detection model of recollection predicts a curvilinear ROC and a linear z-ROC, whereas a categorical recollection model instead predicts a linear ROC and a curvilinear z-ROC. For source recollection tasks, the ROC is typically curvilinear, but the z-ROC is often curvilinear as well (Yonelinas & Parks, 2007). The curvilinearity of the z-ROC has been taken as evidence that recollection is a categorical process. Figure 4 shows the z-ROC from both versions of this experiment, and they both exhibit the curvilinearity that is often seen on source memory tasks. However, the curvilinearity is apparent *even though* recollection is a continuous process (as shown in Figures 1 and 2), not because it is a categorical process. As discussed later, this curvilinear anomaly does not seem to imply categorical recollection except insofar as memory can be so weak that, for some items, source information is completely absent.

Discussion

All of the major dual-process models of recognition memory hold that familiarity is a continuous process (one that involves a decision criterion), whereas recollection is a categorical process (one that does not involve a decision criterion). The results shown in Figures 1 and 2 suggest instead that recollection, like familiarity, is a continuous process. If recollection and familiarity are both continuous processes, then it stands to reason that both processes jointly contribute to individual recognition decisions. In fact, unless they provided completely redundant information, an efficient memory system would (either by design or by learning) combine them to yield an aggregated memory strength signal. An aggregated signal like that would be more diagnostic of prior occurrence than either signal alone.

The idea that individual recognition decisions are based on an aggregated memory strength signal has potentially far reaching implications. First, the traditional signal-detection view of recognition memory involves two unequal-variance Gaussian distributions and a decision criterion (Figure 5). This longstanding model – with its singular memory strength axis – has been widely regarded as being at odds with the similarly longstanding dual-process model of recognition. However, if continuous recollection and familiarity signals are combined into a memory strength variable, then these two venerable models are naturally reconciled (Wixted, 2007). This view implies that a decision criterion is just as relevant to the recollection process as it is to the familiarity process.

Second, much neuroimaging research that seeks to identify the neural correlates of recollection and familiarity is guided by dual-process models that implicitly or explicitly reject the signal-detection model of recognition memory. That is, they are predicated on the assumption that individual recognition decisions are based either on categorical recollection or on continuous familiarity (never on both processes together). If that assumption is wrong, then the results of these studies would need to be reinterpreted. For example, using a 6-point confidence scale (1 = Sure New through 6 = Sure Old), Daselaar, Fleck and Cabeza (2006) found that activity in the posterior hippocampus was similar for confidence ratings of 1 through 5 but was significantly elevated for confidence ratings of 6. Based on the assumption that recollection is a categorical all-or-none process that always yields the highest level of confidence, these authors concluded that the posterior half of the hippocampus selectively subserves the recollection process. But if recollection is associated with lower degrees of confidence as well, then these results would instead suggest that activity in the posterior hippocampus is detectable when memory is strong, not when it selectively involves recollection (Squire, Wixted & Clark, 2007). A similar reinterpretation would apply to many studies that have used confidence ratings or the related Remember/Know procedure to identify the neural correlates of recollection and familiarity (e.g., Eldridge, Knowlton, Furmanski, Bookheimer & Engel, 2000; Montaldi, Spencer, Roberts & Mayes, 2006; Vilberg & Rugg, 2007; Yonelinas, Otten, Shaw & Rugg, 2005).

Because continuous memory processes are generally well characterized by signal-detection theory, and because signal-detection theory predicts a linear z-ROC, the curvilinear z-ROCs evident in Figure 4 (which are typical of source memory procedures) would seem to be an unexpected result. A model that assumes that recollection is a categorical high-confidence-or-none process predicts this z-ROC anomaly, and such evidence has often been taken to support that model (e.g., Yonelinas & Parks, 2007). However, the evidence summarized in Figures 1 and 2, which show that recollection is a continuous process, suggests that some other explanation applies. Indeed, recent work on the shape of source memory ROCs has provided that explanation. Specifically, Slotnick and Dodson (2005) showed that the curvilinear shape of the source memory z-ROC is a consequence of the fact that, for items with weak old/new memory (i.e., for decisions made with low levels of old/new confidence), source information is absent. The standard signal-detection model of source memory includes no provisions for items like these (i.e., items for which no degree of source information is available), so the prediction of a linear z-ROC does not apply when those items are included in the analysis. When such items are excluded from the source ROC analysis, the shape of the ROC corresponds closely to the predictions of signal detection theory (Slotnick & Dodson, 2005). Critically, this is true even though all strong recollection-based decisions (i.e., all old/new decisions made with high confidence) remained in the source ROC analysis. These findings show that recollection-based ROCs are fully compatible with signal-detection theory and that the curvilinear z-ROC often found with source memory procedures arises because of the inclusion of weak items with no source memory, not because recollection is a threshold process that always yields high confidence.

Finally, despite the common assumption that source memory procedures tap recollection, it could be argued that the continuous relationship between confidence and accuracy means that source decisions are based on familiarity. For example, it could be argued that, at test, participants rely on a generate-recognize strategy by mentally simulating each test item, first as originating from Source A (e.g., in red) and then as originating from Source B (e.g., in blue). To make a decision, the participant might choose the imagined source that yields the higher feeling of familiarity. Although generate-recognize theory has been mostly abandoned as an explanation of recall (e.g., Tulving & Thompson, 1973), its possible role in source memory procedures has not been ruled out. If source memory decisions are based on familiarity, then neuroimaging studies that rely on source memory procedures to study the neural correlates of recollection would need to be reconsidered. Moreover, the curvilinear z-ROCs shown in Figure 4, which are often attributable to categorical recollection, would also need to be explained in some other way (e.g., as being based on a combination of continuous source-familiarity and discontinuous source-recollection).

The most parsimonious interpretation of the present results is that recollection is a continuous process, one that can be associated with low levels of confidence and accuracy (contrary to all prior dual-process accounts) as well as high levels of confidence and accuracy. It seems reasonable to suppose that recollection underlies the strongest possible memories (as all dual-process models would stipulate), but the key point of departure here is that recollection also plays a role in weaker memories. This interpretation is consistent with evidence from the Remember/Know procedure in which Know responses, which are often thought to reflect familiarity-based decisions made with high confidence, are instead associated with low confidence and above-chance levels of source recollection (Wais, Mickes, & Wixted, 2008). It is also consistent with recent single-unit recording evidence suggesting that recollection and familiarity are summed by neurons in the hippocampus (Rutishauser, Schuman & Mamelak, 2008) and with recent modeling evidence directly testing the idea that the memory signal for individual items is based on both recollection and familiarity (Starns & Ratcliff, in press). If recollection is a continuous process, and if the recollection and familiarity signals are aggregated into a unidimensional memory strength variable, then dual-process theory and signal-detection theory are naturally compatible accounts.

Figure Captions

Figure 1. Proportion correct as a function of confidence in the color version of the source memory task. The error bars represent 95% confidence intervals.

Figure 2. Proportion correct as a function of confidence in the location version of the source memory task. The error bars represent 95% confidence intervals.

Figure 3. Proportion correct as a function of confidence in the color version of the source memory task for 6 individuals who were tested over a large number of trials. Possible below-threshold values are indicated by light gray bars, and possible above-threshold values are indicated by dark gray bars. The straight lines represent least-squares fits to the presumptive below- and above-threshold data.

Figure 4. z-ROC data for the color version of the task (upper panel) and location version of the task (lower panel). The curvilinearity of both ROCs is visually apparent and is also evident in the fact that the quadratic coefficient of the best-fitting 2^{nd}-order polynomial is positive (and is equal to 0.20 in both cases).

Figure 5. An illustration of the standard unequal-variance signal-detection model of recognition memory. In the illustrated version of the model, memory strength is construed as a joint function of recollection and familiarity.

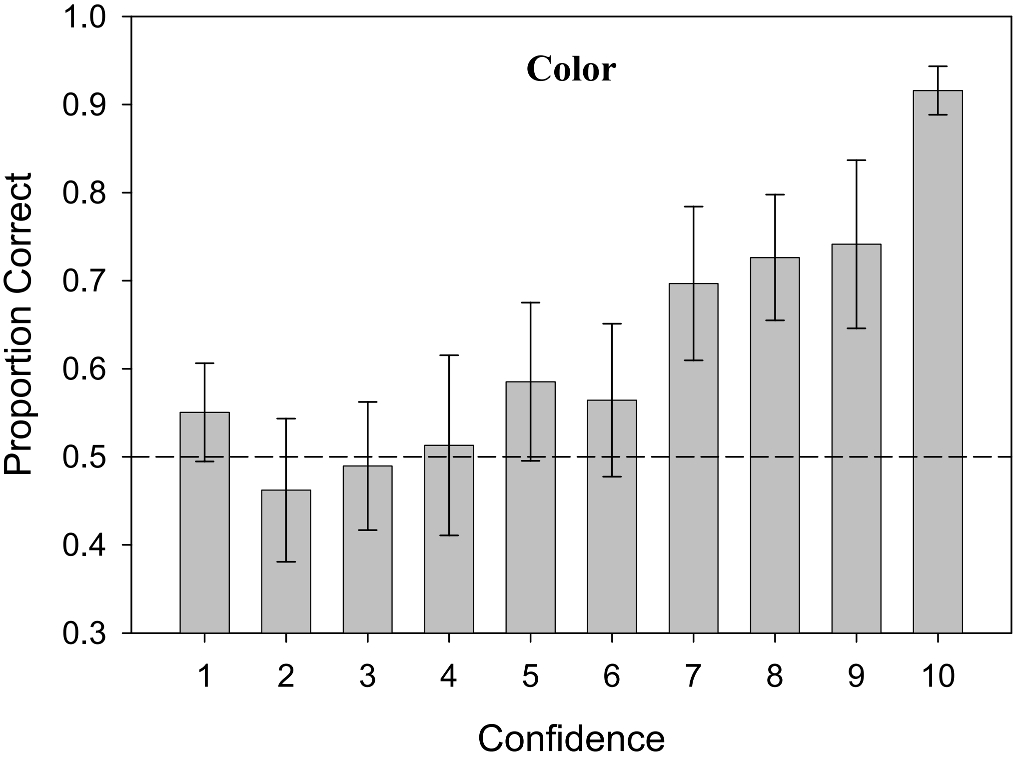


Figure 1. Proportion correct as a function of confidence in the color version of the source memory task. The error bars represent 95% confidence intervals.

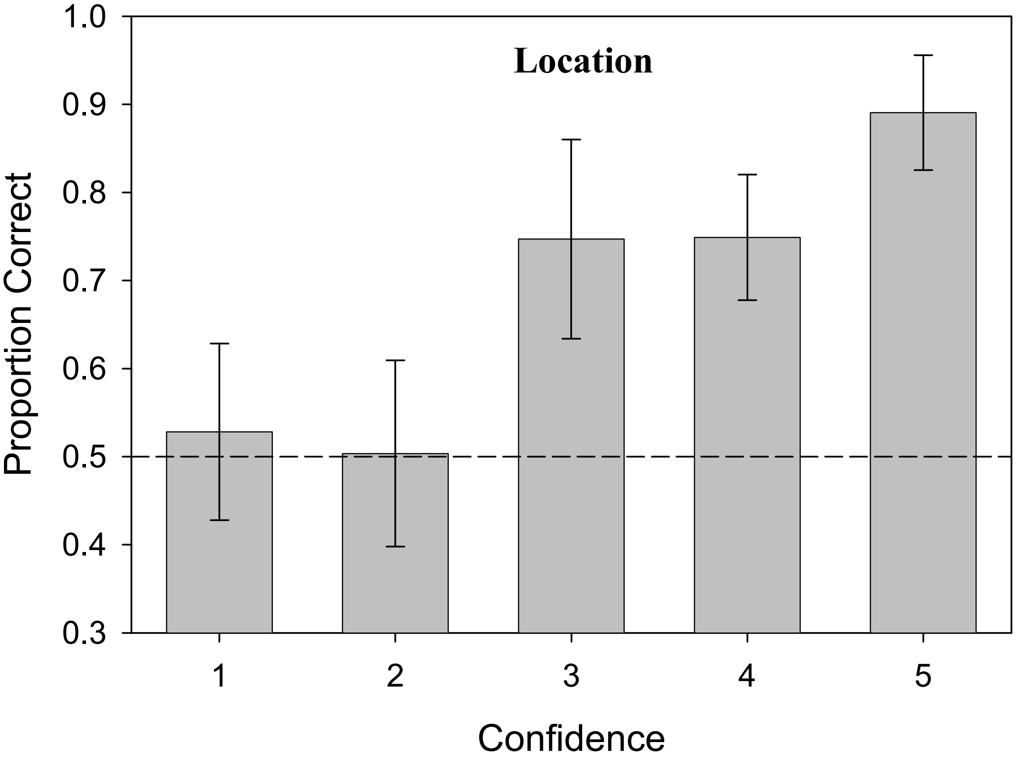


Figure 2. Proportion correct as a function of confidence in the location version of the source memory task. The error bars represent 95% confidence intervals.

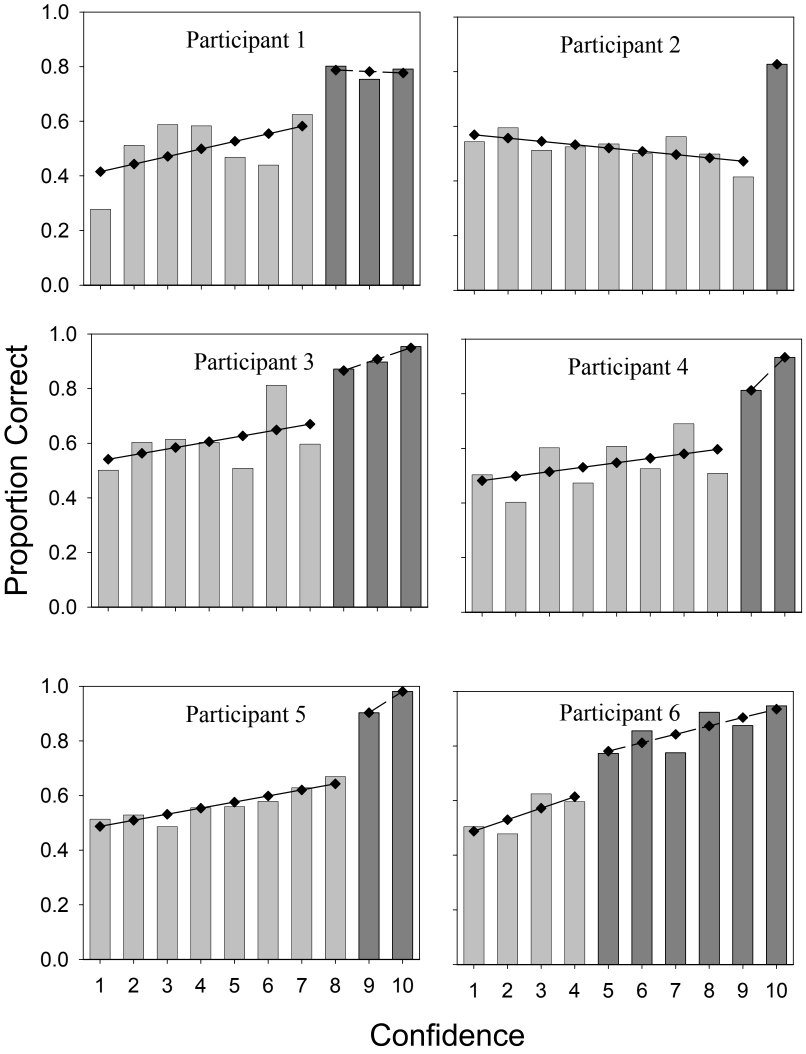


Figure 3. Proportion correct as a function of confidence in the color version of the source memory task for 6 individuals who were tested over a large number of trials. Possible below-threshold values are indicated by light gray bars, and possible above-threshold values are indicated by dark gray bars. The straight lines represent least-squares fits to the presumptive below- and above-threshold data.

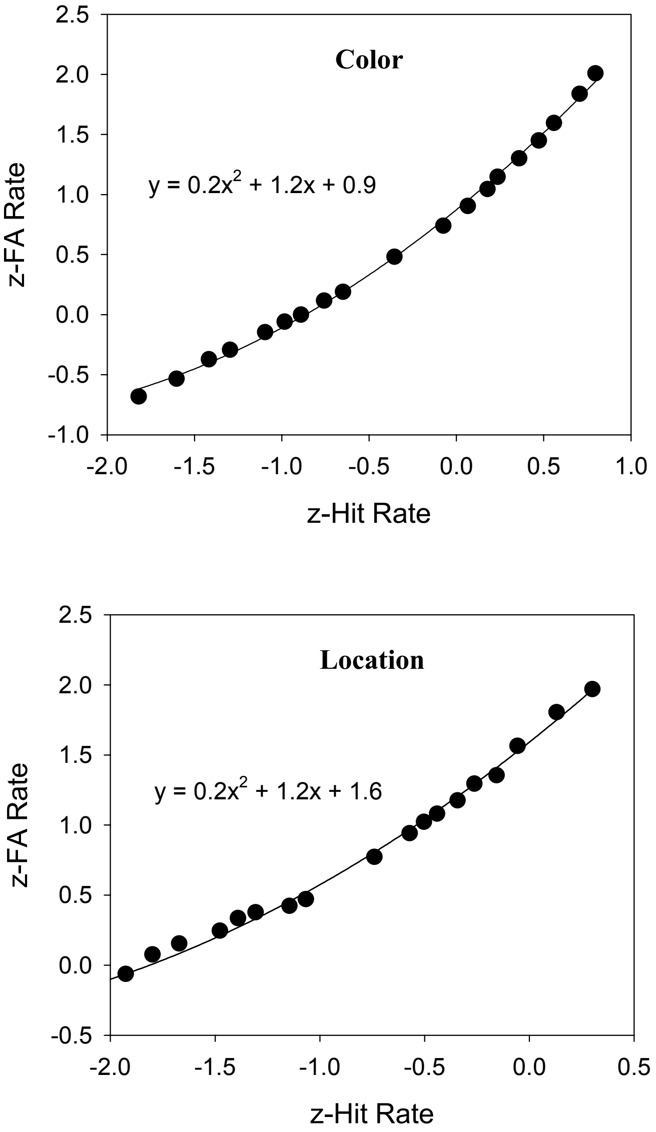


Figure 4. z-ROC data for the color version of the task (upper panel) and location version of the task (lower panel). The curvilinearity of both ROCs is visually apparent and is also evident in the fact that the quadratic coefficient of the best-fitting 2nd-order polynomial is positive (and is equal to 0.20 in both cases).

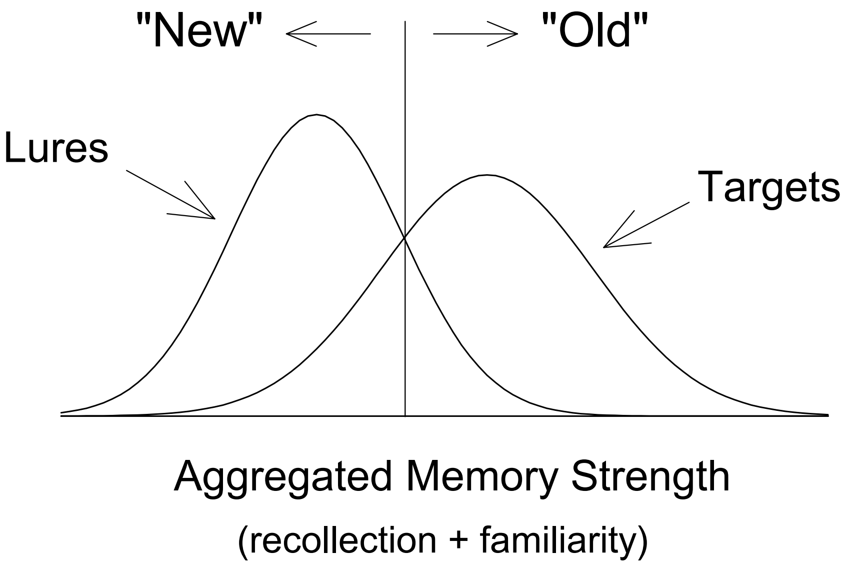


Figure 5. An illustration of the standard unequal-variance signal-detection model of recognition memory. In the illustrated version of the model, memory strength is construed as a joint function of recollection and familiarity.

Chapter 3. Continuous Recollection vs. Unitized Familiarity in Associative Recognition.

Mickes, L., Johnson, E. M., & Wixted, J. T. (in press). *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

Abstract

Recollection has long been thought to play a key role on associative recognition tasks. Evidence that associative recollection might be a threshold process has come from analyses of the associative recognition Receiver Operating Characteristic (ROC). Specifically, the ROC is not as curvilinear as a signal-detection theory requires. In addition, the z-ROC is usually curvilinear, as a threshold recollection model requires, not linear, as a signal-detection model requires. In Experiment 1, word pairs were strengthened at study, which yielded a curvilinear ROC and a linear z-ROC (in accordance with signal-detection theory). This result suggests that associative recognition performance was based on a continuous variable, one that likely consists either of unitized familiarity or continuous recollection. The Remember/Know procedure and an unexpected cued recall test suggested that the more curvilinear ROC in the strong condition was mainly due to increased recollection. In Experiment 2, word pairs were presented for an old/new recognition decision before being presented for an associative recognition decision. When pairs consisting of items not recognized as having been seen on the list were removed from the analysis, the ROC again became curvilinear, the z-ROC again

became linear, and most associative recognition decisions were associated with Remember judgments. These findings suggest that the curvilinear z-ROC often observed on associative recognition tests results from noise, as a mixture signal-detection model assumes, and that recollection is a continuous process that yields a curvilinear ROC that is well characterized by signal-detection theory.

Continuous Recollection vs. Unitized Familiarity in Associative Recognition

Recognition memory decisions are widely thought to be based either on the recollection of specific contextual details or on a context-free sense of familiarity (Atkinson & Juola, 1973; Jacoby, 1991; Mandler, 1980; Yonelinas, 2002). To investigate the nature of the recollection process, associative recognition procedures are commonly used, and this is true in both experimental psychology (e.g., Clark, 1992; Clark & Hori, 1995; Hockley, 1992; Hockley & Consoli, 1999; Hockley & Cristi, 1996; Nobel & Shiffrin, 2001; Verde, 2004; Xu & Malmberg, 2007; Yonelinas, 1997) and cognitive neuroscience (e.g., Giovanello, Verfaellie, & Keane, 2003; Habib & Nyberg, 2008; Haskins, Yonelinas, Quamme, & Ranganath, 2008; Sauvage, Fortin, Owens, Yonelinas, & Eichenbaum, 2008; Speer & Curran, 2007; Turriziani, Fadda, Caltagirone, & Carlesimo, 2004; Stark, Bayley, & Squire, 2002). In a typical version of this procedure, participants first study a list of word pairs and then, on a subsequent recognition test, try to discriminate intact pairs from rearranged pairs. Intact pairs consist of words that appeared together on the study list, whereas rearranged pairs consist of words that appeared on the study list as part of different pairs. Item familiarity is thought to offer no help on this task because, having been seen on a recent study list, the items of both intact and rearranged pairs are, on average, equally familiar. As such, it is often assumed that accurate performance is achieved by recollecting a word's associate at study.

Because of its presumed reliance on recollection, the associative recognition task has been used to address questions about the nature of the recollection process, such as whether recollection is a *categorical* or a *continuous* process (Kelley &

Wixted, 2001; Yonelinas, 1997). The categorical view holds that a pair of items on the recognition test either occasions recollection (e.g., one item elicits recollection of its paired associate at study) or it does not. If it does, a correct intact or correct rearranged decision is made with high confidence and high accuracy. If it does not, the only recourse is to rely on some other process, such as item familiarity (thought to be of little help on this task) or to guess randomly. A continuous view of recollection, by contrast, holds that a pair of items on an associative recognition test can occasion any degree of recollection. The more recollection occasioned by the pair, the higher the confidence and the higher the accuracy (Kelley & Wixted, 2001).

Categorical recognition processes are usually conceptualized in terms of high-threshold theory, according to which memory strength either exceeds a threshold or does not exceed a threshold (Green & Swets, 1966; Macmillan & Creelman, 2005). In an associative recognition task, two recollection thresholds would apply, one which can only be exceeded by intact pairs and another which can only be exceeded by rearranged pairs (an account known as "double high-threshold" theory). By contrast, continuous recognition processes are usually conceptualized in terms of signal-detection theory, according to which the memory strengths of targets and lures are represented by normal distributions with different means and, possibly, different variances. In an associative recognition task, these two distributions might represent degrees of recollection associated with intact pairs (the "target" distribution) and rearranged pairs (the "lure" distribution).

It has long been known that a categorical process predicts a linear Receiver Operating Characteristic (ROC), whereas a continuous signal-detection-process predicts a curvilinear ROC (Green & Swets, 1966). An ROC is a plot of the hit rate vs. the false alarm rate across different levels of bias for a single memory strength condition. Multiple pairs of hit and false alarm rates representing different levels of bias are usually obtained by asking subjects to provide confidence ratings for their recognition decisions. A pair of hit and false alarm rates can be computed for each level of confidence, and those values can be plotted against each other to construct an ROC. Item recognition ROCs are almost invariably curvilinear (e.g., Egan, 1958; Ratcliff, Sheu, & Gronlund, 1992), and this has long been taken to mean that memory strengths are continuously distributed across items. Moreover, item recognition z-ROCs (i.e., a plot of the z-transformed hit rate vs. the z-transformed false alarm rate) are typically linear, which is also consistent with a Gaussian signal-detection account. By contrast, a categorical high-threshold process predicts that the z-ROC will be curvilinear (Macmillan & Creelman, 2005).

Unlike item recognition tasks, Yonelinas (1997) reported that the associative recognition task yields a nearly linear ROC (and a curvilinear z-ROC). Because this was the first linear ROC ever reported for a recognition memory task, and because it occurred for a task that was thought to depend largely on recollection, this finding offered compelling support for the idea that recollection is a categorical process. Moreover, these findings were broadly consistent with the Dual-Process Signal-Detection model (DPSD), which applies to a variety of memory tasks and which assumes that, depending on the task, performance is supported by a recollection process or a familiarity process or both (Yonelinas, 1994). Whereas recollection is assumed to be a categorical threshold process, familiarity is assumed to be governed by a continuous equal-variance signal-detection model. For a purely recollection-based task, which is what the associative recognition task is sometimes thought to be, this model reduces to double-high-threshold theory and predicts a linear ROC.

Accounting for Curvilinear Associative Recognition ROCs

Since the original report of a linear ROC, most studies have found that associative recognition procedures yield curvilinear ROCs, not linear ROCs (e.g., Healy, Light, & Chung, 2005; Qin, Raye, Johnson, & Mitchell, 2001; Verde & Rotello, 2004). This is especially true under conditions in which associative memory is strong (Kelley & Wixted, 2001). In a previous experiment that set the stage for the current research, Kelley and Wixted (2001) manipulated the strength of unrelated word pairs by presenting some pairs once and other pairs multiple times during study. The weak pairs yielded the typical pattern. That is, the ROC was curvilinear, but not as curvilinear as it should be according to a Gaussian signal-detection model, and the z-ROC was curvilinear. However, the strong pairs yielded a symmetrically curvilinear ROC and an essentially linear z-ROC that was accurately described by the equal-variance signal-detection model.

The reliably curvilinear shape of associative recognition ROCs has been taken as evidence that recollection is a continuous signal-detection process after all (Kelley & Wixted, 2001; Wixted, 2007). However, an alternative possibility, originally noted by Yonelinas (1997) and advanced more forcefully in recent years (e.g., Haskins et al., 2008), is that the curvilinearity of the associative recognition ROC instead reflects "unitized" familiarity -- that is, the familiarity of the pair considered as a unit. According to this idea, although the individual items of intact and rearranged pairs are equally familiar, intact pairs, having been presented on the study list, have a higher degree of unitized familiarity than rearranged pairs, which have not been previously seen together. If participants relied on the continuous unitized familiarity signal to discriminate intact from rearranged pairs, the DPSD model predicts that the associative recognition ROC will be curvilinear (because familiarity of any kind is construed as a signal-detection process in this model). Thus, the notion of unitized familiarity brings the DPSD model into line with curvilinear associative recognition ROCs. Using this concept, the DPSD model can even accommodate the symmetrically curvilinear ROC (and linear z-ROC) that is sometimes observed when associative memory is strong (Kelley & Wixted, 2001). For that result, the explanation would be that performance was based almost exclusively on unitized familiarity, in which case the model reduces to an equal-variance signal-detection model and predicts precisely that outcome. The first goal of the research we present here was to discriminate between these two interpretations (i.e., the continuous recollection vs. unitized familiarity interpretations) of curvilinear associative recognition ROCs and linear z-ROCs that are particularly evident when associative recognition memory is strong.

Accounting for Curvilinear Associative Recognition z-ROCs

In most associative recognition studies, overall memory strength is relatively weak (e.g., the pairs are presented once each at study), and the z-ROC is more reliably curvilinear. This result indicates that the ROC is not as curvilinear as it should be, according to signal-detection theory. Accounting for that phenomenon is a second matter of debate. The standard Gaussian signal-detection model cannot accommodate this result because it predicts that the ROC will be distinctly curvilinear and the z-ROC will be linear. Moreover, a pure double-high-threshold model, which does predict a curvilinear z-ROC, cannot account for the fact that the ROC is also typically curvilinear (not linear) when memory is weak. However, the full DPSD model can account for all of these ROC and z-ROC patterns by assuming that threshold recollection and unitized familiarity both play a role in associative recognition. That is, if unitized familiarity plays a role for some pairs on the recognition test, then it would impart some degree of curvilinearity to the ROC. If a threshold recollection process plays a role for other pairs, then it would impart some degree of curvilinearity to the z-ROC as well. Thus, the DPSD model can generally fit weak associative recognition ROCs better than a pure signal-detection model can. If the ROC is more in accordance with signal-detection theory when memory is strong, the DPSD model can accommodate that result by assuming that decisions are largely based on unitized familiarity (a signal-detection process).

An alternative interpretation holds that the z-ROC is curvilinear when memory is weak because associative information is simply not available for some of the pairs, perhaps because the information for those pairs was not encoded at study. The standard two-distribution signal-detection model has no provision for a subset of intact and rearranged test pairs without any associative information. If associative information is not encoded for some of the pairs, then the relevant signal-detection model would be a "mixture" model involving (at least) three Gaussian distributions (cf. DeCarlo, 2002), one for intact pairs, one for rearranged pairs, and one for intact and rearranged pairs for which no associative information was encoded. The third distribution is, in essence, a noise distribution. Such a model predicts a curvilinear z-ROC even though the memory strength signal is always construed as a continuous process (DeCarlo, 2003). Note that the mixture signal-detection (MSD) model does not require a fundamentally different theoretical interpretation than the standard two-distribution signal model. It merely allows for the possibility that some pairs were not encoded at study, an idea that makes no new theoretical claim about the nature of memory. The second goal of the research we conducted was to discriminate between these two interpretations (i.e., the mixture signal-detection vs. categorical recollection interpretations) of the relatively linear associative recognition ROCs and curvilinear z-ROCs that are reliably observed when memory is weak.

Figure 1 illustrates how the two competing models explain a too-linear ROC and a curvilinear z-ROC for the weak condition (the typical result) and a symmetrically curvilinear ROC and linear z-ROC for the strong condition, which is what Kelley & Wixted (2001) observed. The continuous recollection account assumes that a mixture model applies in the weak condition (Figure 1A) and that a standard signal-detection model applies in the strong condition (Figure 1B). A standard signal-detection model would apply in the strong condition if study time were such that at least some associative information was encoded for all of the pairs on the list. In the weak condition, the noise distribution is represented by a dotted curve, and it represents intact and rearranged pairs for which no associative information was encoded (denoted Intact-/Rearr-). The key theoretical consideration is that the memory strength axis in this version of the MSD model always reflects a continuous memory signal, which was assumed by Kelley & Wixted (2001) to reflect continuous recollection. This differs from the interpretation offered by the DPSD model, which is illustrated in Figures 1C and 1D.

To account for the curvilinear z-ROC seen when memory is weak, the DPSD model assumes that a probabilistic recollection process plays a role. With probability R, an intact pair or a rearranged pair elicits recollection, in which case a high-confidence correct decision is made (Figure 1C). The probability of recollection can differ for intact and rearranged pairs, but this illustration assumes a common recollection probability for the sake of simplicity. With probability 1-R, the decision is based on unitized familiarity, which is governed by an equal-variance signal detection model. The contribution of unitized familiarity accounts for the fact that weak ROCs are usually somewhat curvilinear. In the strong condition, responding is based solely on unitized familiarity (Figure 1D). That is, R would equal 0 to yield a perfectly linear z-ROC and a symmetrically curvilinear ROC (though R would be greater than zero if the z-ROC had any degree of curvilinearity in the strong condition or if the ROC were at all asymmetrical).

INSERT FIGURE 1 ABOUT HERE

In our first experiment, we investigated why the associative recognition ROC becomes more curvilinear (and the z-ROC more linear) as memory strength is increased. More specifically, the first experiment tested whether strengthening pairs on an associative recognition task results in decisions based on continuous recollection (as in Figure 1B) or unitized familiarity (as in Figure 1D). We did this by combining the strength manipulation used by Kelley and Wixted (2001) with the

Remember/Know procedure (cf. Hockley & Consoli, 1999; Verde, 2004) and with an unexpected cued recall test administered after the recognition test.

The second experiment was designed to shed light on why associative recognition ROCs have the shape they do when memory is relatively weak. More specifically, Experiment 2 was designed to assess whether the shape of the ROC typically associated with that condition (namely, a curvilinear z-ROC) reflects the inclusion of pairs for which no associative information is available (as in Figure 1A) or whether it reflects categorical recollection (as in Figure 1C). We did this by asking for old/new judgments for the pairs before asking for an intact/rearranged decision, which allowed us to construct multiple associative recognition ROCs that were conditional on confidence in the old/new decision (cf. Slotnick & Dodson, 2005). If a mixture signal-detection model applies, then it should be possible to isolate the no-information pairs and exclude them from the analysis. In particular, word pairs consisting of items that are not even recognized as having appeared on the study list are unlikely to contain any information about whether the pair is intact or rearranged. Once those pairs are excluded, the remaining pairs should yield a symmetrically curvilinear ROC and a linear z-ROC (which is the prediction made by the model shown in Figure 1A when the noise distribution is removed). The Remember/Know procedure was also employed in this experiment to help determine whether the items remaining in the analysis involved decisions based on recollection or on unitized familiarity.

Experiment 1

In Experiment 1, the strength of word pairs was manipulated within list by presenting some pairs once and other pairs 5 times. At test, participants were asked to indicate confidence in their intact/rearranged decisions and to subjectively indicate whether their decisions were based on recollection or on familiarity using the Remember/Know procedure. If the ROC data replicate the results reported by Kelley and Wixted (2001), then the weak condition will yield a relatively linear ROC (and curvilinear z-ROC) that is poorly described by signal-detection theory whereas the strong condition will yield a curvilinear ROC that is accurately described by an equal-variance signal-detection model. Note that the strong ROC is not required by any model to be symmetrically curvilinear. Instead, our investigation capitalizes on the fact that it sometimes is, as it was for the random word pairs used by Kelley and Wixted (2001) and for the associatively related word pairs used in the present experiments. Under those conditions, the competing models offer very different interpretations based on either unitized familiarity or continuous recollection.

According to one common view, Remember responses reflect recollection-based responding, whereas Know responses reflect familiarity-based responding (Bodner & Lindsay, 2003; Eldridge, Sarfatti, & Knowlton, 2002; Rajaram, 1993; Yonelinas & Jacoby, 1995). If so, then if the more curvilinear ROC in the strong condition reflects decisions based primarily on unitized familiarity (Figure 1D), then that condition should be associated with a higher proportion of Know judgments relative to the weak condition. If, instead, the more curvilinear ROC in the strong condition reflects decisions based on increased continuous recollection (Figure 1B), then that condition should be associated with a higher proportion of Remember judgments relative to the weak condition.

The logic of this analysis rests on the assumption that the usual process-pure interpretation of Remember/Know judgments is correct (i.e., that Remember responses reflect recollection and Know responses reflect familiarity). However, much recent evidence suggests that they actually reflect different degrees of memory strength (Donaldson, 1996; Dunn, 2004, 2008; Wixted & Stretch, 2004). According to this idea, when memory strength exceeds a high Remember criterion, then a Remember judgment is made, but if it only exceeds a lower Know criterion, then a Know judgment is made. Even if this interpretation is correct, it is still likely that Remember judgments indicate more recollection than Know judgments do. Wais, Mickes and Wixted (2008), for example, recently reported that source recollection associated with Remember judgments was higher than that associated with Know judgments (even though source accuracy associated with Know judgments was also above chance). The question of whether Know judgments largely reflect familiarity is more controversial, so the prediction that Know judgments will increase with unitized familiarity stems from one particular perspective, which holds that Know judgments reflect familiarity and that this holds true for unitized familiarity.

Even though the different accounts agree that Remember judgments reflect a higher degree of recollection than Know judgments, a test of the validity of that assumption was included in this experiment. Specifically, to determine whether recollection was greater for Remember judgments than for Know judgments, we also exposed participants to an unexpected cued recall test after the recognition test was completed. This test not only allowed us to determine whether Remember judgments were, in fact, associated with higher degrees of associative recollection than Know judgments but to also determine whether or not confidence in the intact-vs.-rearranged recognition decisions was in any way related to later cued recall success. If associative recognition decisions in the strong condition are based largely or exclusively on continuous recollection, one would predict that confidence in the recognition decision would be directly related in continuous fashion to later cued recall accuracy, and this should be true for both intact and rearranged pairs. If recognition decisions in the strong condition are instead based on unitized familiarity, which is thought to be and is typically modeled as being independent of recollection (e.g., Haskins et al., 2007), then no such relationship would be required.

Method

*Participants***.** Fifteen undergraduates from the University of California, San Diego participated for psychology course credit.

*Materials***.** The words were drawn from 28 different categories taken from category norms provided by Van Overschelde, Rawson, & Dunlosky (2004). Both intact and rearranged pairs were constructed using words from the same semantic categories (e.g., *ruby* was paired with *diamond*; *violin* with *cello*; *dentist* with *lawyer*, etc.). The number of pairs on the list drawn from a given category ranged from 2 to 6. Within category pairs were used to facilitate the formation of associations during study and to determine whether the effects reported by Kelley and Wixted (2001), who used unrelated word pairs, would generalize to related materials. The intact or rearranged status of the word pairs was counterbalanced across participants.

Instructions and stimuli were displayed for each participant on a NEC MultiSync LCD1760NX monitor, and powered by a Dell Dimension 4550. The presentation of stimuli and the recording of responses was controlled using E-prime software (www.pstnet.com; Psychology Software Tools).

Procedure. After consenting, participants were given instructions, presented with a list of word pairs to study, and completed the recognition test. During the study phase, 112 semantically related word pairs were randomly presented for 2 seconds either one time (56 pairs in the weak condition) or five times (56 pairs in the strong condition). The repetitions were randomly scattered throughout the list. During the subsequent recognition test, 112 pairs were shown, 56 of which were intact, and 56 of which were rearranged (28 weak pairs and 28 strong pairs, with half of each being intact and the other half being rearranged). The rearranged words were coupled with another semantically related word of equal strength (e.g., two words pairs: *flamingo* --- *penguin* and *vulture* --- *eagle* would be presented once and then at test, rearranged with one another: *flamingo* --- *eagle*). The word pairs always maintained their original left-right positions. Participants indicated on a 6-point scale whether each pair was intact or rearranged (i.e., definitely rearranged = 1, probably rearranged = 2, maybe rearranged = 3, maybe intact = 4, probably intact = 5, and definitely intact = 6). Next, participants indicated if they remembered or knew whether the word pairs were intact or rearranged. The instructions for making these judgments were adapted from Gardiner, Ramponi, and Richardson-Klavehn (1998). Participants were instructed to make a remember judgment when they consciously recollected details of the study episode, whereas a know judgment was to be made when the decisions was based on a sense of familiarity in the absence of recollection. We did not specifically draw their attention to the difference between recall-to-accept (for an intact pair) and recall-to-reject (for a rearranged pair). However, participants were given a surprise cued recall test in which they were provided with one of the words of a pair (always the left word) and asked to type in its original associate. This test allowed us to verify their understanding of a Remember judgment for both intact and rearranged pairs.

Results

Two participants performed at chance levels in the weak condition but were retained in the following analyses because, in both cases, their performance was above chance in the strong condition (excluding the participants who performed at chance in the weak condition had a negligible effect on the results). The basic hit, miss, correct rejection and false alarm rates are considered first, followed by an analysis of the ROC data and then by an analysis of the Remember/Know and cued recall data. An α level of .05 was used for all statistical tests.

Hit, Miss, Correct Rejection and False Alarm Rates. The hit rate (defined as the probability of correctly declaring an intact pair to be intact) increased with strength (Figure 2A), $t(14) = 6.42$, but the false alarm rate (the probability of incorrectly declaring a rearranged pair to be intact) was unaffected by strength (Figure 1B), replicating a pattern that has been repeatedly observed in the past (Buchler, Light & Reder, 2008; Gallo, Sullivan, Daffner, Schacter, & Budson, 2004; Jones & Jacoby, 2001; Kelley & Wixted, 2001; Malmberg & Xu, 2007; Verde & Rotello, 2004). The miss rate (equal to 1 minus the hit rate; Figure 2A) and correct rejection rate (equal to 1 minus the false alarm rates; Figure 2B) for the weak and strong conditions are redundant data but are shown to facilitate comparison with the Remember/Know results presented later.

INSERT FIGURE 2 ABOUT HERE

ROC Analyses. The group ROC data (Figure 3A) and z-ROC data (Figure 3B) were first analyzed in theory-neutral fashion by fitting 2nd-order polynomials to the data via least squares. A linear function would have a quadratic coefficient of zero, whereas a curvilinear function would have a quadratic coefficient different from zero. Because the least squares method minimizes error in the vertical direction, whereas ROC and z-ROC data are associated with error in both the vertical and horizontal directions, we performed each fit twice, reversing the axes for the second fit, and then averaged the absolute values of the resulting quadratic coefficients.

INSERT FIGURE 3 ABOUT HERE

It is visually apparent that the strong ROC data are more curvilinear than the weak ROC data (Figure 3A), whereas the reverse is true of the z-ROC data (Figure 3B). These apparent trends are reflected in the quadratic coefficients as well. For the ROC data, the quadratic coefficients for the weak and strong functions were -0.95 and -4.64, respectively (negative values indicate an inverted U-shaped function, and values closer to 0 indicate a more linear function). For the z-ROC data, the corresponding values were 0.36 and 0.06, respectively (positive values indicate a U-shaped function, and, again, values closer to 0 indicate a more linear function). These findings also replicate the results reported by Kelley and Wixted (2001).

Next, we performed model-based analyses by fitting the MSD model and the DPSD model (both of which are illustrated in Figure 1) to the ROC data by means of maximum likelihood estimation. Each model was fit simultaneously to the data from the weak and the strong conditions, and we fit them to both the group and individual-participant data. The purpose of this model-fitting exercise was to reveal how each model interprets the change in the shape of the ROC as strength increases. A question that might be asked at the outset is whether the MSD model or the DPSD model is really needed to fit associative recognition ROC data. Would the traditional unequal-variance signal-detection (UVSD) model suffice? The answer is no because its ability to describe the group ROC data was very poor, $\chi^2(11) = 69.93$. As described below, both the MSD model and the DPSD model provide much better fits.

The MSD model has four significant parameters that are scaled with respect to the mean and standard deviation of the noise distribution (arbitrarily set to 0 and 1, respectively): μ_{Rearr} is the mean of the rearranged Gaussian distribution, μ_{Intact} is the mean of the intact Gaussian distribution, σ is the standard deviation of the intact and rearranged distributions, and λ (which was constrained to vary between 0 and 1) is the probability that associative information was encoded for an intact or rearranged pair. Thus, λ represents the probability that a memory strength value for an intact pair is drawn from the intact distribution and the probability that a memory strength value for a rearranged pair is drawn from the rearranged distribution, whereas $1 - \lambda$ is the probability that a memory strength value for a test pair (whether intact or rearranged) is drawn from the noise distribution. As λ approaches 1 (as it might in the strong condition and as illustrated in Figure 1B), this model reduces to the equal-variance signal-detection model and predicts a symmetrically curvilinear ROC and linear z-ROC. Although the intact and rearranged distributions illustrated in Figure 1A have the same standard deviation as the noise distribution, there is reason to believe that they will differ in practice. Kelley and Wixted (2001) included new pairs on a recognition test, which are conceptually similar to pairs that were not encoded, and found that the standard deviation of the intact and rearranged pairs based on ROC analysis was much greater than that of the new pairs. Thus, the σ parameter was included to allow for that possibility here.

In addition to estimating the four theoretically significant parameters mentioned above, the locations of 5 confidence criteria were also estimated (which must be done no matter which model is fit to the data). For both the group and the individual fits, the confidence criteria estimates were fixed across strength conditions because the weak and strong pairs were intermixed at test. One participant was excluded from the group ROC analysis because of an extreme response bias (e.g., in the strong condition, this participant's hit and false alarm rates were 1.0 and 0.82, respectively). For the individual fits, one participant did not use one of the confidence ratings for either intact or rearranged pairs in either the weak condition or the strong condition, so adjacent confidence categories were collapsed and degrees of freedom were reduced accordingly.

Based on the group ROC analysis, the interpretation of the ROC data in terms of the MSD model is as expected in light of the polynomial regressions described earlier. Specifically, according to the MSD model, the effect of the strength manipulation was mainly to decrease the proportion of pairs drawn from the noise distribution, thereby resulting in a more curvilinear ROC and more linear z-ROC as strength increased (Table 1). For the group ROC fits, the estimates of μ_{Rearr} and μ_{Intact}

did not differ systematically as a function of strength. Indeed, constraining their values to be constant across conditions did not significantly worsen the fit, $\chi^2(2)$ = 3.19. The same was true of σ (i.e., constraining its value to be equal across strength conditions did not significantly affect the quality of the fit), $\chi^2(1) = 3.19$. By contrast, the estimate of λ differed considerably as a function of strength, and constraining its value to be constant across strength conditions did significantly worsen the fit, $\chi^2(1)$ = 29.69. Table 1 shows the estimated parameter values for the most parsimonious fit of the MSD model. In this model, only λ varies as a function of strength (with the values of the other parameters being the same in the weak and strong conditions).

The same basic story emerges for the individual fits. To prevent runaway estimates that would otherwise occur in a few cases, μ_{Rearr} was limited to a minimum -4.0 for these fits (this limit was reached for 4 participants in the weak condition and for 2 in the strong condition), and μ_{Intact} was limited to a maximum of 7.0 (this limit was reached for 2 participants in the weak condition and for 2 in the strong condition). These limits were chosen in light of the estimates obtained for these values in the group fits. Compared to a fit of the full model (with all parameters free to vary across strength), constraining all of the parameters except for λ to be the same across strength conditions did not significantly worsen the fit, $\chi^2(46) = 30.72$. By contrast, constraining λ to be the same across strength conditions while leaving all other parameters free to vary did significantly worsen the fit, $\chi^2(15) = 47.48$. Thus, as with the group fits, the most parsimonious MSD model involves two levels of λ (one for the weak condition and one for the strong condition), with all other parameters fixed across strength conditions. The average parameter estimates for this model are shown in Table 1, and it is clear that both the group and individual fits suggest that the role of the noise distribution is negligible in the strong condition (because λ is close to 1.0). Thus, according to the MSD model, the primary effect of strengthening word pairs at study was to reduce to approximately zero the proportion of pairs at test for which no associative information was available. This has the effect of creating a more curvilinear ROC (and more linear z-ROC) in the strong condition compared to the weak condition.

As a pure mathematical account, the mixture model assumes that associative memory strength is a continuously distributed variable, but it does not necessarily specify what that variable might be in terms of dual-process theory (and its validity does not depend on the validity of the distinction between recollection and familiarity). Kelley and Wixted (2001) argued that the diagnostic memory strength variable in an associative recognition task reflects recollection, and that idea was further tested in the present experiment using both the Remember/Know procedure and cued recall tests (the results of which are presented below). For the moment, the point is that the MSD model interprets a curvilinear z-ROC to reflect the contribution of a noise distribution, and it interprets the change in the shape of the z-ROC from a curvilinear shape to a more linear shape to reflect a reduced contribution a noise (in the manner depicted in Figure 1B).

INSERT TABLE 1 ABOUT HERE

The DPSD model was fit to the group and individual ROC data in the same way that the MSD model was fit. The interpretation of the ROC data in terms of the

DPSD model is also as expected in light of the polynomial regressions described earlier. For a single strength condition, the DPSD model has three significant parameters, R_{Intact}, R_{Rearr}, and d', where R_{Intact} represents the probability of above-threshold recollection for an intact pair, R_{Rearr} represents the probability of above-threshold recollection for a rearranged pair, and d' represents unitized familiarity. More specifically, d' in this model represents the distance between the means of the intact and rearranged unitized familiarity distributions (which are assumed to be Gaussian in form and to have the same variance). If d' is close to zero (as might be the case in the weak condition), then this model reduces to the double high-threshold model and predicts a linear ROC. If d' increases (as it might for the strong condition), then the model predicts a more curvilinear ROC.

INSERT TABLE 2 ABOUT HERE

Because d' in this model represents the standardized distance between the means of the intact and rearranged unitized familiarity distributions, there are 4 Gaussian distributions altogether when strength is manipulated within list (2 for the weak condition and 2 for the strong condition). The mean and standard deviation of the rearranged unitized familiarity distribution in the weak condition were set to 0 and 1, respectively, and the other familiarity distributions were scaled with respect to those values. The standard deviation of the intact familiarity distribution was also set to 1 in the weak condition (i.e., the weak condition was represented by an equal-variance familiarity model), and its estimated mean is represented by d'_{weak}. The standard deviations of the strong intact and rearranged unitized familiarity distributions were represented by the parameter σ. Thus, the strong condition was also represented by an equal-variance familiarity model, but the strong distributions were not required to have the same standard deviation as the weak distributions. Two additional parameters were needed to specify the means of the intact and rearranged distributions in the strong condition, and these were defined by the parameters δ (the mean of the strong rearranged distribution) and d'_{strong} (the mean of the strong intact distribution minus the mean of the strong rearranged distribution divided by σ). These 4 parameters (d'_{weak}, d'_{strong}, σ, and δ), plus the 4 recollection parameters ($R_{\text{Intact-strong}}$, $R_{\text{Rearranged-strong}}$, $R_{\text{Intact-weak}}$, and $R_{\text{Rearranged-weak}}$), plus the 5 confidence criteria made for a total of 13 parameters. However, in practice, only 10 were needed to fit the ROC data, which is the same number of parameters as the MSD model.

According to the fit of the DPSD model to the group ROC data, the effect of the strength manipulation was to greatly increase unitized familiarity (as measured by the model's *d'* parameter) while having no discernable effect on recollection. Indeed, compared to the fit of the full model, the quality of the fit was not significantly affected when the 4 recollection parameters (which were all limited to the range of 0 to 1) were constrained to a single recollection value, $\chi^2(3) = 4.97$. The single recollection parameter is hereafter denoted R. If R is constrained to equal 0, the fit becomes much worse, $\chi^2(1) = 26.62$, so the model does suggest that recollection occurred even though it did not differ significantly for intact or rearranged pairs or for the weak and the strong conditions. In contrast to the recollection parameters, when the familiarity parameter was constrained to be equal across conditions (i.e., $d'_{\text{weak}} = d'_{\text{strong}}$), with all other parameters free to vary, the fit was dramatically worse, $\chi^2(1) = 27.75$. Thus, 2 *d'* values are needed for an adequate fit, one for each strength condition. The value of σ (the standard deviation of the strong intact and rearranged distributions relative to the weak intact and rearranged distributions) was estimated to be 1.61. When its value was constrained to equal 1, the fit was significantly worse, $\chi^2(1) = 10.39$. Thus, according to this model, the strong intact and rearranged distributions had a higher variance than the corresponding weak distributions. The value of δ (the mean of the strong rearranged distribution) was estimated to be -0.15, which did not differ significantly from zero. Table 2 shows the parameters of the most parsimonious DPSD model (in which only the *d*' parameter differs as a function of strength). The upshot of this model-fitting exercise is that the main effect of the strength manipulation, according to the DPSD model, was to greatly increase the role played by unitized familiarity.

Similar conclusions were reached when the full model was fit to the individual ROC data. Compared to a fit of the full model, constraining the four recollection parameters to a single value (R) while allowing *d*' to vary across strength conditions did not significantly worsen the fit, $\chi^2(45) = 50.53$. By contrast, constraining *d*' be the same across strength conditions while leaving the recollection parameters free to vary (and to differ for intact and rearranged pairs) did significantly worsen the fit, $\chi^2(15) = 47.48$. Thus, as with the group fits, the most parsimonious DPSD model involves parameters that are fixed across strength conditions, except for *d*', which varies across strength. The average parameters estimates based on the individual fits for this model are shown in Table 2. As with the group fit, the individual fits suggest that the effect of strengthening word pairs at study was to increase unitized familiarity while having a negligible effect on associative recollection. This has the effect of creating a more curvilinear ROC (and more linear z-ROC) in the strong condition compared to the weak condition.

As a final check, we examined the parameter estimates obtained from fitting the full DPSD model (with all parameters free to vary) to the individual subject data. Although this model is clearly overparameterized, we used it to look for any trends that might be evident in the recollection parameters. According to the individual fits, the average estimate of R_{Intact} decreased nonsignificantly as strength increased (from .33 to .30), whereas the average estimate of R_{Rearr} increased significantly with strength (from .31 to .48), $t(14) = 2.34$. The increase in R_{Rearr} reflects the fact that the strong ROC is slightly more asymmetrical than the weak ROC. When R_{Intact} and R_{Rearr} were averaged together for each participant to estimate the overall proportion of test pairs that involved recollection, the estimate increased with strength from 0.32 to 0.39, a difference that did not approach significance. By contrast, the familiarity parameter (d') exhibited a highly significant increase, $t(14) = 4.89$. Indeed, the estimated increase in familiarity as a function of strength was dramatic, as the magnitude of d' was more than five times higher in the strong condition compared to the weak condition.

We also analyzed the group and individual ROC data with simpler versions of the MSD and DPSD models. In these models, σ was set to 1. In addition, for the DPSD model, δ was set to 0. Although the fits were worse, the conclusions were unchanged. Thus, the ROC data are interpreted by the MSD model to mean that strengthening word pairs at encoding reduced the contribution of the noise distribution and by the DPSD model to mean that strengthening word pairs at encoding substantially increased unitized familiarity while having a much lesser effect on recollection. Conceivably, an effect of strength on recollection would have been detected had more observations been obtained for each participant (thereby increasing statistical power), but it is worth noting that the number of observations was sufficient to detect a highly significant effect of strength on familiarity.

Remember/Know judgments. The MSD and DPSD models agree that in the strong condition, performance was governed in a more pronounced way by two continuous distributions -- one for intact pairs and one for rearranged pairs -- than in the weak condition. According to one view, these two distributions reflect associative recollection (Kelley & Wixted, 2001); according to the other, they reflect unitized familiarity (Haskins et al., 2008). If the strengthening of word pairs at encoding increased recollection at retrieval (by decreasing the contribution of the noise distribution, as the MSD model assumes) then, according to a standard interpretation of Remember/Know judgments, the Remember hit rate should increase as well. If the strength manipulation instead increased unitized familiarity (while having a much lesser effect on recollection, as the fit of the DPSD model suggests) then the Know hit rate should increase with strength. Similar predictions apply to the rearranged pairs.

For the intact pairs, the Remember hit rate increased significantly as strength increased, $t(14) = 6.01$, whereas the Know hit rate decreased slightly (Figure 4A). For the rearranged pairs, the Remember correct rejection rate also increased significantly as strength increased, $t(14) = 3.71$, whereas the Know correct rejection rate decreased significantly, $t(14) = 4.10$ (Figure 4B). This outcome is more consistent with the idea that recollection was higher in the strong condition than the weak condition than with the idea that unitized familiarity selectively increased with strength.

Recollection and familiarity are often estimated using the independence Remember/Know method. This method provides valid estimates if recollection is a threshold process, if Remember and Know judgments are process pure, and if recollection and familiarity are independent processes. Recollection is estimated by subtracting the Remember false alarm rate from the Remember hit rate (for the intact pairs) and by subtracting the Remember miss rate from the Remember correct rejection rate (for the rearranged pairs). Using that method, the mean estimate of recollection increased as a function of strength from .39 to .64 for the intact pairs and from .34 to .58 for the rearranged pairs. Overall, this method estimates that the proportion of decisions based on recollection for all test pairs (intact and rearranged) increased significantly from .36 to .61, $t(14) = 4.93$. Thus, the interpretation of the data in terms of the Remember/Know procedure is that performance in the strong condition was largely based on associative recollection for both intact and rearranged pairs. This result is in accordance with the continuous recollection interpretation of the MSD model and is somewhat harder to reconcile with the unitized familiarity interpretation provided by the DPSD model (particularly for the intact pairs). However, the familiarity estimates computed using the independence Remember-Know method did accord reasonably well with the unitized familiarity estimates provided by the DPSD model. Specifically, familiarity d' value was estimated to be 0.67 in the weak condition (averaged across estimates obtained from intact and rearranged pairs) and 1.40 in the strong condition (values that are not far from the values obtained by fitting the DPSD model, as shown in Table 2). Note that the estimated increase in familiarity does not mean that, according to the Independence Remember/Know method, familiarity was actually used on a greater proportion of trials in the strong condition. Instead, it means that familiarity was a more diagnostic signal on those occasions when recollection failed. As indicated above, the independence Remember/Know method – but not the DPSD model – estimates that recollection was used for 36% of the pairs in the weak condition and 61% of the pairs in the strong condition.

INSERT FIGURE 4 ABOUT HERE

The effect of the strength manipulation on Remember/Know judgments made to rearranged pairs (Figure 4B) is interesting in light of the fact that the manipulation had no measurable effect on the overall correct rejection and false alarm rates (Figure 2B). The most common interpretation of that result is that associative recollection and item familiarity were both higher in the strong condition compared to the weak condition. The increased associative recollection acts as a force to correctly declare the pair to be rearranged; the increased item familiarity acts as a force to incorrectly declare the pair to be intact (even though there is no reason why it should). These two opposing forces effectively cancel each other, thereby accounting for the fact that the strength manipulation appears to have no effect on responses to rearranged pairs. In agreement with this interpretation, Buchler et al. (2008) found that the false alarm rate to rearranged pairs increased dramatically with strength when the two items of the strong rearranged pairs had been repeated 5 times, but always as part of different pairs. This would serve to greatly increase item familiarity without increasing the ability to confidently recollect the original pairing for the items (because they were previously paired with so many different items). As such, the false alarm rate should – and did – increase.

Despite the fact that the false alarm rate remained unchanged as a function of strength in the present experiment, the Remember/Know data show that the manipulation did have an effect after all, and the increased proportion of Remember responses to correct rejections in the strong condition is consistent with the idea that associative recollection was, indeed, higher in that condition. However, item familiarity is also presumably higher in that condition, which would serve to offset the effect of increased recollection if high item familiarity is inappropriately taken by the participant to indicate that the pair is intact.

Cued Recall. For intact and rearranged pairs, cued recall for the originally paired item was higher in the strong condition (M = .76 and .66, respectively) than in the weak condition (M = .57 and .44, respectively). An analysis of variance performed on these data yielded significant effects of strength $F(1,14) = 59.6$ and pair status, $F(1, 14) = 24.1$, but the interaction was not significant. The fact that cued recall was higher for items that had earlier appeared as part of intact pairs suggests that recall performance for rearranged pairs may have suffered interference because participants saw the incorrect associate during the recognition test (indeed, the incorrect associate was mistakenly recalled for 13% of the rearranged pair cues) and/or that recall performance for intact pairs was facilitated because participants saw the correct associate during the recognition test. However, these effects did not interact with strength.

In both the weak condition and in the strong condition, cued recall was higher for correct Remember judgments than for correct Know judgments, and this was true for both intact pairs (R-Hit vs. K-Hit; Figure 5A) and rearranged pairs (R-CR vs. K-CR; Figure 5B). An analysis of variance performed on the data for the intact pairs yielded a significant effect of type of judgment (Remember vs. Know), $F(1, 11) = 17.9$, but no effect of strength. Thus, although Remember hits were more common in the strong condition (and Know hits less common) compared to the weak condition, the cued recall advantage for Remember hits was similar in both conditions. An analysis of variance performed on the data for the rearranged pairs yielded a significant effect of type of judgment (Remember vs. Know), $F(1, 13) = 5.1$, and a significant effect of strength, $F(1, 13) = 6.4$, but no hint of an interaction. That is, Remember judgments were associated with higher cued recall accuracy than Know judgments for rearranged pairs, but cued recall accuracy for both Remember and Know judgments was higher in the strong condition compared to the weak condition. The effect of strength in this case could mean that a rearranged pair had to achieve a higher degree of recollection to be given a Remember or a Know judgment in the strong condition than in the weak condition because, in the strong condition, recollection was working against the opposing force of increased item familiarity.

INSERT FIGURE 5 ABOUT HERE

The fact that cued recall accuracy was substantially higher for Remember judgments than for Know judgments validates the notion that Remember judgments made during the recognition test reflect greater recollection than Know judgments, and it is consistent with the idea that recollection was higher in the strong condition relative to the weak condition (because Remember judgments increased as strength increased). This validity test is important because a common concern is that inadequate instructions could lead participants to say "remember" when familiarity is strong instead of when the recognition decision is based on recollection. These findings are consistent with the idea that a Remember judgment was made when recollection was strong.

Cued recall performance was generally low for Remember and Know misses (Figure 5A), though it was relatively high for Remember and Know false alarms (Figure 5B). This asymmetry may again reflect the fact that high item familiarity is taken by the participant as evidence that a rearranged pair is intact. Thus, even though recollection for some of those pairs is reasonably high (encouraging a rearranged decision), strong item familiarity for those same pairs may have induced the participant to respond "intact" anyway.

A question of particular interest concerns the relationship between confidence in the intact vs. rearranged recognition decision and the probability of later successful cued recall. If associative recognition decisions are based on continuous recollection, one would expect that relationship to be strong. Figure 6A shows the relationship for the intact pairs, and Figure 6B shows the relationship for the rearranged pairs. The data were collapsed across adjacent confidence ratings to reduce variability.

For the statistical analysis of these results, the data were also collapsed across the weak and strong conditions because a more fine-grained analysis would have too many missing values. For the intact pairs (Figure 6A), higher confidence that the pair was intact was associated with greater cued recall performance, $F(2, 26) = 32.83$. For the rearranged pairs (Figure 6B), the highest levels of confidence that the pair was rearranged were also associated with the highest cued recall accuracy, but the relationship was somewhat flatter thereafter. Still, the effect of confidence was significant, *F*(2, 28) = 4.14, reflecting the fact that recall accuracy was highest for the most confident rearranged decisions. Once again, the flattening of the confidence-accuracy function (relative to the relationship obtained for intact pairs) may reflect the fact that high item familiarity is inappropriately interpreted by the participant as evidence that the pair is intact. If so, then recollection could be fairly strong for some rearranged pairs (which, alone, might support a confidence rating of 1, 2 or 3) but the effect is offset by high item familiarity (leading to a rating of 4, 5 or 6). Whether or not that explanation is correct, the main point is that confidence in an intact-vs.-rearranged recognition decision for a test pair is clearly related the degree of later recollection for that pair.

INSERT FIGURE 6 ABOUT HERE

Overall, the Remember/Know and cued recall data suggest that recollection increased substantially with strength for both intact and rearranged pairs and that the increased recollection was associated with a more curvilinear ROC. In addition, particularly for intact pairs, confidence in the intact/rearranged recognition decision was strongly related to the accuracy of later cued recall (in continuous fashion). All of these findings are consistent with the idea that recollection is a continuous process that yields a curvilinear ROC. This interpretation is easily reconciled with the MSD model depicted in Figure 1A and 1B. By contrast, the findings are less consistent with the DPSD model, according to which the strengthening of pairs on the study list dramatically increased unitized familiarity while also modestly increasing recollection for rearranged pairs only. Evidence from the Remember/Know procedure offered no suggestion that the effect of the strength manipulation was mainly attributable to increased unitized familiarity for the intact pairs, as the DPSD model suggests. Instead, the evidence suggests that recollection increased substantially for intact pairs as well as rearranged pairs. In addition, even for the strong condition, the relationship between confidence in the recognition decision for intact pairs and later cued recall was very strong (as if recollection had a pronounced role in the recognition decisions for those pairs).

It is, of course, possible that unitized familiarity increased with strength as well (as the independence Remember/Know estimates imply), but there is nothing to suggest that this was the dominant force in the strong condition. Indeed, the Remember/Know data suggest that more than 60% of the decisions in that condition were based on recollection, and the cued recall data are consistent with that estimate as well (with average cued recall accuracy in the strong condition exceeding 70% correct). The recollection estimates provided by the DPSD model are much lower than that because the ROC in the strong condition was quite curvilinear and was essentially symmetrical. Had a more pronounced asymmetry been observed in the strong condition, then the recollection estimates provided by the DPSD model would have been higher as well (and the MSD model would need to be modified to accommodate that asymmetry). However, for these materials, a relatively symmetrical ROC was obtained, as Kelley and Wixted (2001) found for the unrelated word pairs they used, and this allowed for a clearer test of the two models. Although the results of Experiment 1 were more sensibly interpreted by the MSD model than by the DPSD model, it could be argued that the failure of the DPSD model to detect an increase in recollection as a function of strength was a Type II error (even though the data were sufficient to detect a highly significant increase in familiarity as a function of strength). Thus, Experiment 2 approached the same issue from a different angle.

Experiment 2

Experiment 1 asked whether the curvilinear ROC (and linear z-ROC) that is reliably observed when associative recognition is strong reflects a continuous recollection process or a unitized familiarity process. The Remember/Know and cued recall results both supported the continuous recollection interpretation and were harder to reconcile with the unitized familiarity interpretation. Experiment 2 asks why, when associative recognition is weak, the ROC tends to be nearly linear (linear enough, at least, to reliably yield a curvilinear z-ROC).

The DPSD model assumes that the curvilinear z-ROC reflects a threshold recollection process (Figure 1C), whereas the mixture model assumes that it reflects the contribution of noise (Figure 1A). An exactly analogous debate has occurred with respect to source memory ROCs. In a source recognition procedure, items on a list are presented with different to-be-recollected details (e.g., some items are presented on the left side of the screen and others on the right). On a later recognition test, items are first presented in source-neutral fashion (e.g., at the center of the screen) for an old/new decision followed by a source recollection judgment (e.g., was the item presented on the bottom or on the top). When confidence ratings are taken for both decisions, the item and source ROCs can be plotted separately, and the source ROC is often conspicuously more linear than the item ROC (Yonelinas, 1999). Because source memory decisions are often thought to be based on recollection, the fact that the source memory ROC is sometimes linear (and the z-ROC is almost always curvilinear) has been taken to support the idea that recollection is a threshold process instead of a continuous process.

An alternative explanation is that the ROC anomalies observed on source memory tests instead reflect a mixture model caused by the absence of source information for a subset of the test items. DeCarlo (2003) demonstrated the viability of the mixture signal-detection account of source memory as an alternative to the notion of threshold recollection by showing, for example, that it predicts the curvilinear z-ROCs that are often interpreted as reflecting recollection. A direct and compelling demonstration of the validity of the mixture model was provided by Slotnick and Dodson (2005), who reanalyzed the linear source ROC data reported by Yonelinas (1999) by partitioning the data according to the confidence ratings made during the old/new phase. That is, instead of plotting one source ROC using all of the data, as Yonelinas (1999) had done, they plotted separate source ROCs according to the confidence in the old/new decisions. Some target items, for example, were not recognized as having appeared on the list (i.e., they were mistakenly declared to be new). The source ROC for these items fell along the chance diagonal, which provides direct evidence for the existence of a noise distribution (i.e., a subset of target items for which no differential source information was available).

Of particular interest was the source ROC constructed solely from items that had received a high-confidence old/new decision. According to the DPSD model, any recollection that occurs during the old/new phase would yield an old/new decision made with high-confidence. Thus, excluding responses made with lower old/new confidence from the source ROC analysis would theoretically exclude only non-recollection-based decisions, so the natural expectation is that the shape of the ROC would remain linear. Instead, the source ROC for that condition (i.e., the condition in which recollection was theoretically preserved) was quite curvilinear and was accurately described by the signal-detection model. Wixted (2007) argued that this is direct evidence for a continuous recollection signal, though Parks and Yonelinas (2007) argued that it instead reflects unitized source familiarity.

In an associative recognition task, a similar analysis is generally not possible because the intact/rearranged decision is not typically preceded by an old/new decision. In Experiment 2, we modified the standard associative recognition procedure by including an old/new question for each pair prior to the intact-vs.-rearranged question. More specifically, for each pair of items, participants were asked to indicate using a 6-point confidence scale whether the items of the pair were new (i.e., neither word appeared on the study list) or old (i.e., both words appeared on the study list), regardless of whether they are now intact or rearranged. After making the old/new decision, they were then asked whether the pair was intact or rearranged (again, using a 6-point confidence scale). One question of interest was whether we would find direct evidence of a noise distribution by plotting the associative recognition ROC for pairs constructed using intact and rearranged pairs consisting of items that were not recognized as having been presented on the list. An ROC constructed from intact and rearranged pairs drawn from the noise distribution should fall along the diagonal (indicating chance performance for these noise pairs). Such a result would provide direct evidence in support of a basic assumption of the MSD model. Another question of interest was whether excluding these items (which would retain any recollection-based decisions) would yield a linear z-ROC or a curvilinear z-ROC. A continuous recollection MSD model predicts a linear z-ROC under these conditions (because the noise items were removed). By contrast, the DPSD model predicts a curvilinear z-ROC (because recollection-based responses would be retained in this analysis). The former outcome would parallel what has been observed on source recognition tasks (Slotnick & Dodson, 2005).

As indicated above, Parks and Yonelinas (2007) argued that the curvilinear source ROC that Slotnick and Dodson (2005) found when the source ROC was constructed using high-confidence old/new decisions (decisions generally though to be based largely on recollection) was indicative of strong, unitized source familiarity, a concept that is similar to the notion of unitized familiarity that has been advanced to explain curvilinear associative recognition ROCs. According to this idea, an item and its source can be unitized in such a way that successful source memory can be based on familiarity (leading to a curvilinear ROC), not threshold recollection, as had been generally assumed in the past. The same idea could be brought to bear on the associative recognition procedure should the results mirror those obtained using a source memory procedure. To investigate that possibility, Experiment 2 also made use of the Remember/Know procedure. If the refined associative recognition ROC constructed using pairs that received high-confidence old/new decisions is curvilinear, then these associative recognition decisions should be largely accompanied by Know judgments if that curvilinearity reflects unitized familiarity. If, instead, this refined ROC reflects continuous recollection, then these associative recognition decisions should be largely accompanied by Remember judgments.

Method

Participants. Thirty undergraduates from the University of California, San Diego participated for psychology course credit.

Materials and Design. The pairs were the same semantically related words as those used for Experiment 1.

Procedure. The procedures were similar to the weak condition of Experiment 1 in that the pairs were semantically related, presented once each for 2 seconds, counterbalanced for intact and rearranged pairs, and remained in their original positions (first or second word of the pair) on the old/new and associative recognition tests. For the old/new test, we introduced 24 new word pairs randomly intermixed with 88 word pairs (44 intact and 44 rearranged) from the study list, for a total of 112 pairs. Participants first indicated whether the word pairs were new (by entering a rating from 1 to 3, indicating increasing uncertainty that the words were not on the list) or old (by entering a rating from 4 to 6, indicating increasing certainty that the words had been on the list). Participants then indicated on a 6-point scale whether each pair was intact or rearranged (i.e., definitely rearranged = 1, probably rearranged = 2, maybe rearranged = 3, maybe intact = 4, probably intact = 5, and definitely intact

= 6). Finally, participants indicated if they remembered or knew whether the word pairs were intact or rearranged.

Results

Six participants performed no better than chance on the associative recognition tests, and they were excluded from further consideration. As such, the analyses presented below are based on 24 participants. Figure 7a shows the group ROC for the associative recognition test using all of the data, irrespective of the old/new confidence ratings (i.e., the data are "collapsed" over the old/new ratings). Because old/new ratings are not typically obtained in associative recognition tasks, this ROC is a typical associative recognition ROC, and it is comparable to the weak ROC in Experiment 1 (because, in both cases, the pairs were presented only once at study). Figure 7b shows the corresponding z-ROC. Once again, 2nd-order polynomials were fit to these data, and both the ROC and z-ROC exhibit curvilinearity. The quadratic coefficient for the ROC was -1.27, whereas the corresponding value for the z-ROC was 0.22 (a typical value).

INSERT FIGURE 7 ABOUT HERE

The main purpose of Experiment 2 was to experimentally identify pairs for which no associative information was encoded and to exclude them from the analysis to see what the shape of the ROC would be for the remaining pairs (for which associative information was encoded). Before excluding any pairs, we first fit the MSD model and the DPSD model to the collapsed ROC data. The full MSD model involves 9 parameters (the 5 confidence criteria, μ_{Intact}, μ_{Rearr}, λ, and σ). The group ROC fit was very good, $\chi^2(1) = 0.89$. In addition, the estimated value of σ was 3.65

(similar to Experiment 1), but the quality of the group fit was not significantly impaired by setting σ equal to 1, $\chi^2(1) = 2.00$, so we used that constraint in all further fits. This had the advantage of equating the number of free parameters between the MSD model and the DPSD model (which no longer has a σ parameter or a δ parameter because there is no within-list strength manipulation). That is, for these fits, the MSD model had 3 significant parameters (μ_{Intact}, μ_{Rearr}, and λ), and so did the DPSD model (R_{Intact}, R_{Rearr} and d').

Table 3 shows the maximum likelihood fit of the MSD model to the group ROC data shown in Figure 7. Also shown are the results obtained from fitting individual ROC data from 21 of the individual participants (3 participants used too few ratings to fit either model). Clearly, for the MSD model, the curvilinearity of the collapsed z-ROC is attributable to the mixture parameter (λ) taking on a value greater than zero, and its value significantly exceeds 0 in both the group fit, $\chi^2(1) = 16.16$, and the individual fits, $\chi^2(42) = 58.44$.

INSERT TABLE 3 ABOUT HERE

Table 4 shows the maximum likelihood fit of the DPSD model to the group ROC data shown in Figure 7. Also shown are the results obtained from fitting individual ROC data from 21 of the individual participants. For the DPSD model, the curvilinearity of the collapsed z-ROC is attributable to threshold recollection (R_{Rearr} and R_{Intact}) taking on values greater than zero. Unlike in Experiment 1, R_{Intact} was significantly greater than R_{Rearr} for both the group fit, $\chi^2(1) = 19.05$, and the individual fits, $\chi^2(21) = 33.11$. In addition, constraining R_{Rearr} and R_{Intact} to zero significantly worsened the group fit, $\chi^2(2) = 19.05$, and the effect was marginally significant for the individual fits as well, $\chi^2(42) = 57.73$, $p = .054$. Thus, for collapsed data, the curvilinearity of the z-ROC is interpreted by this model to reflect the contribution of threshold recollection.

INSERT TABLE 4 ABOUT HERE

To further investigate the nature of the curvilinear z-ROC, we constructed three separate associative recognition ROCs (and corresponding z-ROCs) that were conditional on the old/new confidence ratings. The first conditional ROC was based on associative recognition decisions that were made to intact and rearranged pairs that had received ratings of 1, 2 or 3 in response to the old/new question. That is, this ROC and z-ROC were based intact and rearranged pairs consisting of items that were not recognized as having appeared on the study list (20% of the intact pairs and 32% of the rearranged pairs fell into this category). The second conditional ROC was based on associative recognition decisions that were made to intact and rearranged pairs that had received ratings of 4 or 5 on the old/new portion of the test. That is, this ROC and its corresponding z-ROC were based on intact and rearranged pairs consisting of items that were recognized as having appeared on the study list with low or medium confidence (7% of the intact pairs and 10% of the rearranged pairs fell into that category). The third conditional ROC was based on associative recognition decisions that were made to intact and rearranged pairs that had received ratings of 6 on the old/new portion of the test (73% of the intact pairs and 58% of the rearranged pairs fell into that category).

The conditional ROCs and z-ROCs are shown in Figure 8. The curves drawn through the ROC data reflect the best-fitting equal-variance signal detection (EVSD) model. This model was used to characterize the data because the DPSD model and the MSD model both reduce to the equal-variance model when they are fit to these data. That is, the recollection parameters of the DPSD model do not differ significantly from zero for any of the three fits and neither does the mixture parameter of the MSD model. The maximum likelihood estimates of d' based on the fit of the EVSD model to the conditional ROCs are shown in Table 5. These estimates show that the associative recognition d' increases dramatically as old/new confidence increases. The curves drawn through the z-ROC in Figure 8 represent the best-fitting 2nd-order polynomial, and the quadratic coefficients for these fits are also shown in Table 5.

INSERT FIGURE 8 ABOUT HERE

For the first conditional ROC (Figure 8A), the points essentially fall along the diagonal line, which indicates that no associative information -- recollection based or familiarity based -- was available for these pairs. The corresponding z-ROC (Figure 8b) was quite curvilinear, with a quadratic coefficient equal to 0.33 (Table 5), but this is probably due to random error. Had the ROC points fallen exactly along the diagonal, for example, the z-ROC would have been linear. When the EVSD model was fit to these associative recognition ROC data (i.e., to the O/N 1-3 conditional ROC), the estimated d' was, not surprisingly, indicative of essentially chance performance (Table 5).

INSERT TABLE 4 ABOUT HERE

Direct evidence in favor of the mixture signal-detection account is provided by the fact that a substantial number of test pairs contain no associative information.

The existence of these items alone – and their existence does not appear to be in doubt -- could cause the collapsed z-ROC to exhibit the curvilinearity that is often interpreted to reflect a threshold recollection process, and it could explain why the DPSD model often fits collapsed ROC better than the standard signal-detection model. Moreover, only the MSD model clearly predicts that the selective removal of these items from the ROC analysis will yield a symmetrically curvilinear ROC. In accordance with this prediction, both the O/N 4-5 and O/N 6 conditional ROCs are well fit by the EVSD model, and the quadratic coefficients for the corresponding z-ROCs are close to 0 (Table 4). Thus, these results are consistent with the model depicted in Figures 1A and 1B. If associative recognition decisions are based largely on recollection, as is typically assumed, these results are also consistent with the idea that recollection is a continuous process.

The fact that associative recognition performance for the O/N 4-5 pairs is intermediate between the O/N 1-3 pairs and the O/N 6 pairs suggests that associative information may not be actually encoded in the categorical fashion that the MSD model envisions (i.e., there may not be a sharp distinction between encoded pairs vs. not encoded pairs). Instead, some pairs receive no attention at encoding, some receive partial attention, and some receive full attention. However, the present data can be effectively characterized by the MSD model because the large majority of the pairs fell into the two extreme conditions. More specifically, 26% fell into the O/N 1-3 condition and 66% fell into the O/N 6 condition, with only 8% falling into the intermediate O/N 4-5 condition.

The DPSD model provides a different interpretation of these conditional ROC results, and it is an interpretation that differs from the one that it offers for the collapsed ROC (according to which threshold recollection played an important role). The existence of the "no information" test pairs represented in Figure 8A is not predicted by the DPSD model but could be accommodated by assuming that (a) the information for these items falls below a recollection threshold and (b) these same items were not associated with high enough item familiarity to be recognized as old (on the initial old/new test) and were also not associated differential amounts of unitized familiarity (which might have yielded above-chance performance on the associative recognition test). In the DPSD model, recollection and familiarity are usually assumed to be independent processes, so this would be an unexpected outcome. However, if the independence assumption is dropped, the subset of pairs with no associative information can be assumed to involve below-threshold recollection, low item familiarity, and no unitized associative familiarity.

It is somewhat more challenging for the DPSD model to explain the conditional ROC derived from old/new confidence ratings of 6 (Figure 8E). This old/new confidence category has special theoretical significance because the DPSD model assumes that any recollection-based decisions that occurred during the old/new phase received a high-confidence rating (i.e., a rating of 6). To the extent that recollection occurs, according to this model, the slope of the old/new z-ROC should be less than 1 -- more so the more recollection that occurs (Glanzer et al., 1999). For these data, the old/new z-ROC has a slope of 0.59, indicating a considerable degree of recollection according to the DPSD model. Given that intact and rearranged pairs were presented for old/new decisions, it is reasonable to suppose that much of this recollection involved associative recollection (the kind that would serve to distinguish between intact and rearranged pairs on the subsequent question).

If associative recollection is a high-threshold process, and if it occurred for many of the pairs that received a high-confidence old/new decision, then the conditional associative recognition ROC constructed from these pairs should be much too linear to be accurately described by the signal detection model (and the curvilinearity of the corresponding z-ROC should be pronounced). However, the O/N 6 associative recognition ROC is clearly curvilinear, and the corresponding z-ROC was again nearly linear, with a quadratic coefficient equal to only 0.03 (Table 4). The only explanation that can be offered by the DPSD model is that these pairs mainly involved unitized familiarity, an idea that curiously leaves little room for recollection in any of the intact-vs.-rearranged decisions, this in spite of the estimates obtained from fitting the collapsed ROC (which were significantly greater than zero). More specifically, the recollection parameters of the DPSD model (R_{Rearr} and R_{Intact}) represent estimates of the proportion of the rearranged pairs and proportion of the intact pairs included in the ROC analysis that involved threshold recollection. For the collapsed ROC, R_{Rearr} and R_{Intact} were estimated to be .08 and .38, respectively, for the group ROC fit and .25 and .33, respectively, for the individual ROC fits (Table 4). When pairs that theoretically do not involve recollection were dropped from the ROC analysis, as was the case for the O/N 6 conditional analysis, these estimated proportions should increase accordingly (again, because, theoretically, only non-recollection based pairs were removed). Indeed, because 35% of non-recollection pairs were dropped from this ROC analysis, the recollection parameters would be expected to increase by approximately 50%. Instead, for that ROC analysis, the estimated values of R_{Rearr} and R_{Intact} did not differ significantly from zero, and the EVSD model adequately characterized the data.

We also fit the EVSD model to the individual O/N 6 associative recognition ROC data from the 16 participants who yielded enough observations to perform such an analysis. As with the group data, the individual ROC data did not differ significantly from the predictions based on the EVSD model, $\chi^2(57) = 69.69$. Thus, the need for recollection parameters disappeared when, theoretically, items not involving recollection were removed from the analysis (an adjustment that should have dramatically increased the recollection estimates if the DPSD model is correct).

The failure to detect the need for recollection parameters in this analysis could reflect a lack of power because the conditional analysis was based on fewer observations than the full analysis. However, offsetting that concern is the fact that the DPSD model predicts that this conditional analysis (which was restricted to pairs that received a high-confidence old decision) should dramatically increase the estimated proportion of recollected pairs (because, theoretically, only pairs not involving recollection were excluded from the analysis) and should cause the z-ROC to be more curvilinear. No hint of this effect was observed.

Because the EVSD model adequately characterizes these conditional ROC data, the DPSD model interprets the above-chance ROC data in Figure 8C to reflect the virtual absence of threshold recollection (an interpretation also implied by the linear z-ROC obtained in that condition). The only way for the DPSD model to explain these results is to assume that associative recognition performance for the high-confidence Old/New pairs was based on unitized familiarity (cf. Parks & Yonelinas, 2007). If so, this suggests that the above-zero recollection estimates from the fit of the DPSD model to the collapsed ROC (Table 3) were not actually measuring recollection because recollection is difficult to detect when the conditional ROCs are separately examined. Thus, in light of the conditional ROC analyses, it seems likely that the recollection parameters estimated from the collapsed ROC analysis were capturing the contribution of the noise distribution envisioned by the MSD model.

Remember/Know judgments. One account views the increasingly curvilinear conditional ROCs and the increasing d' values associated with increasing old/new confidence (Table 4) to reflect increasing associative recollection (namely, the continuous recollection interpretation of the MSD model). The other views that same phenomenon as to reflect increasing unitized familiarity (namely, the DPSD model). If the former interpretation is correct, one would expect Remember judgments for associative recognition decisions to increase with old/new confidence. If the latter interpretation is correct, one would instead expect Know judgments for associative recognition decisions to increase with old/new confidence. Figure 9 shows the Remember/Know data collapsed across intact and rearranged pairs (R+ indicates correct Remember judgments to intact and rearranged pairs; R- indicates incorrect Remember judgments to intact and rearranged pairs), and it is clear that correct Remember judgments increased dramatically for the strongest conditional ROC, whereas Know judgments decreased dramatically (K+ indicates correct Know judgments to intact and rearranged pairs; K- indicates incorrect Know judgments to intact and rearranged pairs).

We again used the independence Remember/Know method to compute estimates of recollection and familiarity (estimates that are valid if recollection is a threshold process and recollection and familiarity are independent processes). The estimates were computed separately for intact and rearranged pairs and then averaged together for the three conditional ROC conditions. The estimated recollection estimates for the O/N 1-4, O/N 4-5, and O/N 6 conditions were .01, .08 and .40, respectively. The corresponding *d'* familiarity estimates were -0.13, 0.85 and 0.99. Thus, as with Experiment 1, the estimates suggest that recollection as well as familiarity play a large role in the stronger condition (which, in this case, was the O/N 6 condition). Even so, the ROC was well characterized by the curvilinear EVSD model. This result supports the continuous recollection interpretation of the curvilinear associative recognition ROC.

INSERT FIGURE 9 ABOUT HERE

General Discussion

We addressed two main issues in this research. The first issue is whether the curvilinearity of an associative recognition ROC, which is more likely to be observed when memory is strong, reflects a continuous recollection process or a continuous unitized familiarity process. The second issue is whether the curvilinearity of an associative recognition z-ROC, which is more likely to be observed when memory is weak, reflects a categorical recollection process or reflects the influence of noise caused by the inclusion of intact and rearranged test pairs for which no associative information is available. The answers to these questions bear on the nature of the dual-process theory of recognition memory as well as on the search for the neuroanatomical correlates of recollection and familiarity (a search that is guided by dual-process theory and that makes considerable use of associative recognition procedures).

With regard to the first question, the results suggest that a curvilinear associative recognition ROC reflects a continuous recollection process, not a unitized familiarity process. More specifically, in Experiment 1, increasing the strength of pairs by presenting them multiple times on a study list greatly increased the curvilinearity of the ROC while also increasing the proportion of correct Remember judgments and increasing cued recall accuracy on a subsequent (unexpected) test. In addition, the degree of confidence in intact vs. rearranged recognition decisions was directly related to the accuracy of subsequent cued recall (as if higher degrees of recollection were associated with higher levels of recognition confidence). All of this occurred even though the DPSD model, when fit to the weak and strong ROC data, suggested that the effect of the strength manipulation was to dramatically increase unitized familiarity while having a lesser effect on categorical recollection. The Remember/Know and cued recall results instead suggest that recollection increased considerably with strength (as intuition would suggest) and that the increased recollection was accompanied by an increasingly curvilinear ROC. The fact that the ROC becomes more curvilinear when recollection increases is consistent with the idea that recollection is a continuous process.

With regard to the second issue, the results suggest that the curvilinear z-ROC that is often observed on associative recognition tasks when memory is relatively weak does not result from a categorical recollection process but instead results from the inclusion of intact and rearranged test pairs for which no associative information is available. That is, the results suggest that a mixture signal-detection applies (Figure 1A). In support of that idea, 25% of the intact and rearranged test pairs consisted of items that were not even recognized as having appeared on the list when they were presented on an old/new recognition test. For the pairs made up of these unrecognized items, intact vs. rearranged recognition performance was at chance. This result provides direct evidence for the existence of a noise distribution that is intermixed with the intact distribution and rearranged distribution. When these no-information items are omitted from the analysis, the ROC becomes curvilinear and the z-ROC becomes linear, leaving virtually no sign of what has in the past been taken as evidence of a categorical recollection process (namely, the curvilinear z-ROC that is reliably observed when collapsed associative recognition ROCs are analyzed). The curvilinear ROC that remains when these noise items are removed from the analysis is associated with an increased proportion of Remember judgments, and this is particularly true of the conditional associative recognition ROC constructed from items that were, on an initial old/new test, recognized as having appeared on the list with high confidence (items that seem likely to be associated with a high degree of recollection). As with the results of Experiment 1 in which strength was manipulated across conditions, this again suggests that strong associative recognition ROCs, which are accurately characterized by signal-detection theory, are largely based on recollection. That, in turn, is consistent with the idea that recollection is a continuous process.

It is important to clarify the difference between a noise distribution as envisioned by the mixture model and the notion of below-threshold recollection, which is how a threshold recollection model would explain the 25% of test pairs in Experiment 2 that were devoid of associative information. The noise items in a mixture model reflect the absence of *differential* recollection for intact and rearranged pairs, but they do not reflect the absence of a recollection signal. According to this model, participants query memory for an associative signal in the same way for every pair, and every pair returns a recollection signal that falls in a range that extends from a strong sense that the pair is rearranged to a strong sense that the pair is intact. This is true even for the subset of pairs drawn from the noise distribution. In a threshold recollection model, by contrast, an associative recollection signal is either present or it is absent. For the pairs that are devoid of associative information, the assumption would be that the recollection signal is absent, in which case responding would be based on some other process (item familiarity, unitized familiarity, or guessing). One problem with this way of thinking is that when these noise items were removed from the ROC analysis in Experiment 2, there was little, if any, sign of above-threshold recollection despite the fact that the pairs remaining in the analysis were associated with considerable associative information. Thus, although the threshold recollection account can accommodate the absence of recollection for some items, the apparent disappearance of recollection-based decisions when the noise items are removed from the analysis presents a dilemma for this account.

These conclusions are the same as those reached by Slotnick and Dodson (2005) in their conditional ROC analysis of source memory. Our Experiment 2 was, in fact, modeled on the conditional ROC analysis that they performed for a source memory procedure. Source memory ROCs typically exhibit the same trends that associative recognition ROCs exhibit (namely, curvilinear ROCs and curvilinear z-ROCs), and these ROC trends have been taken to reflect a categorical recollection process (Yonelinas, 1999). However, Slotnick and Dodson (2005) found that the cuvilinearity of the z-ROC was an artifact caused by the inclusion of weak items for which no differential source information was available. When those noise items were excluded from the analysis (e.g., when the source ROC was constructed using only items that had received high-confidence old/new decisions), the source ROC was curvilinear and the z-ROC was linear. This again suggests that recollection is a continuous process and that prior ROC evidence suggesting otherwise instead reflects a mixture signal-detection model (not threshold recollection).

Implications for Dual-Process Theory of Recognition

Prior dual-process theories of recognition memory have all implicitly or explicitly adopted the categorical view of recollection, as noted by Mickes, Wais, and Wixted (2009). If recollection is a categorical process, then it makes sense to assume – as all prior dual-process theories have – that individual recognition memory decisions are based either on recollection (if it occurs) or on familiarity (if recollection does not occur). This common version of dual-process theory is hard to reconcile with the longstanding signal detection view, which assumes that recognition memory decisions are based on a unidimensional and continuously distributed memory strength variable. If recollection is a continuous process, however, then the idea that recollection and familiarity are aggregated into a unidimensional memory strength variable becomes viable. Otherwise, one would be faced with the unusual choice of choosing one process or the other when, for example, recollection for a particular item was relatively weak and familiarity for that item was also somewhat weak. Unless these two memory signals were completely redundant, combining them into a single signal would often be more efficient than relying on either one alone. A dual process theory that assumes an aggregated and continuously distributed memory strength variable is naturally reconciled with signal detection theory (Wixted, 2007).

Starns and Ratcliff (2008) recently conducted a study of both item and associative recognition in which they compared the ability of several models to characterize associative recognition performance, including the DPSD model. The old/new item recognition decisions were accompanied by Remember/Know judgments, and the associative recognition decisions were made using a 6-point confidence scale. They found that the signal-detection model fit their item and associative recognition data much better than the DPSD model. Moreover, associative recognition performance was found to be a strong predictor of item recognition performance, which they interpreted as clear evidence that item recognition decisions are based on a memory signal that involves a combination of the available memory processes. The idea that recognition decisions are based on an aggregated memory signal is a relatively new conceptualization of the dual-process theory of recognition memory, and it is the only conceptualization that renders dual-process theory compatible with a unidimensional signal detection model.

Implications for Cognitive Neuroscience

A continuous view of recollection would have far reaching implications for the interpretation of neuroimaging studies and lesion studies that have attempted to identify the neuroanatomical basis of recollection and familiarity. In recent years, a great deal of evidence has accumulated in support of the idea that recollection is subserved by the hippocampus, whereas familiarity is subserved by the adjacent perirhinal cortex (e.g., Eichenbaum, Yonelinas & Ranganath, 2007). However, the interpretation of virtually all of this evidence hinges on the validity of the categorical view of recollection. If recollection is instead a continuous process, then the same evidence would not imply that adjacent structures in the medial temporal lobe selectively subserve recollection and familiarity (Squire, Wixted and Clark, 2007).

A specific strategy used in some of these studies involved measuring the linearity of the ROC (and curvilinearity of the z-ROC) to quantify recollection and familiarity. The more linear the ROC (and the more curvilinear the z-ROC), the more it was assumed that responding was based on categorical recollection. For example, Haskins, et al. (2008) arranged an experimental task that was thought to encourage unitized familiarity for some pairs and recollection for others. More specifically, in the unitized familiarity condition, participants were asked how well the two words, considered as a compound word, fit a presented definition (e.g., how well does SLOPE-BREAD fit the definition "A pastry eaten by mountain-climbers"). In the associative recollection condition, they were asked how well the two words of a pair served to complete a sentence that contained two blanks (e.g., how well do STEAM and TOKEN complete the sentence "The ______ for the bath cost one _____"). The evidence that the manipulation achieved its intended effect was based on the fact that the shape of the associative recognition ROC was more curvilinear in the compound condition compared to the sentence condition. In accordance with the DPSD model, this result was interpreted to mean that compound word recognition was based on unitized familiarity. Moreover, activity in perirhinal cortex was selectively elevated at encoding for words learned as compounds. This result was taken to mean that perirhinal cortex selectively subserves the (unitized) familiarity process, but that conclusion depends on the assumption that a curvilinear associative recognition ROC reflects unitized familiarity. The present results indicate that this assumption is not generally valid. Instead, viewed in light of our results, the interpretation would be that the sentence condition may have had greater involvement of a noise distribution (thereby accounting for the relatively linear shape of the ROC in that condition) and that this may partly explain why elevated activity was not evident for that condition.

In a similar vein, Sauvage et al. (2008) recently used an associative odor recognition procedure with rats in which ROC data were collected using a biasing manipulation. Control rats yielded a linear ROC (and curvilinear z-ROC), whereas rats with hippocampal lesions yielded a curvilinear ROC (and linear ROC). The interpretation of this result was taken from the human literature on associative recognition in which these two trends were interpreted in terms of the DPSD model to reflect recollection-based responding for the control rats and familiarity-based responding for the hippocampal rats. However, if those trends were due to the same forces that yield linear and curvilinear associative recognition ROCs in humans, the results of the present study suggest that, for whatever reason, the behavior of the control rats was governed by a mixture model (more so than the behavior of the hippocampal rats). Alternatively, the biasing manipulation they used to produce the ROC data may have introduced elements that render the comparison to confidence-based ROC data meaningless (Wixted & Squire, 2008).

The Role of Familiarity in Associative Recognition

Nothing in the present research indicates that unitized familiarity never plays a role in associative recognition. Indeed, if the two items making up a test pair happen to be BLACK and BERRY, it seems reasonable to assume that the two words, considered as a unit, may yield a familiarity signal. Our point is not that unitized familiarity never plays a role. Instead, our point is that previous results that have been interpreted as indicative of unitized familiarity (namely, a curvilinear associative recognition ROC) do not specifically reflect that process. Instead, a continuous recollection signal plays a major role in producing curvilinear ROCs. As such, our findings reinforce the long-held assumption that successful performance on an associative recognition task is largely based on recollection.

Cohn and Moscovitch (2007) recently drew a distinction between associative reinstatement, such as distinguishing between intact and rearranged pairs based on recollection, and associative identification, a kind of associative familiarity. The latter is revealed in old/new recognition tests of the kind used in Experiment 2 in which participants were presented with intact pairs, rearranged pairs and new pairs and are required to indicate which words are old (regardless of whether they were intact or rearranged) and which are new. The phenomenon of interest is that on this old/new test, intact pairs are more likely to be judged old than rearranged pairs even though the items of those pairs are equally familiar. This phenomenon was observed in our experiment: the hit rate for intact pairs in Experiment 2 was 0.80, whereas the hit rate for rearranged pairs was 0.72. Cohn and Moscovitch (2007) attribute this difference to associative familiarity, which is a concept that is not identical to but is not unlike unitized familiarity (cf. Speer & Curran, 2007). While this familiarity process may have played some role in distinguishing old pairs from new pairs, it is not clear whether it plays a role when participants attempt to distinguish intact pairs from rearranged pairs (though it might).

Our findings are silent on the question of whether or not unitized familiarity plays any role in associative recognition decisions, but they reinforce the notion that *item* familiarity plays a role (albeit not a useful one). Although item familiarity does not help to distinguish intact from rearranged pairs because all of the items are equally familiar, Kelley and Wixted (2001) found evidence that subjects nevertheless take item information into account and combine it with associative recollection in arriving at a decision. Specifically, they found that strong intact pairs had a much higher hit rate than weak intact pairs (as any model would predict), but the false alarm rate to rearranged pairs was constant across strength conditions. We replicated that pattern in Experiment 1, and it has been observed in a number of other studies as well (Buchler, et al., 2008; Gallo, et al., 2004; Jones & Jacoby, 2001; Kelley & Wixted, 2001; Malmberg & Xu, 2007; Verde & Rotello, 2004).

The constant false alarm rate across strength conditions has been interpreted to mean that the increased familiarity of the strong items disposes subjects to declare the pair to be intact (even though item familiarity is not diagnostic of the pair's status), whereas increased recollection disposes the subjects to correctly declare the pair to be rearranged. These opposing forces cancel each other and leave the false alarm rate unchanged. The present study lends new support to that interpretation. Specifically, while the overall false alarm rate (and the overall correct rejection rate) remained constant as a function of strength in Experiment 1, Remember judgments associated with correct rejections increased significantly with strength. Thus, the absence of an effect of strength on the Remember correction rejection and false alarm rates does not mean that memory for rearranged pairs was unaffected by the strength manipulation. The change in Remember/Know judgments to these pairs across strength conditions shows that memory was, indeed, affected. This result provides direct evidence that recollection was higher in the strong condition, which should have led to a lower false alarm rate to rearranged pairs. Higher item familiarity in the strong condition was, presumably, a countervailing force that served to prevent that from happening and, as such, prevented a mirror effect that might otherwise have occurred (cf. Buchler et al., 2008). The mirror effect refers to a common finding in which conditions associated with better memory exhibit both a higher hit rate and a lower false alarm rate than conditions associated with worse memory.

Note that a mirror effect is generally observed on associative recognition tasks when different classes of memorable and non-memorable pairs are used (Greene, 1996; Hockley, 1994). Thus, for example, Greene (1996) reported that intact word pairs exhibit a higher hit rate than intact nonwords, and rearranged word pairs exhibit a lower false alarm rate than rearranged nonword pairs. In this case, words presumably have a considerable recollection advantage over non-words, whereas the familiarity differences for the words and nonwords (all of which were recently seen on a list) may be less pronounced. If so, the recollection advantage of words would yield a higher hit rate and a lower false alarm rate (cf. Xu & Malmberg, 2007).

In summary, the results of these two experiments suggest that memory strength on an associative recognition task is a continuously distributed variable that is a joint function of diagnostic associative recollection and non-diagnostic item familiarity. Because the aggregated memory processes are continuous, performance is well characterized by a signal-detection model (though a mixture signal-detection model applies when memory is weak). Thus, a curvilinear z-ROC cannot be interpreted to indicate the presence of threshold recollection, and the recollection parameter estimates obtained from fitting the DPSD model to associative recognition ROC data may capture the influence of a noise distribution, not threshold recollection. Because the results suggest that both recollection and familiarity are continuous variables, it seems reasonable to suppose that they are aggregated into a unidimensional memory strength signal. If so, then dual-process theory and signal-detection theory, which have long been viewed as rival accounts, are fully compatible.

Table 1. *Maximum likelihood parameter estimates and chi-square goodness-of-fit statistics based on a fit of the MSD model to the weak and strong ROC data from Experiment 1.*

	μ_{Rearr}	μ_{Intact}	**σ**	$\lambda_{\textbf{Weak}}$	$\lambda_{\textbf{Strong}}$	χ^2	**df**
Group	-2.08	4.78	3.09	0.62	1.00	22.9	10
Individual	-1.86	4.53	2.63	0.55	0.96	176.8	147

Table 2. *Maximum likelihood parameter estimates and chi-square goodness-of-fit statistics based on a fit of the DPSD model to the weak and strong ROC data from Experiment 1.*

	R	δ	σ	d'_{Weak}	d'_{Strong}	χ^2	**df**
Group	0.37	-0.26	1.61	0.20	1.33	27.6	10
Individual	0.30	-0.04	1.75	0.35	1.62	176.0	146

Table 3. *Maximum likelihood parameter estimates and chi-square goodness-of-fit statistics based on a fit of the MSD model to the associative recognition ROC data from Experiment 2.*

	μ_{Rearr}	μ_{Intact}	λ	*df*	χ^2
Grp	-1.09	2.14	0.51	2	2.89
Ind	-1.43	2.36	0.62	37	51.25

Table 4. *Maximum likelihood parameter estimates and chi-square goodness-of-fit statistics based on a fit of the DPSD model to the associative recognition ROC data from Experiment 2.*

	R_{Rearr}	R_{Intact}	***d'***	***df***	χ^2
Grp	0.08	0.38	0.63	2	3.15
Ind	0.25	0.33	0.46	37	51.96

Table 5. *Parameter estimate and chi-square goodness-of-fit statistics from a maximum-likelihood fit of the EVSD model to the ROC, and quadratic coefficients from a least-squares fit of a* 2^{nd}*-order polynomial to the z-ROC, for the collapsed associative recognition ROC and for the three conditional ROCs from Experiment 2.*

ROC Type	**EVSD**		**z-ROC**
	d'	$\boldsymbol{\chi^2(4)}$	Quadratic Coefficient
Collapsed	1.17	19.56	0.22
O/N 1-3	0.09	5.5	0.33
O/N 4-5	0.51	0.51	0.04
O/N 6	1.63	4.28	0.03

Figure Captions

Figure 1. An illustration of the competing models of associative recognition ROC data. The MSD model assumes that the decision axis represents associative recollection and that some intact pairs and some rearranged pairs are drawn from a nondiagnostic noise distribution (labeled Intact-/Rearr) in the weak condition (**A**). If a symmetrical curvilinear ROC (and linear z-ROC) is observed in the strong condition, this model assumes that all items are drawn from diagnostic intact (Intact+) or rearranged (Rearr+) distributions (**B**). The DPSD model assumes the decision axis represents unitized familiarity and that that some intact pairs and some rearranged pairs are associated with threshold recollection (which occurs with probability R) in the weak condition (**C**). In the strong condition, all pairs are drawn from the unitized familiarity distributions (**D**).

Figure 2. Hit, miss, correct rejection (CR) and false alarm (FA) rates obtained in the weak and strong conditions of Experiment 1. A hit is defined as the proportion of intact pairs that were correctly declared to be intact, whereas a false alarm is defined as the proportion of rearranged pairs that were incorrectly declared to be intact. The miss rate is 1 minus the hit rate, and the correct rejection rate is 1 minus the FA rate.

Figure 3. ROC data (**A**) and z-ROC data (**B**) for the weak and strong conditions of Experiment 1. The curves represent least-squares fit of a 2nd-order polynomial.

Figure 4. Hit, miss, correct rejection and false alarm rates associated with Remember judgments (R-Hit, R-miss, R-CR and R-FA, respectively) and Know judgments (K-Hit, K-miss, K-CR and K-FA, respectively) in the weak and strong conditions of

Experiment 1, with the appropriate measures shown separately for intact pairs (**A**) and rearranged pairs (**B**).

Figure 5. Proportion of items recalled on an unexpected cued recall test administered after the associative recognition test. The recall data are shown separately for hit, miss, correct rejection and false alarm rates associated with Remember judgments (R-hit, R-miss, R-CR and R-FA, respectively) and Know judgments (K-hit, K-miss, K-CR and K-FA, respectively) in the weak and strong conditions of Experiment 1, with the appropriate measures shown separately for intact pairs (**A**) and rearranged pairs (**B**).

Figure 6. Proportion of items recalled on an unexpected cued recall test administered after the associative recognition test as a function of confidence in the intact vs. rearranged recognition decision for intact pairs (**A**) and rearranged pairs (**B**) for the weak and strong conditions of Experiment 1. To reduce variability, the data were averaged over adjacent values on the 6-point confidence scale.

Figure 7. Associative recognition ROC data (**A**) and z-ROC data (**B**) from Experiment 2. The curves represent least-squares fit of a second-order polynomial.

Figure 8. Associative recognition ROC and z-ROC data from Experiment 2 for pairs consisting of items not recognized as having appeared on the list (**A** and **B**, respectively), for pairs of items recognized as having appeared on the list with low or medium confidence (**C** and **D**, respectively), and for pairs of items recognized as having appeared on the list with high confidence (**E** and **F**, respectively). The curves drawn through the ROC data in the left panels represent the best-fitting EVSD model

according to maximum likelihood estimation; the curves drawn through the z-ROC data in the right panels represent the least squares fit of a 2nd-order polynomial.

Figure 9. Proportion of correct and incorrect Remember judgments (R+ and R-, respectively) and correct and incorrect Know judgments (K+ and K-, respectively) for the three conditional ROCs shown in Figure 8.

Figure 1

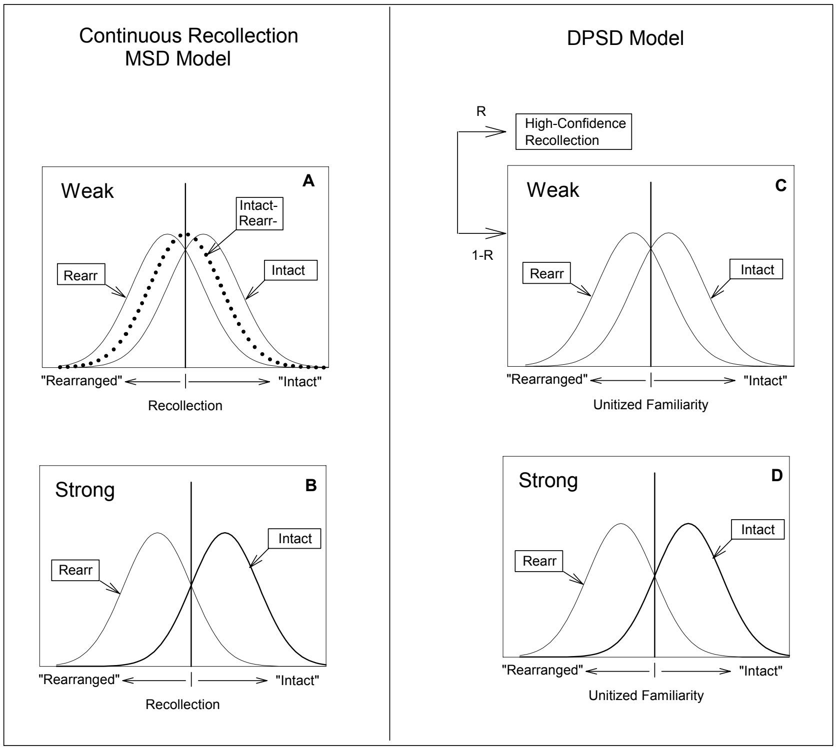


Figure 1. An illustration of the competing models of associative recognition ROC data. The MSD model assumes that the decision axis represents associative recollection and that some intact pairs and some rearranged pairs are drawn from a nondiagnostic noise distribution (labeled Intact-/Rearr) in the weak condition (**A**). If a symmetrical curvilinear ROC (and linear z-ROC) is observed in the strong condition, this model assumes that all items are drawn from diagnostic intact (Intact+) or rearranged (Rearr+) distributions (**B**). The DPSD model assumes the decision axis represents unitized familiarity and that that some intact pairs and some rearranged pairs are associated with threshold recollection (which occurs with probability R) in the weak condition (**C**). In the strong condition, all pairs are drawn from the unitized familiarity distributions (**D**).

Figure 2

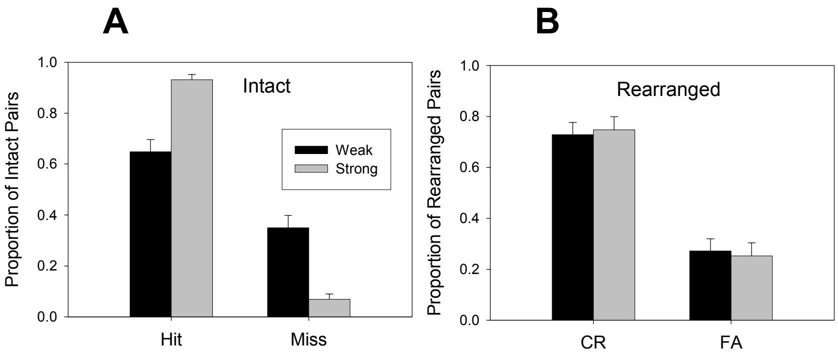


Figure 2. Hit, miss, correct rejection (CR) and false alarm (FA) rates obtained in the weak and strong conditions of Experiment 1. A hit is defined as the proportion of intact pairs that were correctly declared to be intact, whereas a false alarm is defined as the proportion of rearranged pairs that were incorrectly declared to be intact. The miss rate is 1 minus the hit rate, and the correct rejection rate is 1 minus the FA rate.

Figure 3

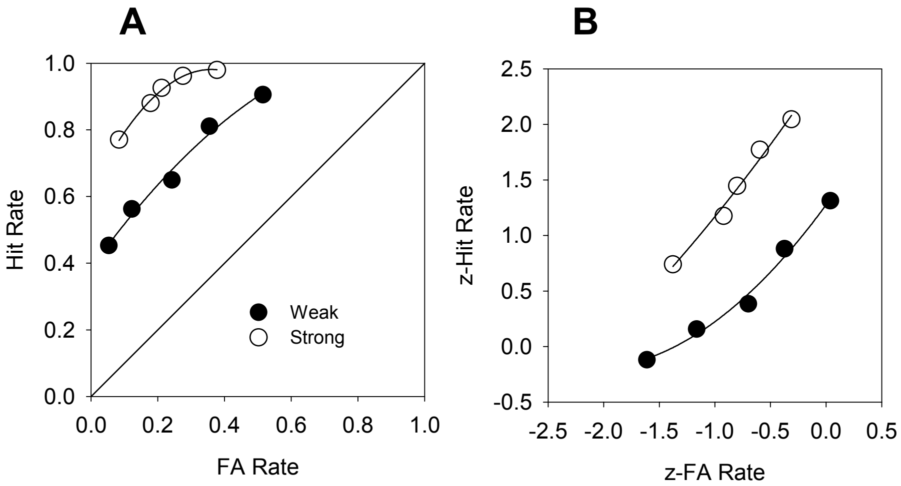


Figure 3. ROC data (**A**) and z-ROC data (**B**) for the weak and strong conditions of Experiment 1. The curves represent least-squares fit of a 2nd-order polynomial.

Figure 4

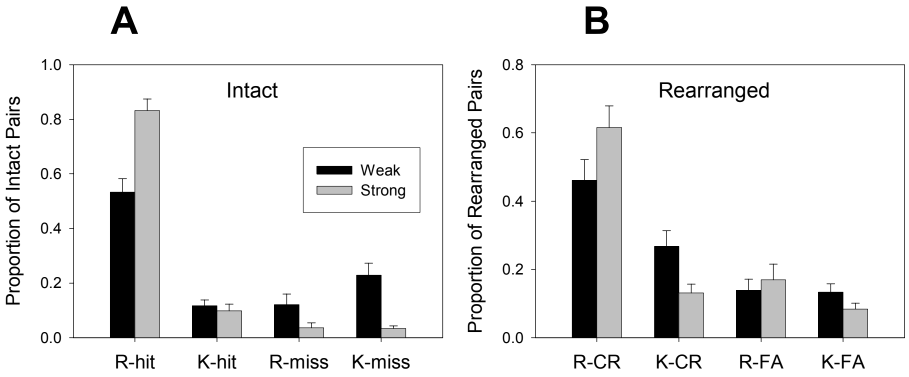


Figure 4. Hit, miss, correct rejection and false alarm rates associated with Remember judgments (R-Hit, R-miss, R-CR and R-FA, respectively) and Know judgments (K-Hit, K-miss, K-CR and K-FA, respectively) in the weak and strong conditions of Experiment 1, with the appropriate measures shown separately for intact pairs (**A**) and rearranged pairs (**B**).

Figure 5

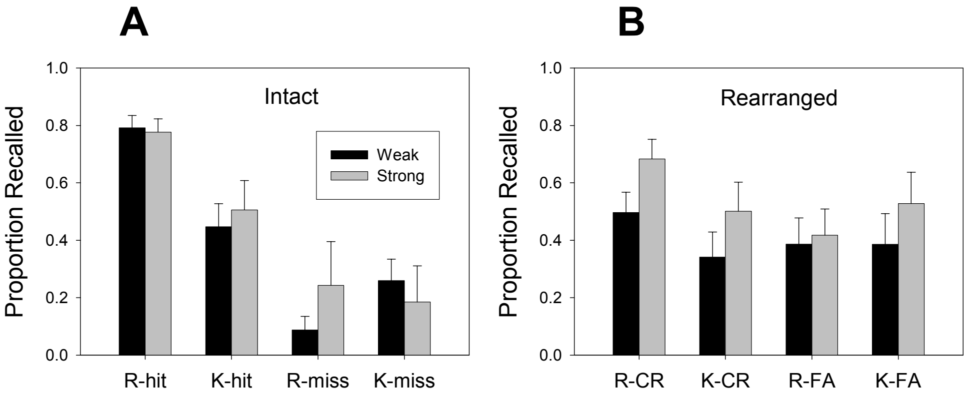


Figure 5. Proportion of items recalled on an unexpected cued recall test administered after the associative recognition test. The recall data are shown separately for hit, miss, correct rejection and false alarm rates associated with Remember judgments (R-hit, R-miss, R-CR and R-FA, respectively) and Know judgments (K-hit, K-miss, K-CR and K-FA, respectively) in the weak and strong conditions of Experiment 1, with the appropriate measures shown separately for intact pairs (**A**) and rearranged pairs (**B**).

Figure 6

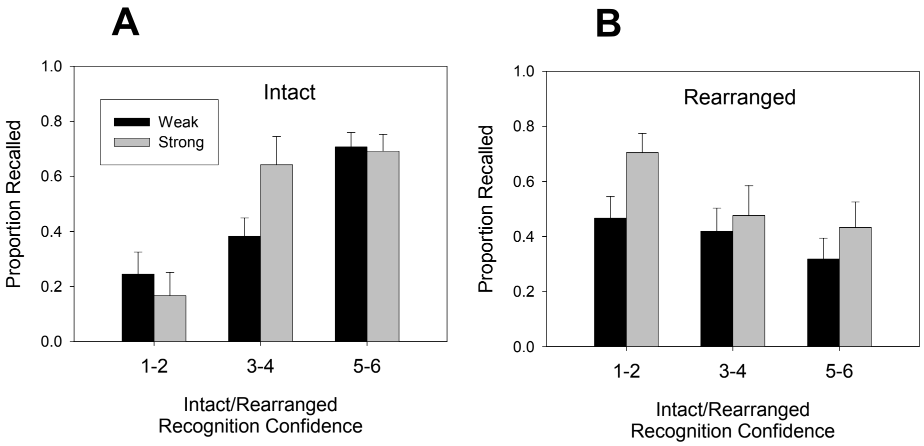


Figure 6. Proportion of items recalled on an unexpected cued recall test administered after the associative recognition test as a function of confidence in the intact vs. rearranged recognition decision for intact pairs (**A**) and rearranged pairs (**B**) for the weak and strong conditions of Experiment 1. To reduce variability, the data were averaged over adjacent values on the 6-point confidence scale.

Figure 7

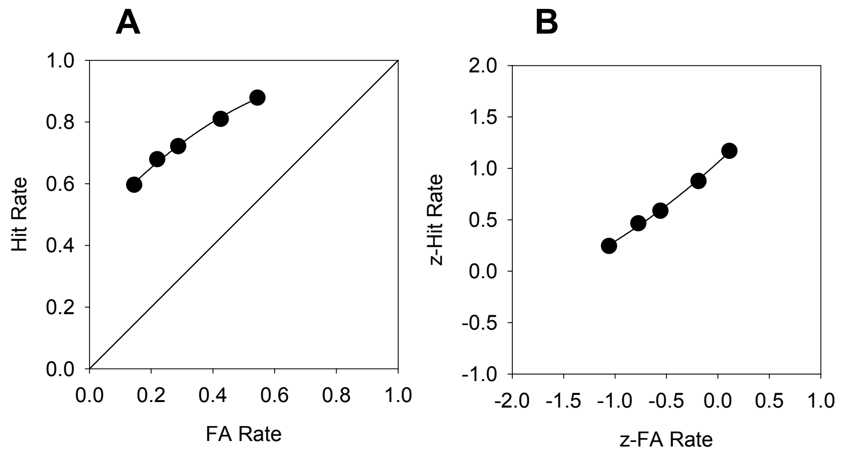


Figure 7. Associative recognition ROC data (**A**) and z-ROC data (**B**) from Experiment 2. The curves represent least-squares fit of a second-order polynomial.

Figure 8

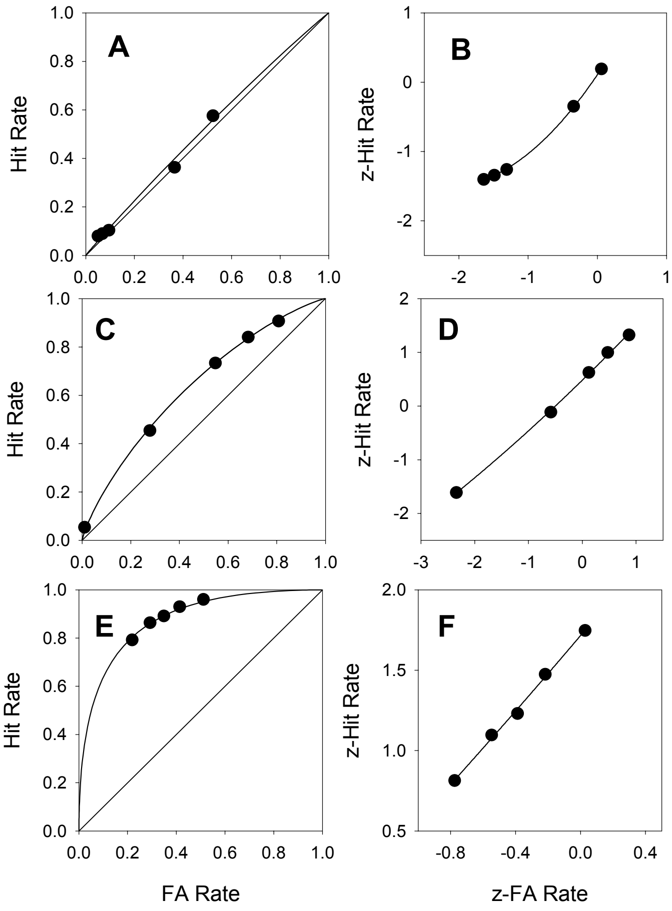
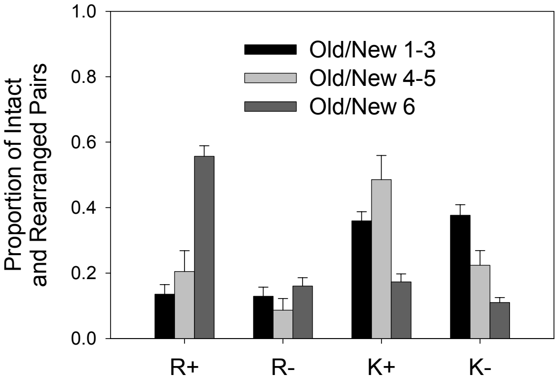


Figure 8. Associative recognition ROC and z-ROC data from Experiment 2 for pairs consisting of items not recognized as having appeared on the list (**A** and **B**, respectively), for pairs of items recognized as having appeared on the list with low or medium confidence (**C** and **D**, respectively), and for pairs of items recognized as having appeared on the list with high confidence (**E** and **F**, respectively). The curves drawn through the ROC data in the left panels represent the best-fitting EVSD model according to maximum likelihood estimation; the curves drawn through the z-ROC data in the right panels represent the least squares fit of a 2nd-order polynomial.

Figure 9



<u>Figure 9</u>. Proportion of correct and incorrect Remember judgments (R+ and R-, respectively) and correct and incorrect Know judgments (K+ and K-, respectively) for the three conditional ROCs shown in Figure 8.

GENERAL DISCUSSION

This body of work supports the unequal variance account of recognition memory and further supports the idea that two memory processes – familiarity and recollection – are continuous and are combined to make recognition decisions. Such an account reconciles two longstanding theories (signal detection theory and dual process theory) that have generally seemed to be irreconcilable.

In Chapter 1, we compared two methods of analysis – a standard Gaussian-based signal detection analysis versus a distribution-free direct ratings analysis, to test the idea that the memory strengths of targets are more variable than the memory strengths of lures. We conducted two experiments that used a new method, one that involved a direct assessment of subjects' memory strength by using a fined grained confidence rating scale. The two methods were surprisingly comparable despite the fact that they rely on completely different assumptions. Even when the underlying memory strength distributions are not assumed to be Gaussian in form, as in the two direct ratings experiments, the target distribution was found to have greater variance than the lure distribution. The observed agreement with the Gaussian-based ROC analysis suggests that confidence ratings can be made using an interval scale. This work also supports the idea that the underlying distributions are approximately Gaussian in form because many other distributional assumptions do not predict that the two methods will agree on the degree of inequality (i.e., that the standard deviation of the lure distribution is .80 times that of the target distribution).

The work completed for the dissertation also lends support to the notion that recollection is a continuous process. These findings help to address an apparent incompatibility between signal detection theory and dual process theory, two accounts that have been at odds with each other for decades. The incompatibility between the two accounts has generally been assumed to be that dual process theory involves two memory processes, whereas signal detection theory seems to involve only one. However, if one allows the possibility that memory strength is a joint function of recollection and familiarity, then there is only one main point of contention between the two theories. That point is in regard to the nature of recollection. All previous dual process models have assumed that recollection is a categorical process (i.e., recollection occurs only with high confidence and high accuracy), whereas the signal detection model must assume that recollection is a continuous process (i.e., confidence and accuracy are graded even when decisions are based on recollection). In Chapter 2, we tested subjects in two experiments using a source memory procedure (thought to be a recollection-based task), and the results indicated that recollection is a continuous process. If so, then signal detection theory and dual process are no longer necessarily at odds with each other because if recollection is continuous, then recollection and familiarity may be combined to yield a continuous memory strength variable.

The nature of recollection was further studied and characterized in Chapter 3, where we compared a signal detection model (specifically the mixture signal detection model) and the DPSD model by testing subjects on an associative recognition task. A typical finding on this task is that ROCs are linear when memory is weak and curvilinear when memory is strong. The DPSD model interprets the linear ROC as being indicative of decisions based on categorical recollection and interprets the increased curvilinearity when memory is strong as being indicative of decisions based on unitized familiarity. By contrast, the signal detection model explains the linear ROC as being indicative of pairs for which no associative information was encoded (i.e., these "noise" pairs are mixed with intact and rearranged pairs for which associative information was encoded), and it interprets the curvilinear ROC in the strong condition as being indicative of continuous recollection (coupled with the relative absence of noise pairs). To test whether the results in the strong condition are due to unitized familiarity or continuous recollection, we employed several methods, including a remember/know procedure, a surprise cued recall test, and a preliminary old/new test. We also analyzed the data using a mixture signal detection model, which involves a third distribution to account for "noise" (i.e., intact and rearranged pairs for which no associative information was encoded at study). The remember/know procedure revealed that the strong memories associated with a curvilinear ROC were due to recollection, not to increased unitized familiarity. The cued recall test further validated this finding. In addition, employing an old/new question prior to the intact/rearranged question enabled us to isolate the noise pairs and remove them from the analysis. These were pairs involving items that were not even recognized as having been presented on the list. Not surprisingly, associative recognition performance for these pairs was no better than chance. Once these word pairs were removed from the associative recognition ROC analysis (leaving only pairs theoretically associated with recollection), even the weak (and ordinarily linear) ROC became clearly curvilinear. A categorical recollection account has no reason to expect this outcome (and no way to accommodate it), but it is just what a continuous recollection account predicts. Taken together, these findings support the idea that recollection is a continuous process.

More evidence that recollection is a continuous process has recently been reported by other laboratories. In a set of experiments similar to those reported here, Slotnick (2010) directly tested the issue of continuous versus categorical recollection. During study, pictures of objects were presented on the left or right of the screen, and during test, subjects first identified whether the objects were originally on the left or right of the screen using a 6-point scale (similar to Experiments 1 and 2 in Chapter 2) and next provided a remember, know, or new judgment. Two signal detection models and two threshold models (including the DPSD model) were fit to the data. The signal detection models clearly fit the data the best, which led Slotnick to this conclusion:

> "The present findings were inconsistent with the threshold models of recollection, but supported the continuous models of recollection that can be described by a classic signal detection model…"

Another level of analysis that was recently reported also supports our claim from Chapters 2 and 3 (Johnson, McDuff, Rugg, & Norman, 2009). This was an fMRI study designed to measure cortical reinstatement of encoding activity patterns at retrieval (thought to be the neural signature of recollection). Previously, cortical reinstatement was believed to occur only when source information was recollected, as indicated by a "remember" response. However, Johnson et al. (2009) scanned subjects at both encoding and retrieval and found that cortical reinstatement of encoding patterns also occurred for Know judgments, but to a lesser degree than for Remember judgments. Based on this, Johnson, et al. 2009 drew the following conclusion:

> “The findings are interpreted as support for a continuous, recollection-related neural signal that has been central to recent debate over the nature of recognition memory processes.”

The prominent DPSD model that has dominated various fields (i.e., psychology, neuroscience, psychiatry) by influencing how researchers interpret recognition memory performance has clearly begun to lose support. Its key flaw is that it assumes that recollection is a categorical process. If recollection is instead a continuous process, the voluminous body of research that has been guided by this influential theory may need to be reinterpreted. All of the findings reported in this dissertation underscore the need for such a reinterpretation.

The human capability for memory is often taken for granted, but it is a truly amazing ability that is generally not appreciated until it is lost or somehow compromised. Much of our personal identity is tied to our memories, and for that reason alone, it is important to gain a basic understanding of how memory works. Applied work with various patient populations can also benefit from advancing basic knowledge about the normal operation of memory. Indeed, in recent years, there has been increased attention paid to memory dysfunction associated with depression, schizophrenia, anxiety (especially Post Traumatic Stress Disorder), and dementia. A deeper understanding of how memory operates in unimpaired individuals may contribute to a better understanding of these various disorders. In this regard, the findings described in this dissertation contribute to a growing literature suggesting that recognition memory processes are continuous in nature and can be usefully conceptualized in terms of signal detection theory.

References

Atkinson, R. C., & Joula, J. F. (1973). Factors influencing speed and accuracy of word recognition. In S. Kornblum (Ed.), *Attention and Performance IV* (pp. 583-612). New York: Academic Press.

Atkinson, R. C., & Juola, J. F. (1974). Search and decision processes in recognition memory. In D. H. Krantz, R. C. Atkinson, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (pp. 243-290). San Francisco: Freeman.

Bodner, G. E., & Lindsay, D. S. (2003) Remembering and knowing in context. *Journal of Memory and Language, 48*, 563–580.

Buchler, N. G., Light, L. L., & Reder, L. M. (2008). Memory for items and associations: Distinct representations and processes in associative recognition. *Journal of Memory and Language, 59*, 183–199.

Clark, S. E. (1992). Word frequency effects in associative and item recognition. *Memory & Cognition, 20,* 231–243.

Clark, S. E., & Hori, A. (1995). List length and overlap effects in forced-choice associative recognition. *Memory & Cognition, 23,* 456-461.

Cohn, M., & Moscovitch, M. (2007). Dissociating direct and indirect measures of associative memory. *Journal of Memory and Language*, *57*, 437-454.

Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology, 33A*, 497–505.

Daselaar, S. M., Fleck, M. S. & Cabeza, R. (2006). Triple dissociation in the medial temporal lobes: recollection, familiarity, and novelty. *Journal of Neurophysiology, 96*, 1902-1911.

DeCarlo, L. T. (2002). Signal detection theory with finite mixture distributions: Theoretical developments with applications to recognition memory. *Psychological Review, 109,* 710–721.

DeCarlo, L. T. (2003). An application of signal detection theory with finite mixture distributions to source discrimination. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 767–778.

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*, 730-740.

Dodson, C. S., Holland, P. W., & Shimamura, A. P. (1998). On the recollection of specific- and partial-source information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1121- 1136.

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition, 24,* 523-533.

Dunn, J. C. (2004). Remember–know: A matter of confidence. *Psychological Review, 111,* 524-542.

Dunn, J. C. (2008). The dimensionality of the Remember-Know task: A state-trace analysis. *Psychological Review, 115*(2), 426-446.

Egan, J. P. (1958). Recognition memory and the operating characteristic (Tech. Note AFCRC-TN-58–51). Bloomington: Indiana University, Hearing and Communication Laboratory.

Egan, J. P. (1975). *Signal detection theory and ROC analysis*. Academic Press: New York.

Eichenbaum, H., Yonelinas, A. P., & Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annual Review of Neuroscience*, *30*, 123-152.

Eldridge, L. L., Knowlton, B. J., Furmanski, C. S., Bookheimer, S. Y. & Engel, S. A. (2000). Remembering episodes: a selective role for the hippocampus during retrieval. *Nature Neuroscience, 3*, 1149-1152.

Eldridge, L. L., Sarfatti, S., & Knowlton, B. J. (2002). The effect of testing procedure on remember–know judgments. *Psychonomic Bulletin & Review, 9*, 139-145.

Gallo, D. A., Sullivan, A. L., Daffner, K. R., Schacter D. L., & Budson, A. E. (2004). Associative recognition in Alzheimer's disease: Evidence for impaired recall-to-reject. *Neuropsychology 18*, 556-563.

Gardiner, J., Ramponi, C., & Richardson-Klavehn, A. (1998). Experiences of remembering, knowing, and guessing. *Consciousness and Cognition*, *7*, 103-107.

Gilhooly, K. J. & Logie, R. H. (1980). Age of acquisition, imagery, concreteness, familiarity and ambiguity measures for 1944 words. *Behaviour Research Methods & Instrumentation, 12*, 395-427.

Giovanello, K. S., Verfaellie, M., & Keane, M. M. (2003). Disproportionate deficit in associative recognition relative to item recognition in global amnesia. *Cognitive Affective Behavioral Neuroscience, 3,* 186-194.

Glanzer, M., Kim, K., Hilford, A., & Adams, J. K. (1999). Slope of the receiver-operating characteristic in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 500– 513.

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

Greene, R. L. (1996). Mirror Effect in Order and Associative Information: Role of Response Strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 687-695.

Gronlund, S. D., & Ratcliff, R. (1989). Time course of item and associative information: Implications for global memory models. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15,* 846–858.

Habib, R., & Nyberg, L. (2008). Neural correlates of availability and accessibility in memory. *Cerebral Cortex, 18*, 1720-1726.

Haskins, A. L., Yonelinas, A. P., Quamme, J. R., & Ranganath, C. (2008). Perirhinal Cortex Supports Encoding and Familiarity-Based Recognition of Novel Associations. *Neuron, 59* (4), 554-560.

Healy, M. R., Light, L. L., & Chung, C. (2005). Dual-process models of associative recognition in young and older adults: Evidence from receiver operating characteristics. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 768–788.

Hockley, W. E. (1992). Item versus associative information: Further comparisons of forgetting rates. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18,* 1321–1330.

Hockley, W. E. (1994). Reflections of the mirror effect for item and associative information. *Memory & Cognition, 22*, 713-722.

Hockley, W. E. & Consoli, A. (1999). Familiarity and recollection in item and associative recognition. *Memory & Cognition 27*, 657–664.

Hockley, W. E., & Cristi, C. (1996). Tests of encoding tradeoffs between item and associative information. *Memory & Cognition, 24*, 202-216.

Humphreys, M. S. (1978). Item and relational information. A case for context independent retrieval. *Journal of Verbal Learning & Verbal Behavior, 17,* 175-188.

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

Jacoby, L. L., Toth, J. P. & Yonelinas, A. P. (1993). Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General, 122*, 139- 154.

Johnson, J. D., McDuff, S. G. R., Rugg, M. D., & Norman, K. A. (2009). Recollection, familiarity, and cortical reinstatement: A multivoxel pattern analysis. *Neuron, 63,* 697-708.

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

Jones, T. C., & Jacoby, L. L. (2001). Feature and conjunction errors in recognition memory: Evidence for dual-process theory. *Journal of Memory and Language, 45,* 82-102.

Kelley, R., & Wixted, J. T. (2001). On the nature of associative information in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 701-722.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection Theory: A User's Guide* (2nd ed.). Mahwah, N.J.: Lawrence Erlbaum Associates.

Malmberg, K. J. (2002). On the form of ROCs constructed from confidence ratings. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 380-387.

Malmberg, K. J. & Xu, J. (2007). On the flexibility and on the fallibility of associative memory, *Memory & Cognition, 35*, 545-556.

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

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,* 509-515.

Montaldi, D., Spencer, T. J., Roberts, N. & Mayes, A. R. (2006). The neural system that mediates familiarity memory. *Hippocampus 16*, 504-520.

Nobel, P. A. & Shiffrin, R. M. (2001). Retrieval processes in recognition and cued recall. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 27*, 384-413.

Parks, C. M. & Yonelinas, A. P. (2007). Moving beyond pure signal-detection models: Comment on Wixted. *Psychological Review, 114* (1), 188-202.

Qin, J., Raye, C. L., Johnson, M. K., & Mitchell, K. J. (2001). Source ROCs are (typically) curvilinear: Comment on Yonelinas (1999). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 1110-1115.

Quamme, J. R., Yonelinas, A. P., & Norman, K. A. (2007). Effect of unitization on associative recognition in amnesia. *Hippocampus, 17* (3), 192-200.

Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition, 21,* 89-102.

Ratcliff, R., Sheu, C. F., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review, 99,* 518-535.

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

Rutishauser, U., Schuman, E. M. & Mamelak, A. N. (2008). Activity of human hippocampal and amygdala neurons during retrieval of declarative memories. *Proceedings of the National Academy of Sciences USA, 105*, 329-334.

Sauvage, M. M., Fortin, N. J., Owens, C. B, Yonelinas, A. P., & Eichenbaum, H. (2008). Recognition memory: Opposite effects of hippocampal damage on recollection and familiarity. Nature Neuroscience, *11*, 16-18.

Slotnick, S. D. (2010). “Remember” source memory ROCs indicate recollection is a continuous process. *Memory, 18,* 27-39.

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

Speer, N. K., & Curran, T. (2007). ERP correlates of familiarity and recollection processes in visual associative recognition. *Brain Research, 1174*, 97-109.

Squire, L. R., Wixted, J. T., & Clark, R. E. (2007). Recognition memory and the medial temporal lobe: A new perspective. *Nature Reviews Neuroscience*, *8*, 872-883.

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*, 238-242.

Staresina, B.P. & Davachi, L. (2006). Differential encoding mechanisms for subsequent associative recognition and free recall. *The Journal of Neuroscience 26*, 9162-9172.

Starns, J. J., & Ratcliff, R. (2008). Two dimensions are not better than one: STREAK and the univariate signal detection model of remember/know performance. *Journal of Memory and Language, 59,* 169-182.

Tulving, E. & Thompson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*, 352-373.

Turriziani, P., Fadda, L., Caltagirone, C., & Carlesimo, G. A. (2004). Recognition memory for single items and for associations in amnesic patients. *Neuropsychologia, 42,* 426-433.

Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language, 50,* 289-335.

Verde, M. F. (2004). The retrieval practice effect in associative recognition. *Memory & Cognition, 32*, 1265-1272.

Verde, M. F., & Rotello, C. M. (2004). Strong memories obscure weak memories in associative recognition. *Psychonomic Bulletin & Review, 11,* 1062-1066.

Vilberg, K. L. & Rugg, M. D. (2007). Dissociation of the neural correlates of recognition memory according to familiarity, recollection, and amount of recollected information. *Neuropsychologia, 45*, 2216-2225.

Wais, P. E., Mickes, L. & Wixted, J. T. (2008). Remember/Know Judgments Probe Degrees of Recollection. Journal of Cognitive Neuroscience, 20, 400-405.

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

Wixted, J. T., & Squire, L. R. (2008). Constructing Receiver Operating Characteristics (ROCs) with Experimental Animals: Cautionary Notes. *Learning & Memory, 15*, 687-690.

Wixted, J. T., & Stretch, V. (2004). In defense of the signal detection interpretation of remember/know judgments. *Psychonomic Bulletin & Review, 11*, 616–641.

Xu, J. & Malmberg, K. J. (2007). Modeling the effects of verbal and nonverbal pair strength on associative recognition. *Memory & Cognition, 35*, 526-544.

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*, 1341-1354.

Yonelinas, A. P., & Jacoby, L. L. (1995). The relation between remembering and knowing as bases for recognition: Effects of size congruency. *Journal of Memory and Language, 34(5),* 622–643.

Yonelinas, A. P. (1997). Recognition memory ROCs for item and associative information: The contribution of recollection and familiarity. *Memory & Cognition, 25,* 747–763.

Yonelinas, A. P. (1999). The contribution of recollection and familiarity to recognition and source memory: An analysis of receiver operating characteristics and a formal model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 1415-1434.

Yonelinas, A. P. (2001). Components of episodic memory: The contribution of recollection and familiarity. *The Philosophical Transactions of the Royal Society, Series B, 356*, 1363-1374.

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

Yonelinas, A. P., Otten, L., Shaw, K. N. & Rugg, M. D. (2005). Separating the brain regions involved in recollection and familiarity-strength in recognition memory. *Journal of Neuroscience, 25*, 3002-3008.

Yonelinas, A. P., & Parks, C.M. (2007). Receiver Operating Characteristics (ROCs) in Recognition Memory: A Review. *Psychological Bulletin, 133*, 800-832.