Behavioral and Neurocomputational Investigations
of Human Recognition Memory and Cognition

Kane William Elfman

Submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy
2015

©Copyright by Kane William Elfman, 2015.
All rights reserved.
Abstract

Models of human recognition memory frequently differentiate between processes of recollection, as the retrieval of qualitative information about a past event, and familiarity, as a more continuous process that matches a currently perceived stimulus against an accumulation of many similar past experiences. Cognitive models have made important advances in our understanding of memory functioning, but have at times struggled to account for some more nuanced empirical findings. Recently, computational models of the brain have provided an alternative, but complementary perspective by simulating networks of neurons that communicate with one another and learn to associate external events with internal patterns of activation. The research presented for this thesis combines behavioral and computational methodologies to investigate aspects of recognition memory and related cognitive domains that have to date been poorly understood.

The current research is presented as four peer-reviewed papers. The first paper, Testing a Neurocomputational Model of Recollection, Familiarity, and Source Recognition (Elfman, Parks & Yonelinas, 2008), tests a computationally motivated hypothesis that, under conditions of high stimulus similarity, the threshold nature of recollection will begin to break down and become more continuous. The results of four behavioral experiments validated this hypothesis and thus supported the model. The second paper, Variations in Recollection: The Effects of Complexity on Source Recognition (Parks, Murray, Elfman & Yonelinas, 2011), builds on recent findings that the retrieval functions associated with perceptually rich stimuli differ from those associated with more simple stimuli. Behavioral experiments supported the hypothesis that the difference between those functions is attributable to recollection effects. In addition, computational modeling work suggested that increased stimulus complexity is akin to increased stimulus overlap and thus offers a theoretically plausible explanation for the observed effects. The third paper, Recollection and Familiarity Exhibit Dissociable Similarity Gradients: A Test of the Complementary Learning Systems Model (Elfman & Yonelinas, 2015) examines how varying the similarity between retrieval cues and their respective study events differentially effects recollection and familiarity performances. Simulations indicated that recollection, which is associated with activity in the
hippocampus, transitions discretely from no retrieval to highly accurate retrieval at a threshold point during a gradual shift from low to high similarity. In contrast, familiarity, which is associated with the medial temporal lobe cortex, demonstrated a more gradual and linear transition. Results from behavioral experiments were consistent with the predictions of the model and validated the computational model. The fourth paper, *A Neurocomputational Account of Memory and Perception: Thresholded and Graded Signals in the Human Hippocampus* (Elfman, Aly & Yonelinas, 2014), explores the hypothesis that the hippocampus, which has traditionally been considered exclusively a memory system, is also critical to the performance of complex perceptual discriminations. Simulations with a standard hippocampal model showed how this single brain region can naturally perform tasks associated with both domains.

The findings from all four papers are discussed as an integrated body of work and critiqued in light of the existing literature. The concluding chapter demonstrates how the current work illuminates our understanding of the processes underlying recognition memory and related cognitive functions, and discusses implications for future research.
Acknowledgments

Thank you to Dr Andrew Pipingas for serving as primary supervisor, who has been both patient and supportive throughout my research. I also appreciate his giving me the latitude to follow my own interests, including facilitating my travel to UC Davis. Thanks also to Dr David Crewther, co-supervisor, for providing additional feedback on the thesis. And thank you to Dr Andrew Yonelinas, co-supervisor at UC Davis. Dr Yonelinas has been incredibly generous, accepting me as an extended visitor in his laboratory, facilitating many research experiments, co-authoring journal papers, and organizing financial support.

Acknowledgement is also given to the members of the Human Memory Lab at UC Davis, past and present, for their feedback and friendship. I also acknowledge the Services for International Students and Scholars at UC Davis for assisting with administrative requirements. Thank you also to the Human Memory Lab and the National Institute of Mental Health for funding much of my research and keeping a roof over my head while at Davis, and to Swinburne University for accepting me into the PhD program and providing additional resources.
Declaration

This thesis contains no material which has been accepted for the award to the candidate of any other degree or diploma, except where due reference is made in the text of this thesis. To the best of my knowledge, this thesis contains no material previously published or written by another person except where due reference is made in text. Where the work is based on joint research or publications, this thesis discloses the relative contributions of the respective authors. I warrant that I have obtained, where necessary, permission from the copyright owners to use any third party copyright material reproduced in the thesis (such as artwork, images, unpublished documents), or to use any of my own published work (such as journal articles) in which the copyright is held by another party (such as publisher, coauthor).

Signed  

Kane Elfman

Dated  

October 15, 2014
Table of Contents

Table of Contents ....................................................................................................... vi
List of Figures (Excluding Papers) ............................................................................ viii
List of Papers for Thesis ............................................................................................. ix

Chapter 1: Introduction ................................................................................................. 1
  1.1 Thesis Overview ............................................................................................... 1
  1.2 Background Theory and Literature ................................................................ 3
    1.2.1 Recollection and Familiarity ..................................................................... 3
    1.2.2 Cognitive Models of Recognition Memory ............................................... 6
    1.2.3 Neurocomputational Properties of Learning ........................................... 15
    1.2.4 The Medial Temporal Lobe and Episodic Memory ............................... 20
    1.2.5 Memory and Perception: An Emerging Perspective ......................... 31

Chapter 2: Testing a Neurocomputational Model of Recollection, Familiarity, and Source Recognition ................................................................................................................. 32

Chapter 3: Variations in Recollection: The Effects of Complexity on Source Recognition ....................................................................................................................... 51

Chapter 4: Recollection and Familiarity Exhibit Dissociable Similarity Gradients: A Test of the Complementary Learning Systems Model .............................................. 66

Chapter 5: A Neurocomputational Account Of Memory and Perception: Thresholded and Graded Signals in the Human Hippocampus ........................................ 85

Chapter 6: Discussion .................................................................................................. 102
  6.1 Summary of Findings .................................................................................... 102
    6.1.1 Predicting the Breakdown of the Recollection Threshold ..................... 102
    6.1.2 Complexity Affects the Characteristics of Recollection ....................... 105
    6.1.3 Similarity Gradients of Recollection and Familiarity ......................... 109
    6.1.4 Memory and Perception Signals in the Hippocampus ..................... 112
  6.2 A Critical Review of Methodology and Frameworks ................................ 115
    6.2.1 On the Use of Cognitive Models ........................................................... 115
    6.2.2 On the Use of Computational Models ................................................... 119
6.3 Concluding Commentary .................................................................122
References ..........................................................................................126
Appendix A: Authorship Indication Forms .............................................139
Appendix B: Journal Reprint Licenses ...................................................144
Appendix C: Ethics Certification .............................................................162
List of Figures (Excluding Papers)

Figure 1. Memory strength distributions for a signal detection familiarity process and a threshold recollection process. ..........................................................7

Figure 2. Illustrations of typical receiver operating characteristic (ROC) curves for an item recognition test (left), and a source recognition test (right). ...............11

Figure 3. Schematic of the hippocampal network used in the current computational models.................................................................23

Figure 4. A conceptual representation of pattern separation and pattern completion. ...........................................................................................................26

Figure 5. Schematic of the hippocampus and neocortex networks based on the complementary learning systems model and applied to learning of a pair of house scenes......................................................................................28
List of Papers for Thesis


Chapter 1: Introduction

1.1 Thesis Overview

Evidence from behavioral experiments, patient testing and neuroimaging studies indicates that recognition memory is subserved by at least two functionally and neuroanatomically distinct processes, commonly referred to as recollection and familiarity (for reviews, see Curran, 2000; Eichenbaum, Yonelinas & Ranganth, 2007; Skinner & Fernandes, 2007; Yonelinas, 2002; Yonelinas et al., 2010). Cognitive models (also referred to as cognitive, quantitative, mathematical, or psychological models) of recollection and familiarity have contributed greatly to our understanding of these processes in so far as they have been shown to correlate with people’s performance on tests of item and source recognition memory. More recently, neurocomputational models of memory have matured to offer compelling, biologically grounded explanations of how these phenomena can arise through dynamic interactions within and between different brain regions and sub-regions (e.g., Hasselmo, Bodelon & Wyble, 2002; Hasselmo & Eichenbaum, 2005; Hasselmo & Wyble, 1997; McClelland, McNaughton & O'Reilly, 1995; McNaughton & Morris, 1987; Rolls, 1996; Rudy & O’Reilly, 1999; Treves & Rolls, 1994). In addition, the development of computationally motivated principles of brain functioning has led to new and unexpected predictions that are helping us to better understand the learning mechanisms of the brain, even extending to new hypotheses regarding the nature of memory dysfunction (Norman & O’Reilly, 2003).

The present dissertation investigates some of the novel memory and cognition performance predictions arising from a combination of behavioral and neurocomputational methodological approaches. The primary aim of the research was to test and, in some instances, expand on current models of recognition memory and related cognitive processes, so as to further our understanding of how these systems function in challenging stimulus environments. The first paper begins by investigating whether a prominent neurocomputational model of memory can account for the confidence functions typically associated with people’s performance on tests of source, or associative, memory. It then goes on to examine an untested
prediction of the model (Norman & O’Reilly, 2003), that under conditions in which learned stimuli are highly similar, the threshold nature of recollection will start to break down and become more continuous. Such boundary conditions are not predicted by cognitive models, and so the study represents an important test of the computational approach. Similarly, the second study investigates findings of atypical recognition behavior, this time regarding the effects of stimulus complexity on performance. Whereas previous research has tended to use relatively simple stimuli for lists designed to test recognition memory, some work has indicated that retrieval functions become quite different for content- or context-rich stimuli (Qin, Raye, Johnson & Mitchel, 2001). Using a computational model of the hippocampus, it is shown that stimulus overlap effects, similar to those observed in the first study, may account for these empirical findings. The third study delves more deeply into the underlying neural mechanisms described by computational models, focusing on the retrieval dynamics predicted by the complementary learning systems (CLS) approach (McClelland, McNaughton & O’Reilly, 1995). Using this model, simulations of the hippocampus and medial temporal lobe cortex are presented, the results of which predict that recollection and familiarity should respond in markedly different ways to variations in the similarity between studied items and their eventual retrieval cues. Specifically, the model predicts very steep, nonlinear transitions for recollection and more gradual, linear transitions for familiarity. Two behavioral experiments were conducted to test the model’s predictions and extend a prominent cognitive model (Yonelinas, 1994). The fourth study takes the same computational model of the hippocampus and memory into new territory by investigating whether it is able to account for recent findings implicating the network in complex scene perception tasks (Aly et al., 2013; Graham, Barense & Lee, 2010; Lee et al., 2005; Lee et al., 2012; Olsen, et al., 2012; Warren et al., 2012).

The following introductory chapter begins with a review of the theoretical framework upon which the current research is based, including an overview of relevant research methods, literature and ongoing controversies. More detailed reviews of the theory and literature can be found within the research papers. Following the background section, the papers included for this dissertation are
introduced in turn, highlighting the specific aims and hypotheses of each study, and my specific contributions to the work, as well as the contributions of any additional authors. The papers are then presented as introduced. Within the papers are thorough descriptions of the different methodologies and experimental designs used, the techniques of analysis, results, and critical discussions. The dissertation is concluded with a discussion chapter that begins with the most valuable insights extricated from each paper. The research is then critically examined as a cohesive and integral body of work. This is followed by a consideration of the limitations of the current research and suggestions for future research in the field.

### 1.2 Background Theory and Literature

#### 1.2.1 Recollection and Familiarity

The conceptual distinction between familiarity and recollection is often highlighted with the example of encountering a man in the street. At first you might recognize the man as someone you know, but at the same time, you are unable to identify who he is or where or when you encountered him. This global, non-specific recognition is the common experience of familiarity. Following further inspection, you might suddenly realize that the man is someone you met at a party last week and his name is Tom. Retrieval of these qualitative pieces of information that are of a specific place and time is referred to as the experience of recollection. Dual-process models (e.g., Jacoby, 1991; Yonelinas, 1997; Yonelinas et al., 2002) assert that these two kinds of experience reflect fundamentally distinct, and largely independent processes. Specifically, familiarity is postulated to be a continuous, strength-based retrieval process. In contrast, recollection is postulated to be a discrete, thresholded process tied to the retrieval of qualitative information associated with a specific stimulus event. Critically, all stimuli invoke a familiarity signal whereas recollection-based retrieval can fail entirely (Yonelinas et al., 2010).

An opposing view is that both experiences can be accounted for with a single-process model (e.g., Donaldson et al., 1996; Squire & Zola, 1998; Wixted, 2007). So for example, recollection may reflect strong memories that are rich in episodic details whereas familiarity reflects weaker, less detailed memories. However, single-process models have been criticized for providing no principled
way of accounting for the behavioral or neuropsychological dissociations observed between recollection and familiarity (Yonelinas, 2002). Moreover, due to a wide spectrum of evidence, some of which is outlined here, this idea is slowly losing influence.

In support of the two processes distinction, numerous behavioral and neuroimaging studies have demonstrated dissociations in tasks that differentially examine familiarity and recollection (Arndt & Reder, 2002; Ratcliff, Sheu & Gronlund, 1992; Yonelinas & Jacoby, 1996). Memory-impaired patients, for instance, have shown severely disrupted performance on tasks that predominantly utilize recollection, whereas performance on more familiarity-relevant tasks was relatively unaffected (Aggleton et al., 2005; Shimamura & Squire, 1991). In a factor analysis of fifty memory-impaired patients, Hunkin et al. (2000) identified separate factors that independently loaded on recollection and recognition (i.e., predominantly familiarity) test scores, respectively. A corroborating view, previously developed by Aggleton and Brown (1999), asserts that recollection is critically dependent on the hippocampus along with connected structures of the fornix, mammillary bodies and anterior thalamus. Familiarity, however, depends on some entirely different regions, including the perirhinal cortex, the dorsomedial thalamus, and areas of the frontal neocortex (Yonelinas et al., 2005).

Further supporting this distinction, Mayes and Roberts (2001) have found that the hippocampus plays only a minor role in item recognition tasks. They argue that, although the hippocampus is important for recollection, it is not involved in making familiarity judgments, the main constituent of recognition performance. This view is also supported in an fMRI study by Eldridge et al. (2000), who showed that the hippocampus is activated for recollection-based responses but not for familiarity-based test responses. The test used by the researchers to examine recollection and familiarity was an introspective procedure that was developed by Tulving (1985). Tulving’s procedure entails testing subjects on a list of studied and non-studied words by having them respond with either ‘remember’, if they recalled the study event, or ‘know’, if they were familiar with the item but did not actually recall the study event. In a similar fMRI study, Henson et al. (1999) found that the left hippocampus was associated with greater activation for ‘remembered' old words
compared with correctly rejected new words. The left parietal cortex also showed greater activation for ‘remember’ responses. In contrast, accurate ‘know’ responses did not elicit hippocampal activation but showed increased activity in the right lateral and medial frontal cortex.

Other evidence for the independence of recollection and familiarity comes from electrophysiological studies of human brain activity during recognition memory performance. For example, ‘remember’ and ‘know’ responses using Tulving’s procedure have been found to correlate with different event related potentials (ERPs) (Duzel et al., 1997). That is, ‘know’ responses were associated with an early temporo-parietal positivity in the N400 range and a late fronto-central negativity, whereas ‘remember’ responses were associated with widespread late bi-frontal activation and a left parieto-temporal positivity. Similar observations have been reported by Donaldson and Rugg (1998), implicating an independent role for recollection in word pair memory tests. Participants studied a list of word pairs (e.g., ‘bike–radio’, ‘carrot–pencil’) and at test were required to judge whether pairs of words had previously appeared together (e.g., ‘bike–radio’) or were re-paired (e.g., ‘bike–pencil’). The ERPs elicited for original pairings, relative to changed pairings, showed sustained positive shifts over the left parieto-temporal and right frontal cortices. In another study by Curran (2000), participants first studied a list of plural and singular words (e.g., ‘cars’, ‘shoe’), and at test they were presented with words in their original plurality (‘cars’), words with reversed polarity (‘shoes’), and new words, and were instructed to respond ‘yes’ to original plurality words, and ‘no’ otherwise. ‘Yes’ responses to original plurality words—equated with high recollection—were most strongly associated with parietal amplitudes, whereas ‘yes’ responses to plurality reversed words and ‘no’ responses to new words—equated with old-new recognition and low recollection—were more strongly associated with a mid-frontal N400 signal. In summary, although ERP studies generally do not indicate specific regions as being independently associated with recollection or familiarity, findings of distinguishing temporal and topographical qualities suggest that the two processes rely at least partially on separate neural structures (for reviews see Curran, 2000; Curran, Tepe & Piatt, 2006; Friedman & Johnson, 2000; Mecklinger. 2000; Rugg & Yonelinas 2003).
1.2.2 Cognitive Models of Recognition Memory

Findings such as the aforementioned have lead researchers to formalize a number of cognitive models of recognition memory (e.g., DeCarlo, 2003; Jacoby, Yonelinas & Jennings, 1997; Kelley & Wixted, 2001; Sherman, Atri, Hasselmo, Stern, & Howard, 2003). Prominent among these is the dual process signal detection (DPSD) model (Yonelinas, 1994; 1997), the starting framework for much of the current research. An illustration of the theoretical strength distributions associated with each process, as they pertain to a list of studied and non-studied items, is presented in Figure 1. As the name suggests, one component of the DPSD model, familiarity, is characterized by signal detection theory. Signal detection is the process of discerning a target (i.e., the “signal”, such as a previously studied item) from irrelevant background information (e.g., non-studied lure items) based on the strength of the signal. The familiarity-based signals, depicted on the left side of the figure, are formally described in DPSD theory by two overlapping Gaussian distributions of equal variance. The dark curve represents familiarity for a set of target items and the light curve represents familiarity for a set of lures. On average, the studied items are associated with a higher strength value. Familiarity can therefore be quantified as a relative measure of discriminability by estimating the distance between the means of the two distributions as a ratio of the variance. This measure of discriminability is referred to as $d'$ ('d prime'; discussed further in the papers). It should be noted that other parametric and non-parametric approaches have also been used to quantify familiarity (see Macmillan and Creelman, 2005). Also, although the DPSD model assumes equal variances of target and lure item distributions, other models assert that many of the phenomena reported in the recognition memory literature could be accounted for by assuming distributions of unequal variance (see the first paper for a discussion).

The right illustration in Figure 1 shows theoretical strength distributions for the recollection component of the DPSD model. The dark and light lines again represent targets and lures, respectively. Unlike familiarity, recollection is a discrete, or state-based, process that can succeed or fail. That is, for some stimulus cues there is a critical threshold above which the relevant memory will be retrieved, whereas for other stimuli there will be no memory retrieval. The sharp peaks on the left
indicate recollection failure. Importantly, under ordinary circumstances, lures rarely produce a signal beyond a certain threshold, so any signal that exceeds the threshold is diagnostic of a target. According to DPSD theory, the threshold is exceeded when any qualitative information associated with the study event—such as when or where it occurred—is evoked. However, the DPSD model stops short of characterizing the strength distribution of items that exceed the threshold and makes no claims as to when recollection failure will occur (note, the third paper of this thesis deals in some detail with these questions). In addition, evidence that recollection does not always present a thresholded function is investigated in research papers 1 and 2.

![Figure 1](image)

**Figure 1.** Memory strength distributions for a signal detection familiarity process and a threshold recollection process. Left: Signal detection theory assumes overlapping Gaussian distributions for targets (dark line) and lures (light line) such that familiarity strength is measured as the distance between the means of the two distributions (d’). Right: Threshold theory does not specify the shape of the memory distributions for recollection, but importantly, some proportion of items is recollected (R) and thus exceed a threshold, and the remaining items fall below threshold (i.e., recollection fails).

**Item and source recognition memory tests.** In classic tests of *item recognition*, subjects are required to determine if an item was previously studied (a *target*) or if it is new (a *lure*). In *source recognition* tests, subjects are required to determine if items were studied in a specific source or context, such as in List 1 versus List 2 or spoken by Person A versus Person B. Despite the apparent similarity
of item and source recognition tests, numerous behavioral and neural studies have indicated that they can be disassociated and that these tests rely on partially distinct memory processes or types of memory (Johnson, Hashtroudi, & Lindsay, 1993; Yonelinas, 2002).

Dual process accounts of item and source dissociations (e.g., Jacoby, 1991; Yonelinas, 1997; Yonelinas et al., 2002) postulate that the two types of tests differ because item recognition relies more heavily on familiarity. That is, in tests of item recognition, while recollection of qualitative information can help identify previously studied items, respondents are also able to make use of quantitative strength information, such as familiarity or recency, in order to discriminate between targets and lures. However, in source tasks, where subjects have been exposed to all the test items equally, familiarity is expected to be less helpful (but see Yonelinas, 1997).

**Receiver operating characteristic (ROC) curves.** One of the documented differences between item and source recognition tasks is the shape of their respective receiver operating characteristics (ROCs). An ROC is a function that relates the proportion of correctly recognized target items (*hit rate*) to the proportion of incorrectly recognized lure items (*false alarm rate*) across variations in a *response criterion*, which is the level of certainty, or confidence, at which a recognition response is classified as affirmative. In item recognition tests, the hit rate is the probability of correctly identifying an old item as old given the current response criterion, and the false alarm rate is the probability of incorrectly identifying a new item as old. For source recognition, the hit rate is the probability of correctly identifying an item as having been paired with a particular source (e.g., the left side of a display, a red background color, male voice, etc.), and the false alarm rate is the probability of selecting an alternative, incorrect source (e.g., the right side of the display, a blue background, female voice, etc.).

A common way to quantify the shape of an ROC is to plot the data points in z-space (zROC, for a review, see Yonelinas and Parks, 2007). The hit and false alarm rates are converted to z-scores by calculating the inverse of the standard cumulative normal distribution assuming a mean of 0 and a standard deviation of 1. If the zROC is linear, the intercept can be used as a rough index of recognition
accuracy. A perfectly symmetrical ROC will have a z-space slope (z-slope) of 1.0, but an ROC that is asymmetric along the diagonal will deviate from 1.0, producing a z-slope that is less than 1.0 if the ROC is pushed up on the left side, and greater than 1.0 if pushed up on the right. In addition, if the underlying distribution of an ROC is not continuous, e.g., a threshold distribution, the zROC will become curved, or U-shaped (see Macmillan & Creelman, 2005).

An illustration of typical item and source recognition ROCs is presented as Figure 2. As the left ROC (A) illustrates, item recognition ROCs are typically curved in probability space. In addition, when plotted in z-space (i.e., as a zROC) the shape is typically linear (see Egan, 1975; Murdock, 1974; Yonelinas, 1994). In contrast, source ROCs (right ROC) are typically more linear in probability space, but they exhibit a pronounced U-shape when plotted in z-space (e.g., Arndt & Reder, 2001; Kelly & Wixted, 2001; Rotello, Macmillan, & Van Tassel, 2000; Slotnick, Klein, Dodson, & Shimamura, 2000; Yonelinas, 1997, 1999; but see Qin et al., 2001). In fact U-shaped zROCs are consistently observed in studies of source recognition, associative recognition (e.g., “Was A paired with B?”), as well as plurality-reversed recognition (e.g., “Was A presented as a singular or plural word?”). For example, a review of the ROC studies examining these tasks by Parks and Yonelinas (2007) showed that U-shaped zROCs were observed in 52 out of 59 conditions from 17 different studies (also see Yonelinas & Parks, 2007; see Migo et al., 2014, for discussion of issues relating to the measurement of source memory).

According to DPSD theory, the ROCs for item and source recognition should in general be quite different. That is, under conditions in which familiarity is very useful for discriminating between targets and lures—such as in an item recognition task, where some items have been studied and some have not—the ROC should be asymmetric and curved, and the zROC should appear relatively linear. However, under conditions in which familiarity is expected to be less useful and performance relies more on recollection—such as in tests of source memory, where all items have been studied and thus are expected to be approximately equal in familiarity—the ROC should be more linear and the zROC should appear more U-shaped. As described above, the empirical literature has strongly supported these predictions. Notably, there are other similar models that make comparable
predictions (e.g., the variable recollection dual process model, Sherman et al., 2003; the some-or-none model, Kelley & Wixted, 2001; and the mixture model, DeCarlo, 2002). Importantly, in all of these models there is assumed to be a threshold process that plays a critical role in source recognition tasks and thus produces source zROCs that are U-shaped.

The use of ROCs in recognition memory research is widespread, however, one must consider both its strengths and limitations when evaluating such data. For example, one particular advantages of ROC experiments is that individual response biases do not impact the shapes of ROC plots or the parameter estimates that are derived from them. Another useful quality of ROCs is that it is easy to directly contrast the fits of different models to a single dataset by using goodness-of-fit statistics (e.g., $\chi^2$ or $G^2$ values). Other indices, including Akaike’s information criterion (Akaike, 1974) and the Bayesian information criterion (Schwarz, 1978), provide relative measures of the goodness-of-fit of one model compared to another. These measures take into account the number of parameters estimated by the model, penalizing models with more parameters. However, a frequently encountered problem is that multiple competing models are found to be exceptionally good fits to the recorded data, such that these statistics do not always help us to resolve whether one model is better than another (for discussion of this issue, see Healy et al., 2005; Macho, 2004; Yonelinas, 1999).

Practical issues to be considered when conducting ROC experiments include the need to utilize a sufficient number of response categories, as well as responses per category, so as to accurately characterize the shape of the ROC. ROC studies commonly examine six levels of response confidence. Fewer categories can make it difficult to determine the shape of the ROC due to the distance between data points, while increasing beyond six categories provides quickly diminishing benefit (Nishisato & Torii, 1970). In addition, evidence suggests that having too few responses per category (less than 50) can lead to high variability in parameter estimates (Macmillan, Rotello & Miller, 2004).

Another limitation of ROCs is the range of performance levels that are practically useful. When piloting ROC experiments, a common difficulty is avoiding either ceiling or floor performances. The problem with results that approach floor is
that the ROC will necessarily approach a slope of 1.0 and a y-intercept of 0.0 (Ratcliff, Sheu & Gronlund, 1992). In contrast, very high performances will tend to kink sharply resulting in unstable parameter estimates (Macmillan et al., 2004). A related problem is extrapolating beyond the observed data points to where the ROC functions intercept with the x- and y-axes. Observed data can easily become truncated toward these extremes, leading to inaccurate estimates. Thus we must be cautious about drawing conclusions solely based on these extreme data points.

Aside from the above practical considerations, potential measurement artifacts should be considered. For example, model simulations have found that if an individual’s response criteria vary across an experimental session, or if some proportion of responses are random, the ROC can exhibit an exaggerated curve (Ratcliff, McKoon & Tindall, 1994). It has also been shown that such effects can interact with overall level of performance and the distribution or responses, to produce different kinds of distortions.

Figure 2. Illustrations of typical receiver operating characteristic (ROC) curves for an item recognition test (left), and a source recognition test (right). The bold dots represent ‘observed’ data points and the continuous lines passing through them represent best fitting functions based on the dual process signal detection (DPSD) model. A) For the item recognition ROC, the left vertical intercept indicates recollection for a proportion of old items ($R_o$) and recollection for new items ($R_n$) is
The level of familiarity-based discrimination for non-recollected items ($d'$) corresponds to the level of curvature. B) The source recognition ROC indicates recollection of the source (‘A’ or ‘B’; e.g., left or right side of the screen) of items, denoted by the intercepts $R_A$ and $R_B$, respectively. The absence of curvature indicates that familiarity was not diagnostic of source (i.e., $d'$ equals zero).

**Systematic manipulation of source ROCs.** An important prediction of the DPSD model is that the degree to which source memory ROCs are linear (and their zROCs U-shaped) can be systematically manipulated. For instance, if the items from two sources differ in overall strength then familiarity can support source memory discriminations. This was first demonstrated by Yonelinas (1999), who showed that if items from one list (source-A) are studied five days before items from another list (source-B) then successful source decisions can be supported by familiarity and the resulting zROCs become flatter.

Other studies have shown that, if items are encoded such that they are integrated or *unitized* with their sources, familiarity can also support source discriminations (Diana, Ranganath & Yonelinas, 2008; Diana, Yonelinas, & Ranganath, 2010; for related results see Staresina & Davachi, 2006). For example, if subjects encode the word ‘elephant’ as coming from a ‘red’ source by imagining a ‘sunburned elephant’, then a red elephant should be more familiar than a green elephant at test. However, if the item and source are studied together but not in a way that promotes unitization, such as in the sentence, “The elephant stood near the red ball,” then familiarity does not support source discrimination. Evidence that encoding conditions promoting unitization also lead to familiarity-based source discrimination is indicated by comparatively more curved ROCs and flatter zROCs.

A critical feature of these source ROC effects is that the manipulation of the shape of the ROC curve, in both cases, is attributable to the introduction or enhancement of familiarity as a source of discriminating information. This manifests as differences in the degree of curvature of the ROCs because, as discussed above, unlike recollection, familiarity is a continuous process that has overlapping, Gaussian distributions. Thus, there is no reason to suggest, from these studies, that recollection ever deviates from the thresholded form.
However, there is other evidence of curved source ROCs (and flattened zROCs) in which no particular manipulation was used to foster the use of familiarity as a basis for source memory decisions, and the resulting zROCs exhibited little if any evidence of a U-shape (e.g., Glanzer, Hilford, & Kim, 2004; Qin et al., 2001; Slotnick, et al., 2000). The most dramatic example of this is a study reported by Qin et al. (2001), who found almost perfectly linear zROCs in a test of source memory for spoken statements. Findings of flat zROCs in source memory, like those found in Qin et al.'s study, have suggested to some that the DPSD model and others like it are inconsistent with the data, and as indicating that recollection is a continuous process that is more in keeping with a pure signal detection model (e.g., Qin et al., 2001; Slotnick & Dodson, 2005; Wixted, 2007). Qin et al. argued, for example, that recollection was the driving force behind the source judgments and, thus, also was the process that produced the linear zROC. If true, such evidence would provide a challenge to the assumption that recollection is a threshold process.

In addition, a computational model of recognition memory was found to predict that recollection memory should become less thresholded and more continuous when stimuli are very similar to one another (Norman & O’Reilly, 2003; discussed further in following sections). Thus, the question of whether recollection can truly be non-threshold, and how that would occur at a neural level, is one of great interest. This problem is addressed at length in the first two papers presented for this dissertation.

**Limitations of cognitive models.** ROCs and cognitive models reveal a great deal about the quantitative characteristics of the processes involved in recognition memory but leave many questions unanswered. For example, as discussed above, there are experimental situations in which source memory performance demonstrates the presence of a familiarity-like signal, and in some instances it is unclear whether the signal is in fact familiarity or rather signifies that recollection is not always thresholded but can transition into a graded signal. Moreover, there is no *a priori* hypothesis as to why recollection might exhibit different characteristics or what neural mechanisms would give rise to this.

Another limitation is that cognitive models often do not make very specific predictions about important experimental variables, such as how recollection and
familiarity respond to subtle changes in the stimulus materials between study and test. That is, when the stimulus cue that prompts memory retrieval is varied in similarity to the originally encoded event, how do the different memory processes perform? This question is important when we consider that in natural settings, the kinds of stimuli that trigger memory retrieval rarely, if ever, match up perfectly with past experience. Certainly, some generalization is required, but this must be balanced against the need for accuracy and relevance so that we can function effectively in a world of innumerable items, features, and contexts. Previous research has explored the effects of stimulus similarity on recollection and familiarity in several interesting ways. For example, a study of hippocampal patient Y.R. found that recollection but not familiarity was impaired on a yes/no recognition test with similar foils, but performance was not impaired on a forced choice test that paired targets with similar foils (Holdstock et al., 2002). This suggests that familiarity can be useful when there is a relatively high level of similarity (Migo et al, 2009). Related research of individual differences in older adults corroborates this view (Migo et al., 2014; see also Bayley, Wixted, Hopkins and Squire, 2008). However, little is known about how the similarity gradient differs for recollection versus familiarity. This question was the motivation for the third paper presented in this thesis and is discussed further in the following computational sections.

Other complex questions are not easily tackled by cognitive models. For example, there is growing evidence that the brain regions involved in long-term recognition memory are also critically involved in other cognitive processes, including short-term memory, and complex scene perception (e.g., Aly, Ranganath, & Yonelinas, 2013; Graham et al., 2010; Lee et al., 2005; Lee et al., 2012; Olsen et al., 2009; Warren et al., 2012). Integrating these findings into existing models of recognition memory is a challenging task that benefits from additional methodological approaches such as computational modeling. This challenge forms the scope of the fourth dissertation paper.

In summary, cognitive models are very valuable tools for describing the characteristics of behaviorally measurable functions like recognition memory, but by themselves they provide very limited insight into the neural mechanisms that give rise to the processes they describe. As alluded to earlier, imaging techniques,
Electrophysiological recordings, and single cell recordings, are making important progress as they identify the specific brain regions involved during these processes, the functional pathways of communication, and numerous other neural characteristics. However, bridging the gap between brain and behavior requires an understanding of how different scales—from neural firing patterns to network-level interactions—interrelate to ultimately give rise to complex psychological processes such as recognition memory. As the following section details, neurocomputational models provide an invaluable tool for testing different hypotheses that relate brain functioning at these different scales to cognition and behavior. Moreover, as demonstrated in the current research, it is an extremely practical companion to cognitive modeling and behavioral experimentation, contributing to a more comprehensive view of the dynamical aspects of recognition memory.

### 1.2.3 Neurocomputational Properties of Learning

In recent years, neurocomputational models have been used to simulate the neural network dynamics of the hippocampus and related memory structures with the goal of identifying and understanding the core mechanisms of long-term memory encoding and retrieval (e.g., Alvarez & Squire, 1994; Hasselmo, et al., 2002; Hasselmo & Eichenbaum, 2005; Hasselmo & Wyble, 1997; McClelland, et al., 1995; McNaughton & Morris, 1987; McNaughton & Nadel, 1990; Rolls, 1996; Rolls & Treves, 1994; Rudy & O’Reilly, 1999). As the current thesis papers demonstrate, computational models are a useful counterpart to cognitive models, helping to validate existing theories and, critically, generate novel predictions of how cognitive processes, such as memory and perception, naturally emerge as a result of distinct patterns of neural connectivity and activation in specified regions.

The models presented in the current research utilized an instantiation of leabra (local, error-driven and associative, biologically realistic algorithm), an algorithm developed by O’Reilly (1996; see also O’Reilly & Munakata, 2000), and implemented in Emergent (Aisa, Mingus & O’Reilly, 2008; formerly PDP++). The algorithm is a coarse mathematical approximation of the biological processes thought to underlie neuronal activation and learning. Technically speaking, the algorithm includes a point-neuron approximation with a discrete spiking function...
and a continuous rate-code output function that simulates groups of neurons, a Hebbian (associative) learning function, an error-driven learning function, and a competitive inhibition function that can be tuned to produce distributed representations with varying levels of sparseness. The following sections provide a cursory outline of the core principles that this algorithm encapsulates and that are specifically relevant to the computational modeling sections of the papers presented for this thesis. For a more in-depth look at the mathematical formulae implemented in the algorithm and their derivations, see the abovementioned references.

**Associative learning.** Current models of the neural basis of associative learning are founded on the principles developed by Hebb (1949), beginning with the concept of the *cell assembly*, which describes a network of neurons that is consistently activated, or repeats the same sequence of firing, during a given cognitive event, such as viewing a particular image. Hebb hypothesized that during this co-activation of neurons, their sensitivity to one another is strengthened and in this way, future occurrences of the stimulus are more readily identified.

Models of associative learning and memory build on these concepts, formalizing, in one form or another, the Hebbian processes of *long-term potentiation* (*LTP*), which is a persistent increase in the weight between two neurons (i.e., the sensitivity of a receiving neuron to the sending neuron), and *long-term depression* (*LTD*), which is a reduction in weight. In the original formulation of leabra, this was accomplished using a *conditional principal components analysis* (*CPCA*) learning rule (adapted from Oja, 1982), described below and used in the current models. It should be noted that numerous Hebbian learning rules have been proposed since Hebb’s original postulation (e.g., Bienenstock, Cooper & Munro, 1982; Grossberg, 1976; Kohonen, 1988; Sejnowski & Tesauro, 1989); which rule is the most biologically plausible is a subject of continuing research and debate. In a relatively recent revision of the leabra algorithm, a markedly different Hebbian rule was formulated to account for recent findings regarding the biological processes involved in synaptic modification. Importantly though, at the broader network level (the level of analysis we are interested in), the two rules can be expected to produce very similar results (Aisa, et al., 2008). Thus, only a cursory account of neural
activation and Hebbian learning rules—as they were used in the current simulations—are presented here so as to prime the reader for the research papers.

Rate-code approximation. The simulation of action potentials is approximated in leabra using an array of biological constraints, including excitatory and inhibitory ion channels, leak currents, noise, and adaptation, among others. Critically, the result of all these parameters is a mathematical function that approximates a very simple idea; that incoming activation (i.e., excitatory and inhibitory postsynaptic potentials) is summed and if over a given period of time the net input provides sufficient excitation (i.e., raises the membrane potential to threshold), the neuron will fire. The summing process can be summarized in a simplified, but intuitively useful, form by the following formula:

$$ V_j = \frac{1}{n} \sum_i w_{ij} x_i $$ (1)

Here, $V_j$ is the membrane voltage potential of a receiving neuron. The impact of each sending neuron is the product of its activation ($x_i$) and weight ($w_{ij}$). The values computed for each sending neuron are summed and standardized by dividing by the total number of neurons ($n$) that synapse with the receiving unit. The final step in this part of the algorithm is rate-code approximation. When the threshold membrane potential is exceeded, activation (i.e., an action potential) will occur at the receiving neuron. Although individual neurons have discrete firing states, small populations of neurons (e.g., 100 or so) can be more efficiently represented with a continuous activation value between 0-1. That is, over a given time interval, we might expect some percentage of neurons in the population to fire. With this approach we can communicate rate values instantly rather than communicating spikes over a period of time, which results in less noise and greater computational efficiency. In the models presented, one of these small neuron populations is referred to as a unit.

Weight updating. Utilizing the above activation rule, the CPCA learning rule can be conceptualized by imagining two equally sized layers of units, an input layer ($x$) and a hidden layer ($y$), where each input unit projects to many hidden units;
therefore, hidden units receive projections from multiple inputs. The weight between a pair of input and hidden units ($w_{ij}$) can be quantified as a value between 0-1 that reflects the probability that the sending unit ($x_i$) is active when the receiving unit ($y_j$) is active. CPCA says that a particular environmental structure is represented by a subset of units in the input layer. Therefore, the level of activation in the hidden layer reflects the probability that a hypothesized structure is actually present in the environment. The CPCA rule achieves this relationship by updating the weight values, and can be formalized—in simplified form—as follows:

$$\Delta w_{ij} = \varepsilon y_j (x_i - w_{ij})$$

(2)

Here the weight, $w_{ij}$, is updated to reflect the activation of the sending unit, $x_i$. If the firing rate of $x_i$ is greater than what has typically been expressed in the past, as indicated by $w_{ij}$, then the weight is adjusted upward (i.e., LTP). Alternatively, if $x_i$ is less than expected, the weight will be adjusted downward (i.e., LTD). Finally, the rate of update is proportional to both the firing rate of the receiving unit, $y_j$, and a learning rate, $\varepsilon$. As explicated in later sections, setting the learning rate too high—regardless of whether this or another learning rule is used—can be detrimental; an optimal learning rate depends on the goal of the memory system (e.g., to record episodes or to extract generalized representations) and its architecture, and thus likely differs across brain regions.

Several critical factors must be addressed by any algorithm that aims to approximate learning dynamics in a biologically realistic way. One natural constraint is the lower and upper bounds within which a weight can change. First, the lower limit cannot fall below zero; this would be equivalent to an excitatory neuron becoming inhibitory, or vice-versa. Second, there is a natural upper limit to weight values. Because the weight values in CPCA reflect a probability, a weight bounding of 0-1 occurs as a matter of principle. Additional factors such as weight change scaling, soft bounding, offset, decay, and learning thresholds have been investigated and discussed in depth and so are not covered here (see O’Reilly & Munakata, 2000).
Other forms of memory and learning. Hebbian, or weight-based memory can be differentiated from activation-based (or online) memory in which a memory is represented by a cell assembly for as long as it is active, or it is primed so as to be easily activated (e.g., Tulving, Schacter & Stark, 1982). In contrast, Hebbian learning creates offline memories because the effect of stimulus exposure can persist long after the activation pattern has ceased—i.e., there are lasting weight changes.

Associative learning can also be differentiated from other kinds of hypothetical offline learning mechanisms or algorithms, such as error-driven learning, in which synaptic weights are adjusted based on the computed difference between two activation states—for example, representing expected and observed outcomes in a goal-directed behavior—and temporal delay learning, in which a pattern of activation or behavior leads to a reward or punishment that occurs at a later point in time (see O’Reilly and Munakata, 2000, for details on the implementation of these mechanisms via leabra). Such mechanisms were not utilized in the current research and are therefore not discussed further.

Inhibitory competition and sparse representations. If excitatory activity in the brain were allowed to occur unhampered, then due to dense reciprocal connectivity throughout the neocortex (Suzuki & Amaral, 1994), activation would quickly cascade out of control. Thus, a core component of many neurocomputational models—and a critical feature of the models used in the current research—is the modulation of excitatory activity through inhibitory forces. A popular concept is the idea of a perpetual ‘tug-of-war’ between excitatory and inhibitory neurons synapsing in the same space, known as inhibitory competition, which results in stable, sparse patterns of activation (Buzsáki, 1984; O’Reilly & Munakata, 2000). Inhibitory neurons typically synapse with nearby excitatory and inhibitory neurons (interneurons). Projections to inhibitory interneurons are classified as feedforward, when originating from a preceding region, and feedback, when originating from the same region. Feedforward inhibition effectively anticipates excitatory activity in a region, whereas feedback inhibition reacts to it. Thus, the role of inhibitory interneurons can be seen as homeostatic, regulating the activity of excitatory pyramidal cells (Bernander, Douglas & Martin, 1991). When there is a lot of
excitatory activity, inhibition will also rise, and conversely, when excitation is low, inhibition will also fall.

In memory systems, inhibition is thought to give rise to a ‘competitive’ learning environment because only the strongest excitatory neurons overcome inhibitory forces to produce action potentials and therefore represent a given stimulus. This process is simplified in leabra using a $k$-winners-take-all ($kWTA$) rule. Like the activation rule described earlier, for computational efficiency $kWTA$ adopts some assumptions about the behaviors of groups of neurons, in this case an entire layer of neurons. Specifically, when there are steady excitatory forces, activation in the receiving layer will eventually stabilize to produce a reliable pattern in which some neurons (or units) are highly active and others are less active. Greater inhibition will result in fewer units becoming highly active. Thus, using the $kWTA$ rule, inhibition can be controlled by varying $k$, which is the number of units that are allowed to be active beyond a certain level (e.g., .25). As explained later, the brain’s ability to express different levels of inhibition in different regions provides it with great power and flexibility when processing and storing events or stimuli.

1.2.4 The Medial Temporal Lobe and Episodic Memory

As touched on earlier, there is wide ranging evidence implicating the involvement of the medial temporal lobe—and in particular the hippocampus—in long-term, episodic memory. Neuropsychological studies indicate that hippocampal damage leads to profound long-term memory impairments and relatively spared perception, working memory, implicit memory, and other cognitive functions (Baddeley & Warrington, 1970; Cohen & Eichenbaum, 1993; Graf & Schacter, 1984; Reed & Squire, 1998; Scoville & Milner, 1957; Zola-Morgan, Squire & Amaral, 1986). Neuroimaging and electrophysiological studies have converged with patient findings by showing a role for the hippocampus in encoding and retrieving long-term memory (Axmacher et al., 2010; Fell et al., 2011; Montaldi et al., 2006; Ranganath et al., 2004; Sederberg et al., 2006; Yonelinas et al., 2005).

A number of computational models have been proposed to explain how the hippocampus is able to encode, maintain, and retrieve memories (McNaughton & Morris, 1987; McNaughton and Nadel, 1990; McClelland, et al., 1995; Rudy &
The model used in the current research is based on widely accepted knowledge of hippocampal anatomy and activation levels. It was instantiated using several components of the leabra algorithm (described above), and closely follows the model used by Norman and O’Reilly (2003). This section begins with a brief overview of hippocampal anatomy, including pathways of connectivity and the roles of different subregions. A discussion of two critical computational mechanisms, often referred to as pattern separation and pattern completion, follows. This leads into an introduction of the complementary learning systems (CLS) model, a prominent theoretical framework that is core to the current research. The CLS model postulates separate hippocampal and cortical memory networks that serve contrasting, yet equally important roles and, moreover, this separation of functions may be needed to provide a satisfactory neurocomputational account of the cognitive functions of recollection and familiarity.

**Anatomy of the hippocampus.** A schematic of the hippocampus and its main neural pathways is presented as Figure 3. The main gateway in and out of the hippocampus proper is the entorhinal cortex (ERc). The ERc receives convergent information from the perirhinal cortex and parahippocampal cortex. Together with the hippocampus, these areas sit atop a hierarchy of interconnected cortical areas, beginning with low-level sensory areas (e.g., the primary visual cortex) and proceeding up through high-level association areas (Lavenex & Amaral, 2000). Additional convergence comes from the prefrontal cortex (Simons & Spiers, 2003). Thus, the hippocampus is perfectly situated to encode associations between the many elements of one’s current cognitive state.

The hippocampus proper comprises the dentate gyrus (DG), and areas CA3 and CA1. The hippocampus is somewhat unique in that the flow of information is unidirectional, with connectivity between subregions forming a loop that begins and ends with the entorhinal cortex. More specifically, large pyramidal axons originating in the superficial layers of the ERc (‘input’) project to granule DG cells and to pyramidal cells making up area CA3 (the perforant path) and area CA1. Area CA3 is also innervated by the DG granule cells (the mossy fiber path), which have a
particularly powerful depolarizing effect on the CA3 neurons. Synaptic connections exist among CA3 neurons (*recurrent collaterals*), as well as from CA3 to CA1 (*Schaffer collaterals*). Finally, CA1 neurons project to the deep layers of the ERc (‘output’), directly and indirectly via the subiculum, completing the loop (note, the subiculum is not included in the schematic as it was not part of the computational model used).

In addition to having powerful projections onto CA3, the DG is unusual in that only about 1% of neurons are active in response to a stimulus. This can be contrasted with the cortex, which is typical around 25%. Area CA3 is also relatively sparse, with around 4% activation, while CA1 is somewhat less sparse with approximately 10% activation (note, these were the values used in the model simulations, set using the kWTA rule discussed earlier). As the following sections explain, these activation values are critical to the functioning of the hippocampus.
**Figure 3.** Schematic of the hippocampal network used in the current computational models. The red arrows indicate projections that are particularly significant during memory encoding, and the blue arrow during retrieval. The purple arrows indicate projections that are involved in both encoding and retrieval, and that perform critical learning via Hebbian weight updates. At encoding, activity spreads from the entorhinal cortex to area CA3, both directly and indirectly via the sparsely active dentate gyrus (producing pattern separation), as well as to area CA1. Hebbian weight changes occur most critically at the recurrent collaterals within CA3 and the Schaffer collaterals from CA3 to CA1.

**Encoding and retrieval in the hippocampus.** The encoding of information in the hippocampus—as it was simulated in the current research (originally from O’Reilly & Munakata, 2000, and Norman & O’Reilly, 2003)—proceeds as follows. First, multimodal information converging at the ERc input layer is construed as a series of independent, non-overlapping *features* (or vectors). Activation spreads to area CA3 both directly and indirectly via the DG. The extreme sparseness of activity in the DG results in highly distinct activation patterns in CA3. Connections between active CA3 neurons (recurrent path) are also strengthened. Thus CA3 helps bind together the disparate features of an event.

Concurrently, the ERc input drives activation in area CA1. Area CA1 is able to map the original ERc input pattern back to the ERc output to produce an identical pattern. That is, it performs *invertible mapping* (in computer science this is equivalent to an *auto-encoder*). Critically, the associations between active CA3 and CA1 neurons are also strengthened so that the original mapping can be reinstated.

At retrieval, recurrent CA3 connections can be helpful if the cue is incomplete or subtly different to the original stimulus, as activation will spread along its strengthened pathways to reproduce the original pattern. Another important feature of retrieval in the model is that the projections from ERc input to CA1 are weakened, or simply turned off at this time. This allows any stored pattern that is completed or reinstated at CA3 to be the dominant signal to CA1.

**Pattern separation and pattern completion.** First proposed by David Marr (1971), the computational mechanisms of *pattern separation* and *pattern completion*
are a critical feature of many memory models. Beginning with the idea that the brain represents the outside world using stable and reproducible patterns of neural activity, pattern separation is the propensity for different brain regions to invoke patterns of activity that are statistically more, or less, distinctive than the stimuli themselves. Pattern completion is the process by which these patterns are later reinstated, commonly associated with memory retrieval. As discussed below, although the two mechanisms are complementary and neither is specific to any single brain region, these mechanisms are also at odds with one another in some crucial respects.

With respect to the hippocampus, pattern separation and completion have been the subject of much computational modeling research (McNaughton & Morris, 1987; O’Reilly & McClelland, 1994; Rolls & Treves, 1994; Rolls, 1996) and experiments using behavioral (Kesner, Lee & Gilbert, 2004; Kirwan, Ashby & Nash, 2014; Rolls & Kesner, 2006), neurophysiological (Leutgeb & Leutgeb, 2007; Leutgeb et al., 2007), and fMRI techniques (Bakker et al., 2008; Kirwan & Stark, 2007; Lacy et al., 2011). In the hippocampal model presented above, pattern separation occurs due to the extremely sparse firing patterns of the DG/CA3, itself the result of relatively strong inhibitory competition. One reason sparse activation leads to greater pattern separation is a matter of statistics. For instance, in the DG where the probability of a neuron firing in response to a given stimulus is 1%, the probability of that same neuron being active for another random stimulus is therefore .01% (i.e., 1% × 1%). By comparison, in a more densely active region such as the cortex, if the probability of activation were 20% there would be a 4% chance of the same neuron being activated for two random events.

Another reason that sparse activation can lead to reduced pattern overlap stems from the presence of very broad and diffuse connectivity between regions and the fact that strong inhibitory competition tends to favor highly conjunctive representations (O’Reilly & McClelland, 1994; O’Reilly & Rudy, 2001). That is, neurons that receive input from many active neurons are more likely to overcome inhibition (i.e., ‘win’) than neurons with fewer active inputs. Because the DG environment is so competitive, only a small variation in the input pattern often results in a dramatic change in the excitatory neurons that win and lose. Thus,
similar input patterns can end up having entirely separate representations. Finally, due to the very powerful ‘detonator synapses’ (McNaughton & Morris, 1987), from DG neurons onto CA3, DG activation will largely determine which CA3 neurons will partake in representing a given event.

So why is pattern separation important in the hippocampus? If the job of the hippocampus is to bind together the various unique associations that make up an event, it must do so quickly (e.g., to encode where you put your coat at a party you must quickly associate the coat with some location). In more typical, densely active networks, a single neuron might be involved in representing many different things. The problem then with rapid learning is that a new representation can render previously learned representations irretrievable—referred to as **catastrophic interference**. Pattern separation helps to overcome this problem by allowing the network to learn patterns quickly without impacting previously encoded representations, even when there is a high level of feature similarity. Determining just how similar items can be without causing catastrophic interference is the subject of the first paper of this thesis.

The job of pattern separation during encoding can be contrasted with that of pattern completion during retrieval. Pattern completion is the reactivation of a stored representation, often from a partial or noisy cue. The recurrent connectivity within CA3 assists in this process as the strengthened connections give rise to **attractor dynamics**. That is, when an input pattern is sufficiently similar to the original pattern, activation will usually inexorably spread to the remaining neurons in the CA3 representation. The final state of activation represents a local energy **minimum** (e.g., analogous to a ball rolling to the bottom of a hill). A consequence of this dynamic is that stored representations will tend to be fully reinstated or else not at all—the subject of the third paper.

As illustrated in Figure 4, there is a natural tradeoff between pattern separation and completion. If pattern separation is too effective, it becomes impossible to generalize from a novel stimulus to a previously encoded one. Consequently, any stimulus, regardless of how similar to its previous state would seem new (e.g., imagine not recognizing your mother because she’d had a haircut). Conversely, if pattern completion is too good then the problem becomes one of
overgeneralization. Consequently, old memories would often be reactivated instead of encoding new representations for novel episodes.

Figure 4. A conceptual representation of pattern separation and pattern completion. The input stimuli (‘A’ and ‘B’) in both cases are partially overlapping, indicating that they bear some similarity. When pattern separation is dominant (left), the representations of those items (i.e., at ‘output’) diverge to different pools of neurons. Thus, they are encoded as separate entities. This may be useful, for instance, when small discrepancies between events or items are diagnostic of important differences (e.g., telling your umbrella from someone else’s). Pattern completion (right) allows a previously encoded pattern (‘A’) to be retrieved when given a new pattern (‘B’), despite some feature differences. One use of this would be to generalize from a novel stimulus to a familiar concept so as to infer its meaning.

**Complementary learning systems.** The complementary learning systems (CLS) model (McClelland, et al., 1995; O’Reilly & Rudy, 2001) posits that some tasks benefit from enhanced pattern separation, and others from enhanced pattern completion, and so to adequately address these varying demands, separate, functionally specialized brain regions are needed; namely the hippocampus, which is optimized for greater pattern separation and the neocortex (referring broadly to the medial temporal lobe cortex and association cortex) which is optimized for pattern completion.

The CLS model treats the neocortex as a simple two-layer network comprising: the lower-level cortex (which for simplicity is modeled as identical to
the hippocampal ERC input), and the higher-level association cortex (or medial temporal lobe cortex). The association cortex is densely activated relative to the hippocampus proper, thereby producing highly overlapping representations. The same kind of Hebbian learning found in the hippocampus is used to encode memories in the neocortex, with one critical difference; learning must be gradual enough that new information can be integrated with previously learned patterns rather than simply erasing them (i.e., catastrophic interference). In doing so, slowly, over time the neocortex is thought to represent the statistical regularities of one’s experiences.

A schematic of both the hippocampal and neocortical components of the CLS model is presented as Figure 5. The figure illustrates how the two networks deal with the encoding of multiple, similar episodes involving scenes with houses. Despite the perceptual and conceptual similarity of the episodes, the hippocampus is still able to create distinct, pattern-separated representations. This might be important if, for example, you were shopping for a house and wanted to compare your impressions of each. Combining the two representations would only create confusion. In contrast, the very purpose of the neocortex is to incorporate new experiences with old. For example, after being in the house market for some time you would likely develop a richer concept of what a house is. As this happens, features that are common to many episodes will be strengthened, via Hebbian LTP, whereas one-off features will slowly fade, via LTD. In fact, the simple ability to recognize a new house as being “a house” exemplifies the role of the neocortex as a system that extracts meaning.
Figure 5. Schematic of the hippocampus and neocortex networks based on the complementary learning systems model and applied to learning of a pair of house scenes. The illustration shows representations of a ‘current episode’ (associated neurons are indicated by blue circles) and a ‘previous item’ (red circles). The patterns in the entorhinal layers of the hippocampus and the lower-level layer of the neocortex can be thought of as a collection of episode features (e.g., colors, shapes, configurations, etc.). Sparse activation in the hippocampus, especially the dentate gyrus, produces highly pattern-separated representations in CA3. Two similar scenes are unlikely to be confused due to the minor differences that do exist. In contrast, the association cortex is more densely activated relative, thereby producing highly overlapping representations and enhancing pattern completion (e.g., the ability to generalize from a novel house to the concept of house).

The CLS model and recognition memory. The CLS model is theoretically aligned with dual process cognitive memory models and in particular, shares many key characteristics with the DPSD model of recognition. Evidence for this was reported by Norman and O’Reilly (2003) with simulation data showing that the CLS model can account for a variety of recognition memory phenomena (e.g., list-length
and list-strength effects, Yonelinas, 2002). Moreover, the test score distributions for old and new items that were associated with the hippocampus and neocortex are theoretically and empirically consistent with the sorts of distributions ascribed to recollection and familiarity, respectively. For example, it was shown that the hippocampal network typically produces a bimodal distribution for old items (i.e., previously learned patterns) at the time of test, in which some items were associated with a high retrieval score—indicated by the number of matching versus mismatching features found when comparing the input and output layers of the ERc—and other items were associated with very low scores indicating retrieval failure; that is, no output pattern was produced. New items, however, were never reproduced at the output layer. Thus, the retrieval of any matching features was highly diagnostic that the current item was old. This is consistent with the DPSD model’s characterization of recollection, which is defined as a threshold retrieval process (Yonelinas, 1994).

Additionally, the distributions associated with the neocortical network corresponded with the signal detection, familiarity component of the DPSD model. That is, when familiarity was measured in the neocortex as the level of sharpness, or contrast, of the activated pattern, the model produced continuous, overlapping Gaussian distributions where old items were associated, on average, with sharper patterns of activation than new items. That is, due to the Hebbian mechanisms of LTP and LTD, when an item is presented to the network for learning, connections between coactive units are strengthened while less active connections are weakened. Also, because the measure is a scalar—that is, it is not associated with the retrieval of specific features—it is consistent characterizations of familiarity as a global match (Norman & O’Reilly, 2003; Clark & Gronlund, 1996).

**Tackling new challenges.** Computational models have the benefit that one can study each component in isolation. In human behavioral experiments, this is impossible, if not in principle then at least in practice. As such, disentangling the processes of recollection and familiarity is always challenging, and to say the least, contentious. While the CLS model offers a compelling account of how the hippocampus and neocortex might contribute to long-term memory, empirical validation is needed. A core aim of the current thesis is to tie the concepts and
predictions of the CLS model to the cognitive constructs of recollection and familiarity by combining the methodologies of computational model simulations and human behavioral experimentation.

As discussed earlier, a limitation of existing dual process models is that they cannot by themselves provide a satisfactory account for why, under some circumstances, tasks that are thought to rely heavily on one process (i.e., recollection), appear to show contradictory characteristics. The CLS model offers an explanation of why this might happen under at least one set of circumstances. Specifically, the simulations by Norman and O’Reilly (2003) made the novel prediction that when the stimuli in a study list are highly related—that is, they share many of the same features—pattern separation in the hippocampus will begin to break down, translating to a collapse of the retrieval threshold and forcing recollection to appear continuous. The first paper in this thesis sets out to test this prediction using a source memory paradigm that is applied first to the CLS model and then in a series of behavioral experiments.

In the second paper, an unexpectedly related problem was investigated; namely, the finding (discussed earlier) that complex stimuli appear to produce more curved ROCs. Evidence from behavioral source recognition experiments, including data from a hippocampal lesion patient, is presented that sheds light on what aspects of recognition memory are affected by differences in complexity. In addition, data from novel hippocampal model simulations provides a compelling account of why memory performance is affected.

Another challenge, tackled in the third paper, is to identify the distinctive retrieval dynamics that are differentially associated with recollection and familiarity. The trade-offs that are core to the CLS model, such as pattern separation and the ability to generalize from novel stimuli, lead to very specific predictions about how the hippocampus and neocortex should each respond to varying levels of study-test item similarity and the gradients that are produced as that similarity is manipulated. For example, a consequence of sparse representations in the hippocampus is that a high degree of similarity is required in order for the hippocampus to pattern complete, whereas the cortex should be able to perform at lower levels of similarity, albeit with less diagnostic accuracy. Thus, a critical test of the model and its
correspondence with recognition memory models is an examination of recollection and familiarity similarity gradients.

1.2.5 Memory and Perception: An Emerging Perspective

As outlined in the thesis overview, the critical involvement of the hippocampus in long-term memory is well established. Moreover, while patients with hippocampal damage have shown profound memory impairment, those same patients have demonstrated relatively normal performance on neuropsychological tests of perception, skill learning, and other cognitive functions (Eichenbaum & Cohen, 2001; Reed & Squire, 1998; Scoville & Milner, 1957; Zola-Morgan et al., 1986). However, recent findings of impaired high-level scene perception with selective hippocampal damage (Graham et al., 2010; Lee et al., 2005; Lee et al., 2012; Warren et al., 2012) and increased hippocampal activation in healthy adults during the performance of challenging scene discrimination tasks (Barense et al., 2010; Lee & Rudebeck, 2010; Mundy et al., 2012) has challenged the predominant view and lead some researchers to hypothesize that the hippocampus likely serves a broader role than previously thought. More recently, Aly, et al. (2013) reported graded patterns of activation in the hippocampus that strongly correlate with scene discrimination performance.

A critical question that has yet to be answered, is whether existing computational models that have been used to explain memory functioning in the hippocampus can also account for the evidence suggesting that hippocampus is involved in making perceptual judgments. Thus, expanding beyond the domain of long-term memory, the fourth and final paper in the current thesis attempts to address this question.
Chapter 2: Testing a Neurocomputational Model of Recollection, Familiarity, and Source Recognition

**Authors:** Kane W. Elfman, Dr Colleen M. Parks, Dr Andrew P. Yonelinas (PI)

**Journal:** Journal of Experimental Psychology: Learning, Memory & Cognition

**Date of publication:** July, 2008

**Abstract:** The authors assess whether the complementary learning systems model of the medial temporal lobes (Norman & O’Reilly, 2003) is able to account for source recognition receiver operating characteristics (ROCs). The model assumes that recognition reflects the contribution of a hippocampally mediated recollection process and a cortically mediated familiarity process. The hippocampal process is found to produce threshold output functions that lead to U-shaped zROCs, whereas the cortical process produces Gaussian signal detection functions and linear zROCs. The model is consistent with several dual process theories of recognition and is capable of producing the types of zROCs observed in studies of item and source recognition. In addition, the model makes the novel prediction that as the level of feature similarity across items increases, the ability of the hippocampus to encode distinct representations for each stimulus will diminish, and the threshold nature of recollection will break down, leading source zROCs to become more linear. The authors conducted 3 new behavioral source experiments that confirmed the model’s prediction. The results demonstrate that the model provides a viable account of item and source recognition performance.

**Author contributions:** Dr Colleen Parks primarily assisted in data collection by preparing many of the test materials, experiment design, and running test sessions. Dr Andrew Yonelinas served as the principal investigator, overseeing the research project. My personal contribution began with the conception of the study, which was inspired by the work of Norman and O’Reilly (2003) who showed how the CLS model predicted novel behavioral effects due to experimental manipulations of stimulus feature overlap. My contribution included the design of behavioral experiments and subsequent analyses of the collected data, including plotting ROCs and deriving parameter estimates from model fitting and multivariate analysis. In addition, I designed and ran the CLS model simulations from which the behavioral
predictions were generated and later compared. I authored the research paper’s entire initial draft, which was then jointly revised by myself, Dr Parks, and Dr Yonelinas.
Testing a Neurocomputational Model of Recollection, Familiarity, and Source Recognition

Kane W. Elfman, Colleen M. Parks, and Andrew P. Yonelinas
University of California, Davis

The authors assess whether the complementary learning systems model of the medial temporal lobes (Norman & O’Reilly, 2003) is able to account for source recognition receiver operating characteristics (ROCs). The model assumes that recognition reflects the contribution of a hippocampally mediated recollection process and a cortically mediated familiarity process. The hippocampal process is found to produce threshold output functions that lead to U-shaped zROCs, whereas the cortical process produces Gaussian signal detection functions and linear zROCs. The model is consistent with several dual process theories of recognition and is capable of producing the types of zROCs observed in studies of item and source recognition. In addition, the model makes the novel prediction that as the level of feature similarity across items increases, the ability of the hippocampus to encode distinct representations for each stimulus will diminish, and the threshold nature of recollection will break down, leading source zROCs to become more linear. The authors conducted 3 new behavioral source experiments that confirmed the model’s prediction. The results demonstrate that the model provides a viable account of item and source recognition performance.

Keywords: hippocampus, receiver operating characteristic, medial temporal lobe, source memory

In item recognition tests, subjects are required to determine if items were studied previously or if they are new to the experiment. In contrast, in source recognition tests, subjects are required to determine if items were studied in a specific source or context, such as in List 1 versus List 2 or spoken by Person A versus Person B. Despite the apparent similarity of item and source recognition tests, numerous behavioral and neural studies have indicated that they can be dissociated and that these tests rely on partially distinct memory processes or types of memory (e.g., Johnson, Hashtroudi, & Lindsay, 1993).

One theoretical account of these dissociations (e.g., Jacoby, 1991; Yonelinas, 1997; Yonelinas et al., 2002) is based on dual process theories of recognition that postulate that recognition memory reflects the contribution of a recollection process that retrieves qualitative information about previous events and a familiarity process that retrieves quantitative memory strength, or recency, information. Because item recognition tests are expected to rely more heavily on familiarity than are source memory tests, it should be possible to observe dissociations between source and item recognition memory.

One of the documented differences between item and source recognition has been in the shape of the receiver operating characteristics (ROCs) observed in these two types of recognition tests. An ROC is a plot of the recognition hit rate against the false alarm rate as a function of response criterion or response confidence. As illustrated in Figures 1A and 1B, item recognition ROCs are curved in probability space and linear in z-space (see Egan, 1975; Murdock, 1974; Yonelinas, 1994, for earlier discussions of this pattern of results). In contrast, source ROCs are more linear in probability space, and they exhibit a pronounced U-shape when plotted in z-space. In fact U-shaped zROCs are consistently observed in studies of source recognition, associative recognition (e.g., “Was A paired with B?”), as well as plurality-reversed recognition (e.g., “Was A presented as a singular or plural word?”). For example, a review of the ROC studies examining these tasks, (Parks & Yonelinas, 2007; also see Yonelinas & Parks, 2007) showed that U-shaped zROCs were observed in 52 out of 59 conditions from 17 different studies.

The U-shaped zROCs seen in source and associative memory studies had originally been predicted by a dual process signal detection model (DPSD) of recognition which assumes that recollection reflects a threshold retrieval process and familiarity reflects a Gaussian signal detection process (e.g., Yonelinas, 1994, 1997, 2001). Because studied items are expected to be more familiar than nonstudied items, classical signal detection theory is assumed to provide a good description of the familiarity process (see Figure 2). The idea is that nonstudied items have some average level of familiarity, but there is item variability such that the nonstudied items form a Gaussian-shaped distribution. Studying an item is assumed to increase its familiarity, and thus the distribution of old items is shifted to the right (i.e., reflecting an increase in familiarity strength). The Gaussian assumption of signal detection theory is directly tested by plotting the ROCs in z-space (Swets, 1964). If the model is correct, then the empirical zROCs should be linear (i.e., a normalized Gaussian ROC becomes linear).
In contrast, because recollection is assumed to reflect the retrieval of qualitative information about a study event (e.g., when or where the item was studied), it is not expected to occur for every item. That is, recollection can fail in the sense that a subject may not always be able to recollect where or when an item was encountered before. Therefore, recollection is assumed to be best described as a threshold process because signal detection theory does not account for memory failures. Threshold theory assumes that some proportion of the items will be recollected but that others will not be recollected (i.e., the latter items will fall below the recollection strength threshold). Threshold theory does not make assumptions about the specific shape of the recollection strength distributions, but two examples of threshold distributions are plotted in Figure 2 (for similar illustrations, see Macmillan & Creelman, 2005). As seen in these distributions, some proportion of items are recollected (R) and thus fall above threshold, whereas the remaining items fall below. For the latter items, recollection is said to fail in the sense that recollection strength no longer provides any useful information for discriminating between targets and lures (i.e., differences in strength are not detectable below the threshold). Unlike signal detection theory, threshold theory predicts a U-shaped zROC (i.e., sweeping a response criterion from right to left in each of the two threshold models in Figure 2 produces a U-shaped zROC). Threshold theory predicts U-shaped zROCs because items that are not recollected are expected to lead to lower confidence responses than are recollected items. Thus, nonrecollected items will lower performance across the middle of the zROC, which will effectively pull the middle of the function down toward the chance diagonal, creating the U-shape.

The important difference between threshold and signal detection processes, and thus between recollection and familiarity, is that one is continuous and one is all or none, because both theories assume that memory strength is continuous and both assume that memory strength can vary from item to item. Rather, the difference is that unlike familiarity, recollection strength can fall below a memory threshold, and thus there should be some items for which the subjects simply fail to recollect.

If the dual process model is correct, then the ROCs for item and source ROCs should in general be quite different. That is, under conditions in which familiarity is very useful in discriminating between targets and lures, as in item recognition where some items have been studied and some have not, the zROC should appear relatively linear, whereas under conditions in which familiarity is expected to be less useful and performance relies more on recollection, such as in tests of source memory where all items have been studied and thus are expected to be somewhat familiar, the zROC should appear more U-shaped. As described above, the empirical literature has strongly supported these predictions. There are now other similar models that make comparable predictions (e.g., the variable recollection dual process model, Sherman, Atri, Hasselmo, Stern, & Howard, 2003; the some-or-none model, Kelley & Wixted, 2001; and the mixture model, DeCarlo, 2002). Importantly, in all these models there is assumed to be a threshold process that plays a critical role in source recognition tasks and thus leads the source zROCs to become U-shaped.

However, a limitation of all of these models is that they are not computational, and so they provide only limited insight into why or how these threshold or signal detection signals arise. That is, they start by assuming that the memory strength distributions are signal detection or threshold in nature and do not provide a computational explanation for why this would be the case or how these processes might be implemented in the brain. Although some of these models do assume that the recollection process is dependent on the hippocampus, whereas the familiarity process is subserved by medial temporal lobe regions such as the perirhinal cortex (e.g., the DPSD and variable recollection dual process models), beyond these basic ideas the neural substrates of these processes have not been specified.
In the current article, we will argue that an existing neurocomputational model of the medial temporal lobes, the complementary learning systems (CLS) model (McClelland, McNaughton, & O’Reilly, 1995; Norman & O’Reilly, 2003), represents a viable candidate for modeling recollection and familiarity and, specifically, for accounting for the types of ROCs observed in various item and source recognition memory studies. We begin by describing the model in more detail. Second, we describe the types of ROCs that the model naturally produces and argue that it is generally consistent with the existing ROC results. Third, we describe a novel set of predictions that the model makes regarding when source zROCs should be U-shaped and when they should become more linear. Finally, we describe three new source ROC experiments conducted to test the predictions of the model.

The CLS Model and Recognition Memory

Recently, Norman and O’Reilly (2003) tested the CLS model and showed how it can be used to explain a wide range of extant recognition data. The CLS model is designed to incorporate anatomical knowledge about the interconnectivity and activation levels of different regions within the medial temporal lobes, and it posits that recognition memory is subserved by two functionally and anatomically separate but complementary structures. Namely, recollection is dependent on the hippocampus, whereas familiarity is subserved by the medial temporal lobe cortex (MTLC), which includes the perirhinal, entorhinal, and parahippocampal cortices. The hippocampus, they argued, is specialized for the rapid encoding of specific events or episodes while minimizing overlap, and potential confusability, with other events that share similar features. In contrast, the neocortex learns more gradually to form stable, highly overlapping representations of the general environment.

The hippocampal component of the CLS model comprises areas of entorhinal cortex, CA1, CA3, and dentate gyrus. An illustration of the hippocampal component is presented in Figure 3A. Each small square in the illustration represents a unit. A unit in turn represents a group of neurons with an activation value, ranging from zero to one, which signifies the proportion of those neurons that are firing at a given moment. Stimuli are presented to the network via the input layer of the entorhinal cortex. In Figure 3A, each stimulus comprises 24 slots of 10 units. A slot can be thought of as a feature dimension, or stimulus attribute (e.g., category, shape, color), with 1 active unit per slot.

Briefly, the recollective process works as follows: At encoding, a stimulus pattern is presented to the input area of the entorhinal cortex. Activity spreads, via feedforward projections, to area CA1 both directly and indirectly via the dentate gyrus, and to CA1. Hebbian mechanisms of long-term potentiation and long-term depression lead to selective strengthening and weakening of weighted connections between and within hippocampal layers. High levels of inhibition within the hippocampal layers, particularly the dentate gyrus, ensure sparse activation and minimal representational overlap between encoded items. Area CA3 is the primary site of pattern storage and retrieval. During retrieval, a partial cue is presented to the entorhinal cortex, which leads to reactivation of the stored representations in CA3. Recurrent connections within CA3 help the system to retrieve complete representations from partial stimulus cues. Area CA1 requires strong input to become active and thus will generally become active only during retrieval when the input from CA3 is relatively complete. Area CA1 then activates at the output layer of the entorhinal cortex the original stimulus. The overall performance of the system can be assessed by comparing the retrieved pattern at entorhinal cortex output with the original stimulus pattern.

In contrast, the neocortical component of the CLS model supports familiarity-based recognition. The neocortical model, as illustrated in Figure 3A, comprises two layers: an input layer and an MTLC layer. The input layer corresponds to projections from various association areas of the neocortex and is also represented here with 24 feature slots. This structure captures the notion that the different features of the input stimulus are represented in different areas throughout the lower level neocortex. Comparatively low inhibition within the MTLC layer (e.g., compared with that in DG and CA3 layers of the hippocampal model) and slow learning over multiple stimulus presentations leads to the gradual formation of stable, overlapping representations. Familiarity is assessed by measuring the relative sharpness of stimulus representations within the MTLC layer. When a stimulus is first presented to the network, random weightings lead to varying degrees of activation. During learning, lateral inhibition leads to bolstering of the most active units and suppression of less active units, resulting in a sharper overall representation. Despite the slow learning rate, single stimulus presentations still lead to reliable and lasting weight changes. Thus, at time of retrieval, studied items will be associated with a smaller number of more highly active units. For a more thorough description of the CLS model and its underlying principles, the reader is encouraged to see Norman and O’Reilly (2003) and O’Reilly and Munakata (2000).

What Types of ROCs Are Produced by the CLS Model?

The first step in testing the model’s ability to account for recognition was simply to determine the types of ROCs the model would naturally produce. Thus, we conducted several initial simulations with the hippocampal component of the model, followed by several comparable simulations using the neocortical component of the model. The parameters used in the current simulations are outlined in the Appendix. Note that no experimental manipulations were implemented in these initial simulations; they were merely the first steps in determining the general shapes of ROCs produced by the two components of the model. To allow for the simulation of source recognition tasks, each study item was made up of two parts; 18 slots were used to represent item features, and 6 were used to represent source features. To create an item set, we began with a random prototype pattern and then manipulated interitem (feature) overlap by varying the number of prototypical features that were shared between items (we used 20% average overlap to begin with and examined the effects of varying this value in simulations we describe later). In this way, the items were not a selection of entirely random features, but rather they shared some degree of similarity, much as study lists in empirical recognition experiments are constructed of items from the same general class (e.g., all items are words or all items are pictures). From the source slots, the six leftmost units were activated when we wanted to represent “left” and vice versa for “right.” The choice of unit position is arbitrary, but this configuration made for easy visual interpretation.
Figure 3. A: Models of the hippocampus and neocortex. The left panel is an illustration of the hippocampal model. A stimulus is first represented in the entorhinal cortex (input). At encoding, activity spreads from the entorhinal cortex to area CA3, both directly and indirectly via the sparsely active dentate gyrus, as well as to area CA1. Patterns are learned through Hebbian weight changes between coactive units, including recurrent connections within CA3. At retrieval, area CA3 pattern-completes from a partial input stimulus and then sends this information to CA1, which converts the representation back to its original pattern at the entorhinal cortex (output). The right panel is an illustration of the MTLC model. A new stimulus will cause varying degrees of activation among MTLC units. As the stimulus is repeated, the representation sharpens, leading to a reliable difference in the overall activation pattern compared with that of novel stimuli. B: Strength distributions generated by the hippocampal and neocortical models (20% feature overlap). The left panel depicts a simulation histogram of left versus right source retrieval distributions produced by the complementary learning systems (CLS) hippocampal model. The right panel depicts simulation histograms produced by the CLS neocortical model. C: Simulation receiver operating characteristics (ROCs; left) and zROCs (right) generated by the hippocampal and neocortical models.
Each simulation consisted of 500 simulated subjects, each of which studied a list of 10 items from the left source and 10 items from the right source in random order. New items were generated for each list and subject. Different learning rates were explored in order to obtain levels of performance that were comparable with behavioral data (see the Appendix for values). Following Norman & O’Reilly (2003), we simulated a study-test lag by including 10 interference items at the end of the list that were not tested at retrieval. In addition, we incorporated encoding variability (variability in how well stimuli are encoded at study). For the hippocampal model, a source recognition test was simulated by presenting each study item to the entorhinal input layer and requiring the network to retrieve and activate at the output layer both the item and its correct source. Performance was then measured by using an adaptation of the match–mismatch rule proposed by Norman and O’Reilly (2003) and described by the formula

\[
\text{score} = \frac{\text{act_win} \times \text{numslots}}{\text{match} - \text{mismatch}},
\]

where \(\text{match}\) is the summed activation of recollected features (at entorhinal output) that match the item cue (at entorhinal input) and \(\text{mismatch}\) is the summed activation of mismatching features; \(^1\) \(\text{numslots}\) represents the total number of feature dimensions (which was always 24). Because source was not included with the item cue, we needed some method of assessing which source was retrieved by the hippocampal model and how well. To do this, we matched the output against ideal left and right source patterns, in turn, and took the higher of the two scores. A left (minus) or right (plus) polarity was then assigned to a final retrieval strength score (combining item and source) according to which source was favored.

The simulations of the neocortical model were conducted in a manner parallel to that used in the hippocampal simulations. However, because the neocortical model does not support recall, we used a “source match” strategy in which we cued the MTLC at test with both the item and source information, first with the left source and then the right. A familiarity score was obtained for each of the left–right presentations, and we took the difference between the two scores. Thus we examined, as a global match, which item–source combination was more familiar. As discussed earlier, presentation of a stimulus to the neocortical model results in sharpening of its representation in the MTLC layer. Familiarity then can be indexed by the average activation of the “winning” units (i.e., those that overcome inhibition), with greater activation corresponding to higher familiarity strength. Here we took the average activation of the top 10% most active units. This is referred to as the act_win measure (Norman & O’Reilly, 2003). Alternative measures of familiarity and their biological plausibility are discussed by Norman and O’Reilly (2003).

Figure 3B shows the strength distributions for simulations performed by using the hippocampal and neocortical models. The hippocampal model histogram demonstrates a threshold-like output function. That is, for items studied on the right side (i.e., the light-colored distribution), there is a large peak on the right side of the figure indicating that those items pattern-completed with the right side source information. In addition, there is a large peak centered at zero indicating that there were also items that did not lead to pattern completion with either left or right source information. Thus, consistent with threshold theory, there were two classes of items: items falling above threshold for which there was an accurate source recollection and items falling below threshold for which there was essentially zero recollected information. Conversely, for the items studied on the left side (i.e., the dark-colored distribution), there was a peak representing left source pattern completions and a middle peak representing retrieval failures. Notably, even though the hippocampal model failed to retrieve the source information of many studied items, when it did pattern-complete, it tended to be accurate (i.e., there were virtually no items that led to the pattern completion of the wrong source). Although this is not a necessary assumption of all threshold theories, it is quite common (see Yonelinas & Parks, 2007).

In contrast, the neocortical model produces two overlapping, Gaussian-shaped distributions, consistent with signal detection models of familiarity. So, unlike the hippocampal component, there was no indication that familiarity reflected a threshold process. That is, the familiarity strength signals of the two sources were Gaussian and were simply shifted left or right relative to one another. Note that subsequent simulations showed that when performance was increased the two familiarity distributions moved further apart, but they still retained their Gaussian shape.

To determine the types of ROCs that are produced by the hippocampal and cortical model components, we converted the underlying strength distributions into ROCs by plotting the proportion of targets and lures that exceeded a response criterion that was varied from conservative to liberal (i.e., moved from right to left). The ROC and zROC curves, shown in Figure 3C, indicate that the hippocampus, which is responsible for recollection, produces ROCs that are approximately linear in probability space (left graph) and U-shaped in z-space (right graph). In contrast, the neocortex, which is responsible for familiarity, produces ROCs that are curved in probability space and linear in z-space.

The distinctive patterns of ROCs produced by the hippocampal and neocortical components of the model appeared to be quite robust in the sense that they were consistently observed across various different parameter values and simulation strategies that we explored. For example, we conducted additional simulations without encoding variability and found that although variability tended to reduce performance, it did not fundamentally affect the nature of hippocampal or neocortical ROCs. We also examined several alternative decision rules for the hippocampus. For example, when making the source recognition decision, rather than considering all of the features of the entorhinal output, we conducted simulations in which we considered only those that corresponded to source information. This led to a decrease in overall performance but still produced U-shaped zROCs. We also implemented a source match strategy as we had used with the neocortical model, but it too produced U-shaped zROCs. In contrast, for the neocortical simulations, the source memory zROCs were always linear and symmetrical, as in Figure 2C.

**Relating the CLS Model to the Existing Literature**

The model simulations showed that the hippocampus consistently produces ROCs that are quite linear in probability space and

1 The method of recording match and mismatch values differs slightly to that used by Norman and O’Reilly (2003), in which they counted the numbers of units active beyond an arbitrary threshold value (rather than summing their actual activation values).
U-shaped in z-space. In contrast, the neocortex produces inverted U-shaped ROCs that are linear when plotted in z-space. These results are in agreement with various quantitative models that have assumed that recollection reflects a threshold process, whereas familiarity reflects a signal detection process. For example, Yonelinas (1994) argued that the recollection of qualitative information sometimes fails and thus can be measured as a probability, whereas relevant familiarity strength information is available for all items and can be described by classical signal detection theory. Similar assumptions have been made in various other models as well (e.g., Kelley & Wixted, 2001; Sherman et al., 2003). The current simulations, however, did not “assume” that recollection reflects a threshold process; rather, they showed that the threshold nature of recollection arose naturally, at least in the current simulations, as a byproduct of the structural properties of the hippocampal network. That is, as outlined above, high levels of inhibition within the hippocampal layers, particularly the dentate gyrus, ensure sparse activation and minimal representational overlap in the hippocampus. In addition, area CA1 requires a reasonably high level of activation from area CA3 in order to become active. This will generally happen only when the pattern activated at area CA3 is relatively complete (that is, the original pattern has been reinstated). Thus, only some items will lead to pattern completion, whereas others will fail to pattern-complete (i.e., they will fall below threshold).

The simulations are also consistent with the quantitative models that assume familiarity produces Gaussian strength distributions, consistent with signal detection theory (e.g., Kelley & Wixted, 2001; Sherman et al., 2003; Yonelinas, 1994). Again however, the CLS model did not assume the strength distributions were Gaussian, as do the quantitative models; rather, it showed that the Gaussian distributions arose naturally from the structural properties of the neocortical network. Slow learning and low inhibition led to largely overlapping representations and thus produced broad Gaussian strength distributions.

One aspect of the simulations that may appear surprising is that the neocortical component of the model was able to support accurate source memory discriminations. By some accounts, source memory tests provide a process-pure measure of recollection (e.g., Glanzer, Kim, Hilford, & Adams, 1999; Slotnick, Klein, Dodson, & Shimamura, 2000), and as such the current findings would not be expected. However, Norman and O’Reilly (2003) argued that the neocortex could learn new associations, although at a much reduced rate to that seen in the hippocampus. In addition, growing evidence suggests that familiarity can support discrimination in tasks like source recognition, at least in cases in which the item information is treated as a single unit that includes the associated source information (e.g., Diana, Yonelinas, & Ranganath, 2008; Quamme, Yonelinas, & Norman, 2007; Yonelinas, Kroll, Dobbins, & Soltani, 1999). Importantly, in the current simulations we treated the item and source information as a single unit (i.e., source features were no different from item features), thus one might expect the neocortex to support at least some degree of accurate source recognition.

In sum, the CLS model was found to be in good agreement with current quantitative models of recognition, and thus it can produce the types of ROCs that are observed in tests of source and item recognition. That is, if one assumes that source recognition relies heavily on the hippocampal recollection component of the model, then one expects to see U-shaped zROCs. In contrast, if one expects item recognition to rely most heavily on the MTLC familiarity process, then one expects to see relatively linear zROCs. Although the simulations described above focused on source recognition, we also conducted simulations of item recognition and found comparable results (also see Norman & O’Reilly, 2003). We emphasize, however, that although the two types of tasks are expected to vary in terms of the relative contribution of recollection and familiarity, we do not expect the item and source tests to be process pure—both processes can contribute to both types of tests. Thus, the observed empirical zROCs are unlikely to be either perfectly linear or exceedingly U-shaped. Rather, one should expect to see a range of functions varying from relatively linear to U-shaped. We also point out that there are other, more subtle aspects of the ROC literature that the model may or may not be able to account for and that we do not directly address here. For example, different variables have been found to influence the degree to which the ROCs are asymmetrical or symmetrical in item and source recognition tasks (see Yonelinas & Parks, 2007). The current simulations show only that the model can account for the general shape of the ROCs and leave questions about more specific subtleties of the ROC shape for future studies. Note also that we chose not to combine the hippocampal and neocortical components into a single simulation framework as it is unclear at this stage how the two components should realistically combine and interact, although this will no doubt be the subject of future research.

A Novel Prediction of the CLS Model

The Threshold Nature of Hippocampal Recollection Should Break Down Under Conditions of High Feature Overlap

The previous simulations demonstrated that the CLS model is capable of accounting for the types of ROCs seen in the source and item recognition literature. Our next goal was to test a particularly novel prediction concerning what happens in the hippocampus when it is presented with many stimuli that share a large number of their features. In general, the hippocampal component of the model operates as a threshold retrieval process, and thus it provides a viable way of accounting for the U-shaped zROCs seen in source recognition tasks. Importantly, however, Norman and O’Reilly (2003) found that the hippocampal component does not always exhibit this threshold nature. Specifically, they conducted simulations of an item recognition paradigm in which old and new items shared either a low or high percentage of prototypical features (i.e., the items were all very similar to each other or they were very distinct). They found that greatly increasing the feature overlap between items significantly hindered the ability of the hippocampal model to assign distinctive representations to each item, making stimulus representations more prone to interference.

The result was a decline in overall performance, but more importantly, instead of lures reliably failing to trigger a response from the hippocampus, there was now a continuous Gaussian-like pattern of retrieval strengths resulting in overlapping old–new item distributions. In order to verify that this pattern of results was reliably predicted by the model, and specifically apparent within a source memory paradigm, we conducted several simulations in
which we directly manipulated the degree of feature overlap. Once we had determined that it was a general prediction of the model, we conducted behavioral studies to directly test it.

Simulating Source Retrieval Across Varying Degrees of Item Feature Overlap

In the current set of simulations, we examined changes in retrieval strength distributions, ROCs, and zROCs as the amount of feature overlap between items was systematically varied. To simulate a low overlap condition, we used an average 20% feature overlap as used in the previous simulations (but applied a higher learning rate to avoid floor performance as overlap was increased). For high overlap, we used 50% feature overlap. We also explored a range of intermediate values and report here on a representative middle-range value of 33%. All other methods of the simulations were the same as those described earlier for the hippocampal component of the model and are described more fully in the Appendix.

Histograms showing retrieval strength distributions at 20%, 33%, and 50% item overlap are presented in Figure 4A. As predicted, at low overlap (20%) the hippocampal model was very effective at attributing the correct source to an item cue. Further, the hippocampus exhibited a threshold-like behavior whereby there was either a total retrieval failure, indicated by the spike in the middle, or a very high strength retrieval, indicated by the spikes at either end. As overlap was increased, the hippocampus began to exhibit a breakdown in threshold retrieval. At 50% overlap, there were relatively few high strength retrievals and few retrieval failures. Instead, two overlapping, Gaussian distributions emerged.

ROCs and zROCs for the low and high overlap simulations are presented in Figure 4B. The plots revealed a pattern consistent with our expectations. At 20% overlap, the zROC exhibited the most curved U-shaped function, with a quadratic coefficient of .52. At 33% overlap and then 50% overlap, the zROCs became more linear, with quadratic coefficients of .35 and .16, respectively.

In addition to the reduction in zROC curvature, the ROCs and zROCs also showed a consistent decline in overall performance as feature overlap was increased. The co-occurrence of these effects was not unexpected since greater overlap theoretically constitutes a more difficult task. However, the question then arises: Are these two effects dissociable or is the variability in zROC curvature really just a by-product of performance level? This issue is further explored in the next set of simulations.

Simulating Source Retrieval With Performance Matched Across Conditions

In the current simulation, we aimed to determine whether the difference in curvature that we saw across the feature overlap manipulation would still be observed under conditions in which we held overall performance constant. To do this, we simply varied the learning rates for the high and low overlap conditions so they led to similar levels of overall performance. That is, the initial learning parameter (.0075) was decreased to .0045 for the low overlap condition and was increased to .0300 for the high overlap condition.

Figure 4. A: Simulation hippocampal source retrieval histograms of the left and right source retrieval distributions for low feature overlap (20%), medium feature overlap (33%), and high feature overlap (50%) from left to right, respectively. B: Corresponding simulation receiver operating characteristics (ROCs; left) and zROCs (right) of the hippocampal model as a function of feature overlap.
The retrieval strength distributions for conditions of low and high feature overlap, presented in Figure 5A, retained the expected effect when matched for performance. While lowering the learning rate for the 20% overlap condition diminished overall performance, the hippocampus still exhibited its threshold nature. Accurate retrieval usually corresponded with high confidence retrievals. Moreover, diminished performance was reflected in an increase in clear retrieval failures (the middle spike) rather than a breakdown in threshold functioning.

Raising the learning rate in the 50% overlap condition improved performance to a level comparable with that of the 20% overlap (with lowered learning) condition, but the distribution still exhibited effects of diminished pattern separation. High strength retrievals disappeared from the distribution and were replaced with more centralized and overlapping, Gaussian-like distributions.

ROCs and zROCs for low and high overlap conditions are presented in Figure 5B. The probability ROCs indicated a very linear function for the low overlap condition as compared with a distinctly curved function for the high overlap condition. The zROCs verify that performance was well-matched across conditions, showing comparable y-axis intercepts. At 20% overlap, the zROC function exhibited a clear U-shape, with a quadratic estimate of .36. At 50% overlap, the zROC function was almost perfectly linear, with a quadratic estimate of .02.

Model Summary

In sum, the hippocampal simulations examining the effects of feature overlap showed that under conditions of very high feature overlap, the model predicts that the threshold nature of the hippocampus will begin to break down, leading source memory zROCs to become less U-shaped than is typically observed.

Source Recognition Experiments: Testing the Feature Overlap Prediction

We conducted three experiments to test the model’s predictions concerning the effects of feature overlap on source recognition ROCs. In each experiment, subjects studied items on the left or right side of a computer screen and then were given a recognition...
memory test in which they indicated if the item had been studied on the left or right. Subjects rated their confidence of each recognition judgment, and these ratings were used to plot ROCs (Yonelinas, 1997). In Experiment 1, the high and low feature overlap conditions consisted of animal names and random nouns, respectively. In Experiment 2, the high and low overlap conditions reflected pictures of houses and random pictures. In Experiment 3, feature overlap was varied by limiting the presentation duration of the study pictures.

The results of each experiment were analyzed by examining the confidence ROCs in each condition. For illustrative purposes, we present the group ROCs, but we conducted our statistical analyses on the individual subject ROCs. The ROCs were plotted on probability coordinates and z-coordinates. As with the model simulations, the shape of the ROC was assessed by conducting a polynomial regression analysis on the zROC and measuring the quadratic term. We predicted that, in line with the simulation results, the quadratic values would be smaller for the high overlap condition (i.e., showing a less pronounced curve) than for the low overlap condition.

Experiment 1

**Source Recollection for High Versus Low Overlap Word Lists**

Experiment 1 was our first attempt to construct behaviorally equivalent conditions of comparable high and low feature overlap. To do this, we used words taken from one category (e.g., animals) as our high overlap condition and a list of random nouns (excluding animals) as our low overlap condition. We assumed that the animal names would share more semantic features than the random words, and thus the threshold nature of recollection should be reduced for the animal words compared with that of the random words (i.e., the animal zROC should be more linear than the random zROC).

**Method**

**Subjects.** Thirty undergraduate psychology students at the University of California, Davis (mean age = 20 years) participated in the experiment for course credit.

**Materials.** Two word lists were compiled to make up low and high overlap conditions. The high overlap list comprised 120 commonly identifiable animal species and subspecies, randomly selected from the Kucera and Francis (1967) word pool, with mean word frequency of 5.68, \(SD = 12.45\). The low overlap list comprised 120 nonanimal nouns, randomly selected from Kucera and Francis, with comparable word frequency to the high overlap list, 5.77 (\(SD = 0.20\)).

**Design and procedure.** All experiments were performed on a PC-compatible computer. The viewing distance was approximately 0.5 m. At the beginning of the session, subjects were informed that they would be presented with a series of words on a computer screen in which half would be presented on the left and half on the right in random order. They were instructed to try to remember the words and on which side of the screen they were presented. To improve performance, subjects were asked to associate the words from the two sources with two distinctive people.

The low and high overlap words were interspersed in a random presentation order and were presented for 3 s each. Subjects were not informed of the semantic structure of the study list.

Immediately following the study phase, subjects received a source memory test for all of the studied words. The words were presented in random order, one at a time, in the middle of the screen. Subjects were instructed to judge whether they thought the item had been on the left or right of the screen at the time of study. They were instructed to press 1, 2, or 3 on the keyboard if they thought it was on the left; 4 if they were sure it was on the left, 2 if they were less sure, and 3 if they were very unsure. If they thought it was on the right, they were told to respond with keys 4, 5, or 6: 6 if they were sure it was on the right, 5 if they were less sure, and 4 if they were very unsure.

Subjects were told to try to use the entire range of response keys. At test, each word remained on the screen for a maximum of 3 s. If they did not respond in this time, they were presented with the next item. The experimental session took approximately 50 min to complete.

**Results**

ROC's and zROCS were plotted for each subject as a function of response confidence, and linear and quadratic regressions were fit to each of the individual zROCs to obtain y-intercept and quadratic estimates, respectively. The y-intercept was used as a rough estimate of overall performance and the quadratic as a measure of the curvature of the zROC. A number of subjects performed very close to ceiling, making differences across conditions difficult to detect. As an attempt to diminish ceiling effects and their potential to distort the ROCs (Yonelinas & Parks, 2007), we analyzed only data from subjects with a y-intercept less than 2.5. Additionally, we conducted outlier analyses to eliminate the effect of extreme values that are occasionally found with quadratic coefficients. In all, 8 subjects were excluded for exceedingly high performance, and 2 were excluded as outliers. Finally, quadratic regression can be conducted only with zROCs that have at least 3 points, which eliminated 1 more subject from the analyses.

Average zROCs are presented in Figure 6. The figure indicates that the overall pattern of results was as expected: The high overlap condition zROC was lower and less U-shaped than that of the low overlap condition. When plotted in probability space (not shown), the ROC for the high overlap condition was lower and more curved than that for the low overlap condition. The average functions, however, sometimes do not provide an accurate reflection of the individual results, so we examined ROCs for each subject and conducted a linearity analysis to derive quadratic estimates on the zROCs for each subject. The quadratic estimates are presented in Figure 6 and show that in the low overlap condition the quadratic was significantly greater than zero, \(r(18) = 2.547, p = .020\), indicating that the zROC was significantly U-shaped. In contrast, in the high overlap condition the quadratic component was not significantly different from zero, \(r(18) = 0.949, p = .355\), indicating that the zROC did not deviate significantly from linear. The figure also suggests that the quadratic for the low overlap condition was larger than that for the high overlap condition; however, the difference failed to reach significance, \(r(18) = 1.216, p = .12\).
As the large error bars in Figure 5 suggest, there was a great deal of variability in the quadratic terms. We examined the results more carefully and found two factors that may have led to this variability. First, high levels of performance in many subjects appeared to lead to ROC points that approached 100%, which produced irregularly shaped zROCs (i.e., neither linear nor continuously curved). Second, in a related study we conducted (Parks & Yonelinas, 2008), we found that quadratic estimates become difficult to reliably estimate when there are fewer than 100 responses per subject/condition. The numbers of trials in the current experiment (i.e., 60 responses per subject/condition) may have been too small to support stable quadratic estimates. Both of these potential shortcomings were addressed in the next experiment.

Experiment 2
Source Recollection for High Versus Low Overlap Picture Lists

Experiment 2 examined the effects of feature overlap by contrasting memory for random photographs and photographs of houses. We assumed that the houses would have greater feature overlap (both perceptual and conceptual) than would random pictures, and so the zROCs should be less U-shaped for the house pictures than the random pictures. We increased the number of trials from 60 to 120 items per subject/condition, used a between-subjects design, and selected presentation durations that were designed to avoid ceiling levels of performance.

Method

Subjects. Forty-eight undergraduate psychology students (mean age = 20 years) participated in the experiment for course credit.

Materials. Two picture lists were compiled to make up low and high overlap conditions. For the low overlap list, 240 random distinct pictures were selected from the Web site Webshots. For the high overlap list, 240 pictures of different houses, photographed from the outside, were taken from Webshots.

Design and procedure. Subjects were randomly assigned to either the high (house pictures) or low (random pictures) overlap condition. The instructions and testing procedure were otherwise the same as those used for Experiment 1. At study, pictures were displayed one at a time on the left or right of the screen in a random order. Pilot studies indicated that subjects performed markedly better on the low overlap list. To achieve more comparable performances across the two conditions, high overlap pictures were displayed for 8 s, with a 500-ms interstimulus interval, and low overlap pictures were displayed for 1 s, with a 100-ms interstimulus interval. At test, subjects had an unrestricted response time limit. The experimental session took approximately 50 min to complete.

Results

ROCs and zROCs were plotted as in the previous experiment. As expected, overall performance was lower in Experiment 2 than Experiment 1, and the ROC analysis indicated that there were no outlier subjects. The average zROCs are presented in Figure 6. Consistent with the findings of Experiment 1, increasing feature overlap led the zROCs to become less U-shaped and the ROCs to become more curved. An analysis of individual subject ROCs confirmed that this difference was reliable. That is, the quadratic coefficients on the zROCs were significantly greater in the low feature overlap condition compared with the high feature overlap condition, \( t(46) = 4.298, p < .001 \). The quadratics were significantly greater than zero in the low overlap condition, \( t(24) = 10.533, p < .001 \); and high overlap condition, \( t(22) = 4.663, p < .001 \), indicating that the zROCs were significantly U-shaped in both conditions.

Despite manipulating the presentation rates at study, overall performance was higher for the low overlap condition than for the high overlap condition. These results are consistent with our first set of simulations, which indicated that feature overlap would decrease performance and lead the zROCs to become less U-shaped. However, our second set of simulations showed that even when overall level of performance was matched the zROC in the high feature overlap condition was less U-shaped than in the low feature overlap condition. To investigate whether this was the case, in the current behavioral experiment we conducted a second analysis in which performance (as indicated by the linear y-intercept of the zROC) was matched by removing the 6 highest performers from the low overlap condition and the 6 lowest performers from the high overlap (houses) condition. The average zROCs and the average quadratic estimates are presented in Figure 6. Consistent with the earlier simulations, the results indicated that even when overall level of performance is matched, the high feature overlap condition leads to a significant reduction in the U-shape seen in the zROCs, \( t(34) = 2.939, p < .005 \).

Experiment 3
Source Recollection for Short Versus Long Stimulus Sampling Duration

For a final behavioral experiment, we further tested the generalizability of the effects already observed by examining a different manipulation of item overlap. In this experiment, only the random picture list was used, and a high overlap condition was created by severely reducing the stimulus sampling duration (the time the item is on screen) for one list. Our reasoning was that if the duration given to study an item is restricted, the number of features that can be extracted from the item that differentiates it from other items will necessarily be reduced. Consequently, items that are briefly presented will be less distinct from one another than those that are presented for longer durations. Importantly, however, we held the stimulus presentation rate constant—we just manipulated the time in which the study item was on screen. In this way, the short sampling duration was expected to reduce the number of item features that were sampled but to still allow adequate processing time for subjects to encode the source information (left or right side of the screen). Note that we conducted additional simulations with the CLS model in which the stimulus sampling duration was manipulated by leaving out varying numbers of features from the item at study. The results were consistent with the other simulations that showed a breakdown in threshold retrieval, although the effect appeared to be somewhat smaller than that seen with the other simulated manipulations.
Method

Subjects. Seventy undergraduate psychology students (mean age = 20 years) participated in the experiment for course credit.

Materials. Four hundred and eighty distinctive pictures were taken from Webshots; half were randomly assigned to the short sampling condition and half to the long sampling condition. The images followed the same visual layout as Experiment 2.

Design and procedure. Subjects were randomly assigned to either the short or long stimulus sampling duration condition. During study, in the short condition pictures appeared on the screen for a duration of 200 ms followed by a 900-ms interstimulus interval. In the long condition, pictures appeared for 1,100 ms with no interstimulus interval. Pictures were displayed one at a time on the left or right of the screen in a random order. The instructions and testing procedure were otherwise the same as those used for the two previous experiments. At test, subjects had a maximum of 10 s to respond to each item. The experimental session took approximately 40 min to complete.

Results

Individual ROCs and zROCs were plotted as in the previous two experiments, and quadratic regressions were fit to each zROC. Four subjects were excluded from analysis for failing to distribute their responses sufficiently to produce meaningful ROCs, and 4 were excluded as outliers. The average zROCs, presented in Figure 6, reveal a pattern consistent with the previous two experiments. The zROCs were slightly more U-shaped for the low overlap condition than the high overlap condition, and the ROCs were slightly more curved for the high overlap condition than the low overlap condition. An examination of individual zROCs indicated that the quadratic coefficient (see Figure 6) was greater in the low overlap (long duration) condition than in the high overlap (short duration) condition, although the difference fell just short of significance, $t(60) = 1.603, p = .057$. The quadratics were significantly greater than zero in the low, $t(32) = 8.416, p < .001$; and high, $t(28) = 8.990, 0.001$, feature overlap conditions, indicating that the zROCs were significantly U-shaped in both conditions.

Discussion

Across all three experiments, conditions of low feature overlap produced zROCs that appeared more U-shaped than under conditions of high overlap. Although quadratic estimates did not show significant changes in all experiments, a consistent trend clearly emerged in comparing the results (Figure 6). In order to assess the effect of high and low feature overlap across the three experiments,

![Figure 6. Summary of behavioral data showing mean quadratic estimates (below) for high and low overlap conditions in Experiments 1 through 3 (error bars represent standard errors). Above are corresponding average zROCs. ROC = receiver operating characteristics.](image-url)
we conducted a combined analysis, which indicated that the low overlap conditions yielded significantly larger quadratic coefficient estimates than did the high overlap conditions, \( F(1, 142) = 9.64, \text{MSE} = 0.058, p < .005 \). However, there were no significant differences between the experiments, \( F(2, 142) = 1.632, \text{MSE} = 0.058, p = .199 \); and there was no evidence of an interaction, \( F(2, 142) = 0.761, \text{MSE} = 0.058, p = .469 \), showing that the feature overlap effects did not differ appreciably across experiments. Thus the results show a direct relationship between the degree of feature overlap and the curvilinearity of source recognition zROCs, as predicted by the CLS model. The model also predicted that increasing feature overlap would lead to a decrease in overall performance, which was the case in each experiment. Importantly, however, the simulations indicated that the change in shape should also be observed when overall performance is controlled. In Experiment 2, we compared the ROCs for subgroups of subjects in such a way that we matched overall performance levels, and we found that the high feature overlap condition still led to a less U-shaped zROC than did the low feature overlap condition. In addition, in Experiment 3 we manipulated feature overlap by limiting the sampling duration of each study item, and as predicted by the model this led the source zROC to become less U-shaped. Note that our simulations had suggested that this manipulation might lead to a more subtle effect than that observed with the other manipulations. This effect did appear to be slightly reduced in magnitude, but this difference should be interpreted cautiously because we found no significant difference in the magnitude of the effects across the experiments.

Because the standard method of assessing linearity (i.e., regression) does not take x-axis variability (i.e., false alarm variability) into account, we verified the quadratic coefficient results by using another method of assessing linearity that does consider false alarm variability. That is, we fit two models to the data by using maximum likelihood estimation, one that predicts only linear zROCs (i.e., the traditional unequal variance signal detection [UVSD] model) and another that predicts U-shaped zROCs (i.e., a UVSD model that included a threshold process [UVSD+]). The traditional UVSD model is nested under the UVSD+ model, which allowed us to compare the fits of the two models by using the G² statistic. The UVSD+ model fit the data better than the UVSD model in all three experiments: Experiment 1, \( F(1, 29) = 22.85, \text{MSE} = 2.70, p < .001 \); Experiment 2, \( F(1, 46) = 49.17, \text{MSE} = 11.81, p < .001 \); Experiment 3, \( F(1, 64) = 80.91, \text{MSE} = 12.49, p < .001 \), indicating that the zROCs had a significant U-shape and that a simple signal detection model does not provide an adequate account of the reported ROCs.

General Discussion

Quantitative models such as the DPSD (Yonelinas, 1994) prove useful in describing the behavioral regularities of recognition memory but provide limited insight into how the physical structure of the brain might give rise to them. Neurally instantiated computational models offer a logical next step, using known biological constraints to simulate how one or more memory processes might emerge naturally and dynamically from a few simple rules. The CLS model represents an important step toward this end. As we have described, the model consists of two distinct memory components, the hippocampus and the neocortex, that perform complementary learning roles. The hippocampus learns quickly and forms detailed, pattern-separated representations of novel stimuli and associations; this is useful for remembering details of specific events. In contrast, the neocortex learns gradually to form stable, highly overlapping representations; it works to incorporate environmental regularities into an established semantic framework. The current study revealed a rather startling convergence between the quantitative dual process models and the CLS model, and it confirmed that the CLS model provides a powerful neurocomputational model of recollection and familiarity.

The Hippocampus Supports a Threshold Recollection Process, Whereas the Neocortex Supports a Signal-Detection-Based Familiarity Process

In our initial set of simulations, we examined the types of strength distributions and ROCs produced by the hippocampal and neocortical components of the CLS model. The hippocampal component, which supports recollection, generally produced threshold strength distributions and thus produced U-shaped zROCs. In contrast, the cortical component of the model, which corresponds to familiarity, produced Gaussian strength distributions and therefore generated linear zROCs.

These results are consistent with various quantitative dual process models that have assumed that recollection reflects a threshold process, whereas familiarity reflects a signal detection process (e.g., Kelley & Wixted, 2001; Sherman et al., 2003; Yonelinas, 1994, 1999, 2001). However, unlike the quantitative models that simply assumed that recollection and familiarity reflect threshold and signal detection processes, respectively, the CLS model showed that the threshold and signal detection properties of recollection and familiarity arose naturally as a byproduct of the structural properties of the hippocampal and neocortical networks supporting recognition memory.

To the extent that the CLS model can produce ROCs consistent with those produced by the current quantitative dual process models, it is in good agreement with the existing ROC literature. If source recognition relies more heavily on recollection than familiarity, then the zROCs should be U-shaped, as is almost always observed in source memory tests. In contrast, if item recognition relies heavily on familiarity, then the expected zROCs should be relatively linear, as is typically observed in those tests.

The Threshold Nature of Recollection Breaks Down Under Conditions of High Feature Overlap

Although it is important for new models to account for the existing literature, it is also essential that they move beyond other existing models and produce novel predictions. The CLS model moves beyond the current quantitative models of ROCs by describing the neurocomputational underpinnings of the processes involved in recognition memory. In addition, our second set of simulations indicates that the CLS model accurately predicts a pattern of results that would not have been predicted by the existing models, namely that the threshold nature of recollection should begin to break down under conditions of high feature overlap.

We tested the model’s feature overlap prediction in three experiments that manipulated the degree of feature overlap in differ-
ent ways, including increasing the number of words from one semantic category (i.e., animal names, Experiment 1), increasing the number of pictures of one type (i.e., house pictures, Experiment 2), and decreasing the opportunity to fully sample the item features of each picture (Experiment 3). The predictions of the model were verified by these experiments: High overlap resulted in source zROCs that were more linear than those found for low overlap conditions, even when overall level of performance was controlled. These differences in the shape of the zROCs support the CLS model’s prediction of a shift from threshold to Gaussian-like strength distributions with increasing feature overlap. We note that the breakdown in threshold nature of recollection was never complete in the current studies, in the sense that the zROCs generally retained a U-shape, even in the high feature overlap conditions. In all conditions, except one in the first experiment, the deviation from linearity was significant. So even in the high feature overlap conditions, the hippocampus does not appear to be adequately described by a signal detection process. Future studies will be useful that determine whether it is possible for the hippocampus to produce truly Gaussian strength distributions under conditions of even more extreme levels of feature overlap.

This shift in distributions from threshold toward Gaussian is best understood in terms of a failure of pattern separation. When feature overlap is low, area CA3 is able to encode distinct representations for each item. Consequently, a test item will either cause activation to spread from CA3 to CA1, resulting in a complete and accurate retrieval of source information, or it will fail to lead to pattern completion. However, if a class of items is made extremely similar, the ability of the hippocampus to perform pattern separation can be exceeded. That is, CA3 is no longer able to extract distinctive representations for each study item. Essentially, because all the items are so similar, an item’s representation in CA3 can be linked to both of the sources in CA1. Consequently, the hippocampus will rarely fail to produce an output (i.e., decreasing the threshold nature of the strength distributions), and there will be considerable overlap in the memory strengths associated with the targets and lures (i.e., enhancing the Gaussian nature of the strength distributions).

The CLS model goes beyond earlier quantitative models of ROCs in specifying the biological underpinnings of the memory processes. In addition, the feature overlap prediction that we tested was derived directly from the CLS model, and without that model there was little a priori reason to expect that the linearity of the source ROCs would be related to this manipulation. There is nothing inherent in the DPSD (Yonelinas, 1994), or any of the other quantitative models, that would lead it to predict that increasing feature overlap would affect the shapes of the ROCs. This, of course, is because those models do not specify the nature of the underlying representations or the mechanics that produce the different strength distributions. There are, undoubtedly, different ways in which these quantitative models could be modified to account for the current results. One possibility is to assume that familiarity becomes more useful in supporting source discriminations as feature overlap is increased. However, we are not aware of any empirical evidence to support such a claim. In fact, Norman and O’Reilly (2003) argued that according to the CLS model, the cortical familiarity process relative to recollection actually becomes less, rather than more, useful as feature overlap increases.

Although quantitative models like the DPSD may be able to fit the results post hoc, it is the CLS model that predicted those results.

If the CLS interpretation of the current results is correct, it points to an important boundary condition for the use of earlier quantitative models to derive estimates of recollection and familiarity from ROCs. That is, the curvilinearity of the ROCs in probability space is used indirectly to determine the contribution of familiarity in these models (e.g., Yonelinas, 1994). If the model were fit to the ROCs observed in the high overlap conditions, the greater degree of curvilinearity in those functions would lead the model to overestimate the contribution of familiarity to performance (and potentially underestimate the contribution of recollection). The model’s parameter estimates should therefore be interpreted very cautiously in conditions of high feature overlap, such as the study lists used here that contained hundreds of similar pictures of houses. In situations like this, it would be prudent to include other measures of recollection and familiarity such as remember/know (Tulving, 1985), process dissociation (Jacoby, 1991), or structural equation modeling methods (e.g., Quamme, Yonelinas, Widaman, Kroll, & Sauve, 2004; Yonelinas, 2002) and to look for convergence across methods to verify the results of the ROC analysis (see Yonelinas, 2002, for a discussion of these various methods).

Does the CLS model supersede the earlier quantitative models? No. As we see it, the two approaches are complementary in furthering our understanding of the processes involved in recognition memory, and they serve very different functions. Whereas the computational models have the advantage of linking more directly to neurobiology, the simple quantitative models have the advantage of serving as measurement models that can be used to estimate the contribution of different processes to memory. Given the complexity of computational models, they are not particularly useful at providing quantitative measures of ROC shape, nor is it obvious how they can be used to derive estimates of recollection and familiarity.

Alternative Theoretical Approaches

The current results pose rather serious problems for pure signal detection theories of recognition (e.g., Donaldson, 1996; Dunn, 2004; Wixted, 2007) and are in much better agreement with various hybrid models that incorporate both signal detection and threshold mechanisms (DeCarlo, 2002; Kelley & Wixted, 2001; Sherman et al., 2003; Yonelinas, 1994, 1997, 2001). In fact, the ROCs observed in the current study join a large body of research that rules against pure signal detection models. If recognition memory judgments are based solely on the Gaussian strength distributions typically assumed by signal detection theory, then the zROCs must be linear. In source recognition tests such as those examined here, as well as in associative recognition and plurality-reversed recognition, the observed zROCs are invariably U-shaped (e.g., see Figures 1 and 5; for reviews, see Parks & Yonelinas, 2007; Yonelinas & Parks, 2007). Importantly, even if the signal detection models are modified to include separate Gaussian strength measures for recollection and familiarity (Rotello, Macmillan, & Van Tassel, 2004; Wixted, 2007), they still fail to produce the types of U-shaped zROCs so often reported in source tests. These results have led many memory researchers to argue that it is necessary to move beyond pure signal detection models.
The U-shaped zROCs observed in the current studies are consistent with a number of hybrid quantitative models. In fact, as mentioned previously, they were predicted by the DPSD, the first hybrid model to supplement a familiarity-based signal detection process with a threshold recollection process (Yonelinas, 1994, 1997, 1999). Other models have also adopted this hybrid approach of incorporating signal detection and threshold theory as well, and in fact, most have made the same assumptions about the nature of recollection and familiarity (e.g., Kelley and Wixted’s, 2001, some-or-none model; Sherman et al.’s, 2003, variable recollection dual process model). These models are therefore in good agreement with the results from the CLS simulations that show that recollection and familiarity, in general, do produce threshold and signal-detection-like strength distributions, respectively. Further, the current simulations show why familiarity and recollection are described well by signal detection and threshold theories: Due to the relatively low inhibition and slow learning rate, the cortex builds highly overlapping representations that in turn lead to highly overlapping Gaussian distributions of targets and lures. In contrast, the representations in the hippocampus are relatively sparse and distinct from one another and therefore can produce threshold strength distributions.

Another dual process hybrid model that is consistent with the observed ROCs is the mixture model (DeCarlo, 2002). This model is formally almost identical to the dual process models, but its parameters are interpreted in somewhat different ways. It includes an equal variance signal detection retrieval process that best maps onto the cortical familiarity process of the CLS model. The second process is a threshold attentional process, which reflects the assumption that only a proportion of items are attended at encoding and only those items will increase in memory strength. Although a true attention process is unlikely to map directly to the CLS model of the medial temporal lobes, the threshold nature of the attention component in the mixture model suggests that it might be related to the hippocampal component of the CLS model. Thus, the hippocampus may simply fail to encode some proportion of the studied items and, thus, may fail to retrieve any information about these items at a later time. However, the CLS model does not limit the threshold nature of recollection to encoding. Although encoding failure should lead to recollection failure, other factors can also lead to retrieval failure. Most obvious is the degree of interference or feature overlap. The current simulations indicated that the shape of the source ROC was directly influenced by the degree to which items were similar. In these simulations, the encoding was held constant across this manipulation, yet recollection failure became less common as feature overlap increased. Other factors such as delay and test manipulations like response deadline also influence the degree to which recollection succeeds or fails (for a review, see Yonelinas et al., 2002), and so we believe that treating the threshold component of recognition as arising solely from encoding failure is insufficient.

The fact that recollection appears signal-detection-like under conditions of high feature overlap is also broadly consistent with previous work motivated by the source monitoring framework (Johnson, Hashtroudi, & Lindsay, 1993). According to that framework, source memory attributes reflect the combined contribution of various aspects of the study event, as well as retrieval and evaluation processes that yield varying degrees of recollection and familiarity. To the extent that source memory decisions reflect the combination of multiple types of information, it should not be surprising that the memory strength functions can appear to be Gaussian. Note that this approach does not explicitly assume that a single type of information is never recollected in a threshold fashion. Thus, the source monitoring framework can account for cases in which the source zROCs are more linear, and the fact that U-shaped zROCs are often found in tests involving a single feature does not necessarily contradict it.

Is feature overlap the only factor that can influence the extent to which source zROCs are U-shaped? Certainly not. There are several other manipulations that influence the shape of the source ROCs, and at least some of them reflect the operation of a very different mechanism. For example, when the familiarity of the two sources differs (e.g., one list is presented more recently than the other), source zROCs become more linear (Yonelinas, 1999), presumably because differences in familiarity are used to discriminate between items from different lists. In addition, when the item is unitized with its source, or with its paired item in an associative test, source and associative recognition zROCs can become more linear (Diana et al., 2008; Quamme et al., 2007; Yonelinas et al., 1999). In these cases, the flattening of the zROC is presumably due to the increased usefulness of familiarity for discriminating between items from different sources. Source zROCs may also become more linear when the complexity of the stimuli increases (e.g., Qin, Raye, Johnson, & Mitchell, 2001), although it is still unclear whether such an effect might be due to increased usefulness of familiarity or something more similar to the current findings, such as an increase in feature overlap that changes the nature of the recollection strength distributions.

Conclusion

The CLS model of Norman and O’Reilly (2003) was found to provide a viable account for recognition memory ROC results and was able to produce threshold recollection strength distributions and Gaussian familiarity distributions consistent with various quantitative ROC models. In addition, a novel prediction derived from that model, which is that the threshold nature of recollection should begin to break down as feature overlap is increased, was supported in new source recognition experiments. The results provide strong support for the CLS model of recognition memory and demonstrate the power of a combined computational–behavioral research approach.

References


The current simulations were based on the CLS model as described by Norman and O’Reilly (2003) and are available at http://compmem.princeton.edu/. Also see O’Reilly and Munakata (2000) for a comprehensive guide to simulating the CLS model. The following notes are a selective description of the rules and parameters (especially those specific to the current simulations).

Hebbian Learning

The Hebbian learning rule, as used by Norman and O’Reilly (2003) and applied in the current simulations is described in simple form by the following equation:

\[ \Delta w_{ij} = \epsilon y_j (x_i - w_{ij}) \]

where \( x_i \) is the activation of the sending unit, \( y_j \) is the activation of the receiving unit, \( w_{ij} \) is the weighting between the two units, and \( \epsilon \) is the learning rate. The initial weight strength is a reflection of the probability that, on average, the sending unit will be active given that the receiving unit is also active. If both the sending and receiving units become active, then we can expect that \( x_i \) will exceed \( w_{ij} \), resulting in a weight increase (long-term potentiation). However, if the receiving unit is active but the sending unit is not, the equation informs us that the weight will decrease (long-term depression). The learning rate determines how quickly the weight is updated.

k-Winners-Take-All (kWTA) Inhibition

The sparseness of activity in a pool of neurons is controlled by inhibitory interneurons that project feedforward and feedback inhibition. This was approximated with a kWTA function (see Norman & O’Reilly, 2003), which provides a current of inhibition consistently across all the units in a specified layer to limit the number of units (\( k \)) that can be strongly active (i.e., greater than .25).

Layer Sizes, Activations, and Connectivity

The parameters used in the hippocampal model are consistent with consensus views of functional architecture in the brain (e.g.,

Table A1
Layer Sizes and Activity Levels

<table>
<thead>
<tr>
<th>Layer/area</th>
<th>Units</th>
<th>Activity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hippocampus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entorhinal cortex (in/out)</td>
<td>240</td>
<td>10.0</td>
</tr>
<tr>
<td>Dentate gyrus</td>
<td>1,600</td>
<td>1.0</td>
</tr>
<tr>
<td>Area CA3</td>
<td>480</td>
<td>4.0</td>
</tr>
<tr>
<td>Area CA1</td>
<td>640</td>
<td>10.0</td>
</tr>
<tr>
<td>Neocortex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Association areas (input)</td>
<td>240</td>
<td>10.0</td>
</tr>
<tr>
<td>Medial temporal lobe cortex</td>
<td>1,920</td>
<td>10.0</td>
</tr>
</tbody>
</table>

Note. In/out = input and output layers, respectively.

Table A2
Properties of Modifiable Projections

<table>
<thead>
<tr>
<th>Projection</th>
<th>( M )</th>
<th>( \text{Var} )</th>
<th>Scale</th>
<th>% Con</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hippocampus</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EC to DG, CA3 (perforant pathway)</td>
<td>.5</td>
<td>.25</td>
<td>1</td>
<td>25</td>
</tr>
<tr>
<td>DG to CA3 (mossy fiber; encode/retrieve)</td>
<td>.9</td>
<td>.01</td>
<td>25/0</td>
<td>4</td>
</tr>
<tr>
<td>CA3 recurrent</td>
<td>.5</td>
<td>.25</td>
<td>2</td>
<td>100</td>
</tr>
<tr>
<td>CA3 to CA1 (Schaffer collaterals)</td>
<td>.5</td>
<td>.25</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>Neocortex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Input to MTLC</td>
<td>.5</td>
<td>.25</td>
<td>1</td>
<td>10</td>
</tr>
</tbody>
</table>

Note. \( M = \) mean initial weight strength; \( \text{Var} = \) variance of initial weight distribution; Scale = scaling of this projection relative to other projections; % Con = percentage connectivity; EC = entorhinal cortex; DG = dentate gyrus; MTLC = medial temporal lobe cortex.

Table A3
Learning Rate Means and Ranges

<table>
<thead>
<tr>
<th>Simulation</th>
<th>Mean ( \epsilon )</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source recognition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hippocampal model</td>
<td>.0045</td>
<td>0–.009</td>
</tr>
<tr>
<td>Neocortical model</td>
<td>.001</td>
<td>0–.002</td>
</tr>
<tr>
<td>Overlap manipulation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hippocampal model, all conditions</td>
<td>.0075</td>
<td>0–.015</td>
</tr>
<tr>
<td>Overlap with matched performance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hippocampal model, low overlap (20%)</td>
<td>.0045</td>
<td>0–.009</td>
</tr>
<tr>
<td>Hippocampal model, high overlap (50%)</td>
<td>.030</td>
<td>.025–.035</td>
</tr>
</tbody>
</table>

Note. \( \epsilon = \) learning rate.

(Appendix continues)
2002). The mossy fiber pathway was also scaled to zero during retrieval to aid pattern completion (Hasselmo, et al., 2002).

Learning Rates

We experimented with various learning rates to achieve levels of performance that were comparable with our behavioral data and to match performances across low and high overlap conditions. We were cautious, however, to constrain learning rates to ranges that are biologically plausible. We also allowed hippocampal learning rates to vary within simulation trials to achieve more realistic retrieval strength distributions. Note that, although this is a form of encoding variability (see Norman & O’Reilly, 2003), it does not translate to a mixture model (DeCarlo, 2002), because all items and their features are encoded to some degree (however, a learning rate of zero was sometimes possible, but highly improbable). Therefore, the zero strength peak observed in the distributions is not, as the mixture model would suggest, the result of encoding failure but rather a failure of retrieval. Table A3 shows the mean learning rates and ranges (uniformly distributed) used for the three simulation sets.

Received February 14, 2007
Revision received December 4, 2007
Accepted December 15, 2007
Chapter 3: Variations in Recollection: The Effects of Complexity on Source Recognition

**Authors:** Dr Colleen M. Parks, Dr Linda Murray, Kane W. Elfman, Dr Andrew P. Yonelinas (PI)

**Journal:** Journal of Experimental Psychology: Learning, Memory & Cognition

**Date of publication:** July, 2011

**Abstract:** Whether recollection is a threshold or signal detection process is highly controversial, and the controversy has centered in part on the shape of receiver operating characteristics (ROCs) and z-transformed ROCs (zROCs). U-shaped zROCs observed in tests thought to rely heavily on recollection, such as source memory tests, have provided evidence in favor of the threshold assumption, but zROCs are not always as U-shaped as threshold theory predicts. Source zROCs have been shown to become more linear when the contribution of familiarity to source discriminations is increased, and this may account for the existing results. However, another way in which source zROCs may become more linear is if the recollection threshold begins to break down and recollection becomes more graded and Gaussian. We tested the “graded recollection” account in the current study. We found that increasing stimulus complexity (i.e., changing from single words to sentences) or increasing source complexity (i.e., changing the sources from audio to videos of speakers) resulted in flatter source zROCs. In addition, conditions expected to reduce recollection (i.e., divided attention and amnesia) had comparable effects on source memory in simple and complex conditions, suggesting that differences between simple and complex conditions were due to differences in the nature of recollection, rather than differences in the utility of familiarity. The results suggest that under conditions of high complexity, recollection can appear more graded, and it can produce curved ROCs. The results have implications for measurement models and for current theories of recognition memory.

**Author contributions:** Dr Andrew Yonelinas acted as principle investigator, overseeing the research and article revisions. The behavioral experiments were devised as a collaboration of all authors and then conducted by Dr Colleen Parks and Dr Linda Murray. Analysis of the data was a joint endeavor. I was solely
responsible for the computational component of the research. This included formulating a computational hypothesis to explain how the hippocampus processes simple versus complex stimuli and thus produces different shaped ROCs. I also conducted simulations using the CLS model to test the hypothesis, and analyzed and reported on the data. I was also responsible for writing the sections of the paper that dealt with computational modeling. I also contributed to the structure and drafting of all other sections of the paper including revisions.
Variations in Recollection: The Effects of Complexity on Source Recognition

Colleen M. Parks
University of Nevada, Las Vegas

Linda J. Murray, Kane Elfman, and Andrew P. Yonelinas
University of California, Davis

Whether recollection is a threshold or signal detection process is highly controversial, and the controversy has centered in part on the shape of receiver operating characteristics (ROCs) and z-transformed ROCs (zROCs). U-shaped zROCs observed in tests thought to rely heavily on recollection, such as source memory tests, have provided evidence in favor of the threshold assumption, but zROCs are not always as U-shaped as threshold theory predicts. Source zROCs have been shown to become more linear when the contribution of familiarity to source discriminations is increased, and this may account for the existing results. However, another way in which source zROCs may become more linear is if the recollection threshold begins to break down and recollection becomes more graded and Gaussian. We tested the “graded recollection” account in the current study. We found that increasing stimulus complexity (i.e., changing from single words to sentences) or increasing source complexity (i.e., changing the sources from audio to videos of speakers) resulted in flatter source zROCs. In addition, conditions expected to reduce recollection (i.e., divided attention and amnesia) had comparable effects on source memory in simple and complex conditions, suggesting that differences between simple and complex conditions were due to differences in the nature of recollection, rather than differences in the utility of familiarity. The results suggest that under conditions of high complexity, recollection can appear more graded, and it can produce curved ROCs. The results have implications for measurement models and for current theories of recognition memory.

Keywords: source, recognition, recollection, familiarity, receiver operating characteristics

Recognition memory is largely agreed to rely on two processes: recollection and familiarity (e.g., Diana, Reder, Arndt, & Park, 2006; Wixted, 2007; Yonelinas, 2002). Dual process theory defines recollection as conscious memory for contextual or episodic details of prior events, whereas familiarity is defined as recognizing something as old without consciously recalling episodic details (Jacob, 1991; Mandler, 1980; Tulving, 1985; Yonelinas, 1994). Despite the emerging consensus on the component processes of recognition memory, there is still widespread disagreement about the nature of recollection and whether it is best described as a threshold process, whereby recollective strength can fall below a memory threshold, or as a signal detection process, whereby memory strength varies in a completely continuous manner. Answering this question is important for testing competing theories of recognition memory and because it affects how one measures recollection. For instance, if recollection is a threshold process, then explaining recollection deficits in aging or amnesia is a matter of explaining why recollection is less likely to occur successfully. On the other hand, if it is a continuous process as described by signal detection theory, then it might be incorrect to conceptualize deficits as less frequent recollection; rather, it may be that the frequency is stable but that the strength of recollection has decreased or the variability of its strength has greatly increased. As it turns out, deciding between these different ways of describing recollection has not been easy because there is evidence supporting both threshold- and signal-detection-based accounts of recollection.

The most common way of assessing the assumptions of threshold and signal detection theories of recollection has been to measure source memory and to examine receiver operating characteristics (ROCs). Source memory refers to the ability to discern the original source of an item, and it is typically expected to rely heavily (though not exclusively) on recollection (Johnson, Hashtrudi, & Lindsay, 1993; C. M. Parks & Yonelinas, 2007; Yonelinas, 2002). In the laboratory, participants will often hear two voices reading words, see the words on different sides of the computer screen, or read words in different colors. Later, they are presented with the studied items and are asked to remember who said it, where it was, or what color it had been and to rate their confidence in their source response. ROCs are then examined by plotting the hit rate against the false alarm rate across different levels of confidence (see Macmillan & Creelman, 2005, for details). The z-scores of each hit and false alarm rate are also typically plotted to produce zROCs. Plotting the ROCs in z-space is useful because signal detection theory predicts linear zROCs.
whereas threshold theory predicts U-shaped zROCs (e.g., T. E. Parks, 1966; also see Macmillan & Creelman, 2005).1 The zROCs observed in source memory tests are almost always U-shaped, in contrast to those seen in item recognition tests, which are often very close to linear (see Yonelinas & Parks, 2007, for a review).

Although signal detection theories can account for the linear zROCs typically found in item recognition, they cannot account for the U-shaped zROCs found in source memory tests. Pure threshold models could account for U-shaped zROCs, but not all source zROCs are as U-shaped as threshold theory predicts. One interpretation of this pattern of findings is that although source memory often relies heavily on recollection, it reflects the contribution of both recollection and familiarity. In line with this interpretation, the dual process signal detection (DPSD) model predicts different degrees of zROC linearity or curvilinearity depending on the relative contribution of recollection and familiarity (e.g., Yonelinas, 1999). According to this approach, recollection is a threshold process, and familiarity is a signal detection process. This combination of processes allows the model to predict zROC shapes that are different from those predicted either by pure threshold or pure signal detection models alone. Specifically, tasks that require greater proportions of recollection for successful performance are expected to result in more U-shaped zROCs. Because many source memory tests force participants to rely primarily on recollection to make accurate source decisions, the model predicts U-shaped zROCs. However, when source tests allow familiarity to serve as a basis for accurate source judgments, the DPSD model predicts zROCs that are closer in shape to those found in item recognition (i.e., more linear).

In line with these predictions, the degree to which source memory zROCs are U-shaped can be systematically varied. For instance, if the items from two sources differ in overall strength, then familiarity can support source memory discriminations. For example, if Source A items are studied 5 days before Source B items, then successful source decisions can be supported by familiarity, and the resulting zROCs become flatter (e.g., Yonelinas, 1999). Similarly, if items are encoded such that they are integrated or unitized with their sources, familiarity can also support source discriminations. For example, if participants encode the item “elphant” as coming from the “red” source by imagining a sunburned elephant, then a red elephant should be more familiar than a green elephant at test. Encoding conditions like this that promote unitization have been found to promote familiarity-based source discriminations, and they lead zROCs to become flatter (e.g., Diana, Yonelinas, & Ranganath, 2008; Diana, Yonelinas, & Ranganath, 2010; for related results, see Staresina & Davachi, 2006).

However, there are also cases in which no particular manipulation was used to foster the use of familiarity as a basis for source memory decisions, and the resulting zROCs exhibited little if any evidence of a U-shape (e.g., Glanzer, Hilford, & Kim, 2004; Qin, Raye, Johnson, & Mitchell, 2001; Slotnick, Klein, Dodson, & Shimamura, 2000). The most dramatic example of this is a study reported by Qin et al. (2001), who found almost perfectly linear zROCs in a test of source memory for spoken statements. Findings of flat zROCs in source memory, like those found in Qin et al.’s study, have suggested to some that the DPSD model and others like it are inconsistent with the data, and as indicating that recollection is a continuous process that is more in keeping with a pure signal detection model (e.g., Qin et al., 2001; Slotnick & Dodson, 2005; Wixted, 2007). Qin et al. argued, for example, that recollection was the driving force behind the source judgments and, thus, also was the process that produced the linear zROC. If true, such evidence would provide a challenge to the assumption that recollection is a threshold process.

Qin et al. (2001) argued that the reason recollection acted in a more continuous manner in their study than in previous studies was because their materials were much more complex than those used previously. In that study, participants watched videotapes of speakers making statements (schema-based statements in one study, facts and feelings in another). Both the studied items (i.e., sentences) and the sources (i.e., auditory and visual information about the speaker) in these experiments contained more information than in typical source memory experiments. Thus, even if the recollection of each aspect of the study item or source behaves in line with threshold theory, taken together they may begin to act in a more signal-detection-like manner. However, those experiments differed from previous studies in various other ways as well (e.g., the analysis was based on aggregate ROCs rather than on individual participant ROCs because there were too few trials collected from each participant). Thus, it is not possible to determine with certainty whether stimulus complexity per se was responsible for the flattened zROCs. In addition, even if complexity was shown to lead to a flattening of the source zROCs, it would not be clear whether complexity led recollection to become more continuous in nature, or whether complexity led familiarity to be more useful in making the source discriminations. That is, highly complex stimuli may facilitate the unitization of the item and source information, and thus the flattening of the zROC could be due to familiarity, rather than recollection.

We examined these issues in a series of experiments geared toward determining whether manipulations of complexity would produce differences like those found between Qin et al. (2001) and Yonelinas (1999). In Experiments 1a and 1b, we manipulated complexity of the stimulus (words vs. sentences) and the complexity of the source (audio vs. video). The simplest condition, in which participants listened to male and female speakers reading single words, mimics many source experiments that have produced fairly linear source ROCs (U-shaped zROCs). The most complex condition, in which participants watched videos of a man and a woman making statements, was designed to match the stimulus properties of Qin et al.’s study. If the complexity of the item or the source is a critical factor, then we should find increasing ROC curvilinearity with increasing complexity. Experiment 1 demonstrated that both stimulus complexity and source complexity can lead to a flattening of the source zROCs. Experiment 2 was designed to determine whether the flattening of the zROCs produced by complexity arose because complexity led recollection to become more continuous, or because it increased the reliance on familiarity-based source memory.

1 Importantly, threshold theories can also predict curvilinear ROCs if one adds additional free parameters for each confidence point, but the standard threshold models of recollection assume that when recollection occurs, it supports high confidence responses and therefore results in a linear ROC (for a discussion of such models, see Malmberg, 2002; Yonelinas & Parks, 2007).
**Experiment 1**

Experiments 1a and 1b were conducted to test the effects of stimulus and source complexity on the shape of source memory ROCs. Stimulus complexity (words vs. sentences) and source complexity (audio vs. video) were crossed, resulting in four experimental conditions in Experiment 1a (audio-word, audio-sentence, video-word, and video-sentence). In Experiment 1b, only the two extreme conditions (audio-word [simplest] and video-sentence [most complex]) were included, but ROCs were examined under conditions in which overall performance was matched to determine whether differences in ROC shape could be attributed to differences in performance.

**Method**

**Experiment 1a.** Ninety-six students (average age = 19 years; 72 women, 24 men) from the University of California, Davis, participated in Experiment 1a for course credit. Twenty-four participants were randomly assigned to each of the four conditions: audio-word, audio-sentence, video-word, and video-sentence.

**Materials and procedure.** The stimulus pool consisted of 230 words and 230 sentences. The words were all nouns, four to eight letters in length, with an average Kucera–Francis frequency of 51.3. The sentences varied in length and consisted of general trivia, such as “There are more Barbie dolls in Italy than there are Canadians in Canada” and “The hippopotamus has a stomach 10 feet long.” Trivia were collected from Symons (2006), Harvey (1987), and Bacon (1979). Both sets of stimuli were divided into lists of 115 items each to create two counterbalance conditions. Six additional items (of both types) were used as buffers—three for recency and three for primacy.

A man and a woman were filmed reading the sentences and words. The digital video was then edited to create both the audio and the video study lists for each of the four experimental conditions (audio-word, audio-sentence, video-word, and video-sentence). Presentation of stimuli by the two speakers was counterbalanced across participants. One order of stimuli was presented to all participants at study; stimuli were pseudorandomly mixed such that the man or the woman read no more than three items in a row. Words presented for test were newly randomized for each participant.

**Procedure.** In the study phase, participants were instructed to attend to the speaker and what he or she said in preparation for a later memory test on which speaker said which word or sentence. The audio and video word lists lasted approximately 10 min, whereas the sentence lists (audio-sentence and video-sentence) lasted approximately 25 min. Duration of the stimuli varied because of different speaking rates and stimulus lengths, but there was a standard 1-s inter-stimulus interval between all stimuli. All study items were tested and were presented as text on a computer screen such that there was no perceptual match between study and test for any of the conditions. Items (words or sentences) were presented one at a time on a computer screen, and participants indicated whether they remembered the man or the woman presenting the item using a 6-point confidence scale; 1 represented high confidence that the item had been presented by the man, and 6 represented high confidence that the item had been presented by the woman. Participants were instructed to use the entire range of response confidence, and they completed the test at their own pace.

**Experiment 1b.**

**Participants.** Thirty-two students (average age = 19.5 years; 23 women, 9 men) from the University of California, Davis, participated for course credit.

**Procedure.** The materials and procedures were the same as those of Experiment 1, with the exceptions that (a) only the extreme conditions from Experiment 1a were tested (audio-word and video-sentence), (b) the manipulation was within-subjects, and (c) in an attempt to equate performance across the conditions, audio-word stimuli were each presented twice (video-sentence stimuli were presented only once as in Experiment 1a). Repetitions of audio-word stimuli were pseudorandomly ordered such that no word was repeated immediately. Participants completed the audio-word condition first followed by the video-sentence condition; this order ensured that participants could not visualize the man and woman in the videos while encoding in the audio-word condition.

**Results**

**Experiment 1a.** An analysis of variance of overall recognition performance (indexed by $d'_r$, a measure of discrimination analogous to $d'$ but appropriate for both symmetrical and asymmetrical ROCs; see Table 1) showed both a main effect of stimulus complexity, $F(1, 95) = 6.3, p < .02$, and a main effect of source complexity, $F(1, 95) = 12.16, p < .002$, but no significant interaction, $F(1, 95) = 1.5, p = .229$, indicating that increases in stimulus complexity and source complexity led to increases in overall source memory performance.

The average ROCs and $z$ROC are presented in Figure 1. The figure shows that increasing stimulus complexity (moving from the circles to the triangles) led the ROCs to become more curved (and conversely led the $z$ROC to become flatter). In addition, increasing source complexity (moving from the two lower to the two upper curves) led the ROCs to become more curved (and led the $z$ROC to become flatter). To quantify these effects, we examined the shape of the ROCs (see Figure 1) in three complementary ways. First, the DPSD model was fit to each participants' ROCs using maximum likelihood estimation to obtain estimates of recollection and familiarity. Because recollection is modeled as a threshold process and familiarity is modeled as a signal detection process, these parameters can also be interpreted as measurements of ROC shape, with familiarity providing an estimate of the curvilinearity and recollection providing an estimate of the linearity of the ROC. The parameter estimates for each condition are shown in Figure 2. The recollection estimates did not differ across conditions, $F(3, 95) = 1.77, p = .16$. The familiarity estimates, on the other hand, showed both a main effect of stimulus complexity.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Overall Recognition Performance ($d'_r$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment</td>
<td>Audio word</td>
</tr>
<tr>
<td>1a</td>
<td>0.61 (0.46)</td>
</tr>
<tr>
<td>1b</td>
<td>1.22 (0.67)</td>
</tr>
</tbody>
</table>

Note. Mean sensitivity ($d'_r$) for encoding conditions in Experiments 1a and 1b (standard deviations in parentheses).
In our second analysis of ROC shape, we compared the fit of the traditional unequal variance signal detection (UVSD) model with the fit of the same model with an added threshold parameter ("UVSD + R") for the two extreme conditions (audio-word and video-sentence). The models were fit using maximum likelihood estimation and goodness of fit was measured using the log-likelihood statistic $G^2$. It was expected that the addition of the extra parameter in the UVSD + R model would result in a better fit (and thus a lower $G^2$) for both conditions. If, however, the video-sentence ROC is more curvilinear than the audio-word ROC, then the audio-word condition should obtain a greater benefit from the addition of the threshold component than the video-sentence condition. This was the pattern that was found: the decrease in $G^2$ values with the addition of the threshold parameter to the model was greater in the audio-word condition than in the video-sentence condition (audio-word = 9.33 and 3.97; video-sentence = 5.35 and 3.30; for UVSD and UVSD + R, respectively; Mann–Whitney U test, $z = -2.18, p < .03$).

In the third analysis of shape, $z$ROCs (see Figure 1) for each participant were submitted to quadratic regression analyses to obtain an estimate of the quadratic coefficient, an estimate of the (U-shaped) curvature present in the $z$ROC. Because most source $z$ROCs are U-shaped, we expected the quadratics to be greater than zero and, in addition, that the complexity manipulation would affect the size of the coefficient, with simpler conditions having larger quadratics than complex conditions. The quadratic coefficients were greater than zero in all conditions: audio-word, $t(21) = 6.08$; video-word, $t(21) = 4.28$; audio-sentence, $t(21) = 2.14$; video-sentence, $t(22) = 6.21$; all ps < .05. An analysis of variance indicated that there was an interaction between stimulus complexity and source complexity, $F(1, 85) = 13.23, MSE = 0.05$.

Figure 1. Receiver operating characteristics (ROCs; left panels) and $z$-transformed ROCs ($z$ROCs; right panels) for Experiments 1a and 1b.

\[ F(1, 95) = 6.62, p < .02, \text{ and a main effect of source complexity,} \]
\[ F(1, 95) = 6.95, p < .02, \text{ such that the familiarity (curvilinearity) of the ROCs was greater when complexity increased. The interaction was not significant} (F < 1). \text{ Thus, the analysis indicated that both stimulus and source complexity led the source ROCs to become significantly more curvilinear.} \]

\[ F(1, 95) = 6.02, \] and a main effect of source complexity, \[ F(1, 95) = 6.95, p < .02, \text{ such that the familiarity (curvilinearity) of the ROCs was greater when complexity increased. The interaction was not significant} (F < 1). \text{ Thus, the analysis indicated that both stimulus and source complexity led the source ROCs to become significantly more curvilinear.} \]

In our second analysis of ROC shape, we compared the fit of the traditional unequal variance signal detection (UVSD) model with the fit of the same model with an added threshold parameter (“UVSD + R”) for the two extreme conditions (audio-word and video-sentence). The models were fit using maximum likelihood estimation and goodness of fit was measured using the log-likelihood statistic $G^2$. It was expected that the addition of the extra parameter in the UVSD + R model would result in a better fit (and thus a lower $G^2$) for both conditions. If, however, the video-sentence ROC is more curvilinear than the audio-word ROC, then the audio-word condition should obtain a greater benefit from the addition of the threshold component than the video-sentence condition. This was the pattern that was found: the decrease in $G^2$ values with the addition of the threshold parameter to the model was greater in the audio-word condition than in the video-sentence condition (audio-word = 9.33 and 3.97; video-sentence = 5.35 and 3.30; for UVSD and UVSD + R, respectively; Mann–Whitney U test, $z = -2.18, p < .03$).

In the third analysis of shape, $z$ROCs (see Figure 1) for each participant were submitted to quadratic regression analyses to obtain an estimate of the quadratic coefficient, an estimate of the (U-shaped) curvature present in the $z$ROC. Because most source $z$ROCs are U-shaped, we expected the quadratics to be greater than zero and, in addition, that the complexity manipulation would affect the size of the coefficient, with simpler conditions having larger quadratics than complex conditions. The quadratic coefficients were greater than zero in all conditions: audio-word, $t(21) = 6.08$; video-word, $t(21) = 4.28$; audio-sentence, $t(21) = 2.14$; video-sentence, $t(22) = 6.21$; all ps < .05. An analysis of variance indicated that there was an interaction between stimulus complexity and source complexity, $F(1, 85) = 13.23, MSE = 0.05$.

---

2 Quadratic coefficients were first submitted to outlier analyses; coefficients that were outside the bounds defined by the mean plus or minus two standard deviations were not included in further inferential analyses.
The critical pattern of results from Experiment 1a was replicated: The average ROC for the complex condition (video-sentence) was more curvilinear than the ROC for the simple condition (audio-word). This pattern was found (in reverse) in z-space as well: the zROC for the complex condition was more linear (less U-shaped) than in the simple condition (see Figure 1). Importantly, Experiment 1b results demonstrate that these differences are not attributable to differences in performance level.

Combined, Experiments 1a and 1b indicate that increasing the complexity of the studied item or increasing the complexity of the source information can lead to more curvilinear ROCs and flatter zROCs. Clearly both stimulus and source complexity can affect the shape of ROCs, but the most dramatically curved ROCs were found when both item and source complexity increased together. Increasing complexity generally led to an increase in overall performance in Experiment 1a, particularly when both stimulus and source complexity increased together. Experiment 1b addressed this issue by examining the effects of complexity while controlling for levels of overall performance and showed the same pattern: The simple condition produced relatively linear ROCs, whereas the complex condition produced curved ROCs.
Experiment 2

Experiment 1 showed that complex study events can lead source ROCs to become more curved, but they do not indicate why differences in complexity produce differences in ROC shape. It could be that source memory judgments in complex conditions rely more on familiarity if the richer information allows the participants to encode the item and source in an integrated manner. Previous work has indicated that familiarity can support source and associative recognition discriminations if the item and source information are unitized or integrated during study (e.g., Diana et al., 2008, 2010; Quamme, Yonelinas, & Norman, 2007; Yonelinas, 1999). Thus, encoding sentences from a video of the speakers may provide a richer study event that more readily allows an integration of the item and source information, thereby increasing the utility of familiarity in source tests.

Another possibility is that the increase in the curvilinearity of the ROC might occur because increasing complexity causes recollection to behave in a more continuous and Gaussian manner (Qin et al., 2001; Yonelinas, 1999). For example, even if the retrieval of each detail is a threshold process (it may simply fail), the culmination of many such details may give rise to graded recollection for the complex event (e.g., Malmberg, 2002). If so, then recollection may influence ROC shape differently depending on the complexity of the stimuli.

To test these alternative accounts, we examined the effects of divided attention (Experiment 2a) and amnesia (Experiment 2b) on source discriminations for the simple and complex source conditions from Experiment 1b. In Experiment 2a, we examined divided attention at test because previous work has indicated that recollection is particularly sensitive to the disruptive effects of divided attention (e.g., Grupposo, Lindsay, & Kelley, 1997; for a review, see Yonelinas, 2002). Thus, if the differences seen in the ROCs in the previous experiment were due to differences in the nature of recollection, but the contribution of recollection was the same, then dividing attention should reduce performance equally in the simple and complex conditions. In contrast, if the exaggerated curve seen in the complex condition was due to an increased reliance on familiarity in that condition, then dividing attention should have a smaller effect in the complex condition than in the simple condition.

We applied similar logic in Experiment 2b; we compared the performance of hypoxic patients with that of healthy controls on the simple and complex source tasks. Previous work has indicated that patients suffering mild hypoxic episodes exhibit specific deficits in recollection (e.g., Yonelinas et al., 2002; for evidence that severe hypoxic patients can exhibit deficits in both processes, see Manns, Hopkins, Reed, Kitchener, & Squire, 2003; Yonelinas et al., 2004). If recollection contributes equally to the simple and complex source tasks, then the patients should show similar deficits relative to controls on the two tasks. In contrast, if performance on the complex source task is supported by familiarity, then the patients should show less of a deficit on the complex than the simple source task.

Method

Experiment 2a: Divided attention.

Participants. Sixty-four students (average age = 19.3 years; 47 women, 17 men) from the University of California, Davis, participated for course credit. Thirty-two participants were randomly assigned to the full attention (FA) condition, and 32 participants were randomly assigned to the divided attention (DA) condition.

Materials and procedure. Each participant completed the audio-word task and then the video-sentence task from Experiment 1b. In the DA condition, a digit-monitoring task was added during the test phase of each task. The digit-monitoring task consisted of an auditory presentation of single digits ranging from 1 to 9 in a random order, with a 1-s interval between digits. Two different orders were created—one for each test phase. The participants’ task was to monitor the series of digits and say “odd” out loud whenever two odd digits were presented consecutively. The experimenter monitored the performance of each participant and hit a desk bell whenever an error was made.

The confidence scale was eliminated from the test phase in both tasks; participants simply pressed a key labeled “man” or a key labeled “woman” for each word or sentence presented. In the DA condition, each response timed out after 5 s if the participant did not respond. This was done to minimize task-switching between the memory test and the digit-monitoring task. Participants were instructed to treat both tasks as important and to try to make as few mistakes as possible. Before each test phase, participants were given a short practice session on the digit-monitoring task.

Experiment 2b: Amnesia.

Amnestic patients. Two amnesic patients were recruited for participation and were paid for their time. Both patients showed normal intelligence and attention scores but exhibited memory impairments, primarily with verbal materials (see Table 2). Patient 1 was a survivor of sudden cardiac arrest and so was expected to

Table 2

Demographic Information and Standardized Test Results for Amnesic Patients

<table>
<thead>
<tr>
<th>Patient</th>
<th>Sex</th>
<th>Age (years)</th>
<th>Edu (years)</th>
<th>IQ</th>
<th>Att</th>
<th>Ver</th>
<th>Vis</th>
<th>Gen</th>
<th>Delayed</th>
<th>Logical</th>
<th>Rep</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M</td>
<td>49</td>
<td>16</td>
<td>96</td>
<td>-0.20</td>
<td>-2.53</td>
<td>2.00</td>
<td>-1.40</td>
<td>-1.13</td>
<td>-3.07</td>
<td>0.73</td>
</tr>
<tr>
<td>2</td>
<td>M</td>
<td>24</td>
<td>16</td>
<td>111</td>
<td>0.20</td>
<td>-1.33</td>
<td>0.33</td>
<td>-0.87</td>
<td>-2.13</td>
<td>-3.27</td>
<td>-1.53</td>
</tr>
</tbody>
</table>

Note. Age = years of age at testing; Edu = years of education; IQ = estimate of Wechsler Intelligence Scale (Wechsler, 1987); WMS–R = Wechsler Memory Scale—Revised (Wechsler, 1987); Att = attention/concentration; Ver = verbal memory; Vis = visual memory; Gen = general memory; Delayed = delayed memory; Logical = delayed logical memory; Rep = delayed visual reproduction; M = male.

* All scores presented are z-scores.
have bilateral damage relatively limited to the hippocampus as a result of cerebral hypoxia (Rempel-Clower, Zola, Squire, & Amaral, 1996); however, because of a defibrillator, the patient could not be scanned. Patient 2 suffered a traumatic head injury acquired in a motorcycle accident and was in a coma for 1 month. A structural magnetic resonance image revealed no apparent brain damage, suggesting that his memory impairment was hypoxia-related.

Healthy controls. Nine healthy control participants were recruited for participation and were paid for their time. The average years of education (M = 16.6 years) and the average age (M = 38.8 years) of the controls were similar to that of the amnesics (M = 16 years, and M = 38.5 years, respectively).

Materials and procedure. The materials and procedures were identical to those used in the Experiment 2a FA condition except that participants were instructed to use a strategy during the study phase that involved thinking about the feminine or masculine aspects of a word or sentence when trying to link stimuli to the female or male speaker. The strategy instruction was added to reduce the variance between participants, because debriefing in previous studies suggested that some participants tried to use encoding strategies, but others did not. Finally, participants in Experiment 2b made confidence responses on a 6-point scale at test.

Results and Discussion

Experiment 2a: Divided attention. Figure 3 shows d' averages for the two tasks in both the FA and DA conditions. Dividing attention reduced source memory performance and had almost identical effects on the simple and complex tasks (i.e., leading to a reduction in d' of 0.51 and 0.53 in the simple and complex conditions, respectively). The decrease in performance because of divided attention was significant, \( F(1, 62) = 17.2, p < .001 \); however, \( d' \) remained above 0 in both tasks: audio-word, \( t(31) = 4.51, p < .001 \); video-sentence, \( t(31) = 6.8, p < .001 \). Results across the simple and complex memory tests did not differ significantly, and there was not a significant interaction between test type and attention condition (both \( F_s < 1 \)).

The similarity in the effects of divided attention on the simple and complex conditions is most consistent with the idea that recollection was supporting source decisions to the same degree in the two conditions. Had the increase in curvature found in the complex condition ROCs been due to familiarity, performance in that condition should have been less vulnerable to the effects of divided attention. Thus, the patterns in Experiment 2a suggest that differences in the complexity or richness of the stimuli may affect the behavior of recollection such that it acts in a probabilistic manner with relatively simple stimuli but becomes more signal-detection-like with more complex stimuli.

Experiment 2b: Amnesia. Recognition performance for patients and controls is shown in Figure 4. Overall performance of the two patients was significantly lower than that of the control group in each condition, \( F(1, 8) = 16.8, p < .004 \), but there were no differences in performance as a function of condition (\( F_s < 1 \)). This pattern of results suggests that the combination of recollection and familiarity that supports source memory in the simple condition is about the same as that which supports source memory in the complex condition.

Because there were only two patients, statistical analyses were limited to the overall recognition scores (above). However, visual inspection of the ROCs and \( z \)ROCs (see Figure 5) supports the conclusions based on the \( d' \) analysis that the amnesic patients' deficits were comparable in the two conditions, thus suggesting that the tasks were equally dependent on recollection. DPSD model estimates were also consistent with this conclusion, showing no evidence that the amnesics were able to rely more on familiarity in the video-sentence condition than in the audio condition (see Figure 5).

Overall, the patients showed similar impairments in the simple and complex conditions. These data are consistent with the claim that recollection is driving performance in both conditions despite the difference in ROC shapes, thus supporting the findings of Experiment 2a.

General Discussion

The aim of the current study was two-fold: to determine whether increasing the complexity of stimuli in a source memory experiment would lead the ROCs to become more curved (and the \( z \)ROCs to become flatter), and if so, whether those changes in shape were due to recollection, familiarity, or both. We varied complexity by mimicking the extreme conditions found in the literature that seem to have been associated with ROCs of different shapes: Our simplest condition (audio-word) matched the typical source conditions that are found to produce fairly linear ROCs, and our most complex condition (video-sentence) matched the richer source conditions found in Qin et al.’s (2001) study. The results of Experiment 1a indicate that both stimulus complexity and source complexity increased the curvilinearity of the ROCs (and decreased the U-shape of \( z \)ROCs), Experiment 1b showed that the effects of complexity on ROC shape remained even when overall performance was matched, thus ruling out the possibility that the increased ROC curve was an artifact produced by high levels of performance. Experiments 2a and 2b examined whether the change in shape was due to changes in recollection or familiarity, by examining the effects of dividing attention and mild hypoxia—two factors known to undermine recollection.

Recognition perfor-
mance in the simple and complex conditions was similarly affected by both of these factors, suggesting that complexity influenced recollection rather than familiarity. Together the results of the current experiments demonstrate that increasing complexity leads to an increased ROC curve (i.e., a flattened zROC) and suggests that this influence on shape is most likely due to changes in the nature of recollection, rather than in the extent to which familiarity contributes to performance.

The debate over the shape of source ROCs has been fueled in part by the mix of findings ranging from very linear to very curvilinear ROCs. The current study addresses this debate by examining one factor that may have led to some of the discrepancies in previous studies. The results make it clear that the shape of ROCs in source memory studies is dependent on the complexity of the stimuli. When the source and stimuli were simple, we observed a fairly linear ROC that was highly U-shaped in z-space, but increasing the complexity of either the source or the stimulus led to more curvilinear ROCs and flatter zROCs. Thus, in addition to factors such as unitization (Diana et al., 2008, 2010) and item strength differences (Yonelinas, 1999), it is now clear that increasing complexity can lead to curved source ROCs.

Unlike unitization and strength differences though, which have been shown to affect ROC shape by increasing the usefulness of familiarity in making source decisions, complexity appears to influence the shape of source ROCs by changing the nature of recollection. If the differences in ROC curvature found in Experiments 1a and 1b had been due to greater familiarity in the complex condition, then dividing attention at test in Experiment 2a should have affected performance in the simple condition more than in the complex condition. In fact though, dividing attention had the same effect on performance in the two conditions. Thus, the results suggest that contributions of recollection and familiarity were the same regardless of complexity and that the differences in ROC curvilinearity in Experiment 1 were due to differences in the behavior of recollection. In addition, the amnesic patients in Experiment 2b who were expected to have severe recollection impairments performed similarly in the simple and complex conditions, suggesting that contributions of recollection and familiarity were similar in the two conditions. Together, these results suggest that recollection generally leads to linear ROCs, but that when materials become highly complex, recollection can produce more curved ROCs. One might argue that the failure to find effects of divided attention and amnesia on ROC shape reflected a lack of statistical power. However, both of these manipulations had sizable effects on overall recognition performance, and there was no indication in the results that the complexity manipulation influenced familiarity. Future studies using convergent methods to verify these conclusions will be useful, but the current results certainly favor the graded recollection account of complexity over a familiarity account of the effect.

These results present a challenge for models, such as the DPSD model (Yonelinas, 1994), that assume that recollection is a threshold process that predicts linear ROCs. Importantly, however, even though complexity was found to lead the ROCs to become more curved, they were inconsistent with pure signal detection models as well. That is, although the zROCs became more linear with increased complexity, they still retained a U-shape across all four conditions, which is inconsistent with the linear zROCs predicted by signal detection theory.

**Why Does Complexity Lead Recollection to Produce Curved ROCs?**

To answer this question, it is useful to examine the response distributions for the simple and complex conditions. Figure 6 shows the distribution of confidence responses for simple and complex conditions when overall performance is matched (i.e., Experiment 1b). Interestingly, the proportion of correct and incorrect high confidence source responses (i.e., 1s and 6s) remained constant across simple and complex conditions, suggesting that complexity did not alter the number of high confidence source recollections. Rather, complexity increased the number of source responses that were associated with the second-to-highest level of confidence (the 5s and 2s, demarcated by arrows in Figure 6) and more importantly, the proportion of correct-to-incorrect responses at this level shifted to favor the correct source. However, because performance was still better for the highest confidence responses than for the 5 and 2 responses, performance was graded across the scale, which was necessary to produce the curved ROCs. Thus,
complexity appears to have led to an increase in the recollection of lower confidence and less diagnostic source information.

It is not clear why complexity led to an increase in less diagnostic source recollection, but we suspect that it might be due to an increase in the numbers of features and associations available in the stimuli. That is, in the complex condition, there are many different aspects of the item and the source that the participants might attend to, and some of these aspects might not be entirely diagnostic of source (e.g., “the sentence is about animals and I remember that the male talked a lot about animals”). Information that is only weakly diagnostic of source may lead to an increase in the proportion of items attributed to a source with lesser confidence. A related possibility is that in the complex condition, participants may be more likely to remember different numbers of associations that can be useful in making the source discriminations, such that they remember two details on one trial (voice and visual information) but only one on another trial (voice). Confidence levels that differ as a function of number of details or associations recalled could lead to the observed pattern of results, if it is the case that details are falsely recollected in a similar way. Future studies will be necessary to further test these accounts.

How Can Current Models of Recollection Account for These Effects?

As discussed earlier, the DPSD model does not account for the effect of complexity because it assumes that recollection produces linear ROCs (U-shaped zROCs). Moreover, recollection cannot be a pure signal detection process, because as with many previous studies, the source zROCs in the current experiments were consistently U-shaped. One potential solution is to use models that integrate the characteristics of these two classes of models. For example, the variable recollection dual process (VRDP) model (Sherman, Atri, Hasselmo, Stern, & Howard, 2003) is based on the DPSD model, but it assumes that the above-threshold items (i.e., recollected items) form a Gaussian recollection strength distribution (for similar models, see DeCarlo, 2003; Kelley & Wixted, 2001; Onyper, Zhang, & Howard, 2010). The DPSD model simply assumes that above-threshold recollection supports high confidence responses, and it does not specify the shape of the recollection strength distribution. Threshold strength distributions are sometimes illustrated as being square, but they can take on any number of different shapes (e.g., see Macmillan & Creelman,
average ROCs in Experiment 2 to illustrate how such an approach sets. For example, when the model is fit to ROC data, there are it tends to be underconstrained by the current type of ROC data model. Because the distribution (VR). When $R_f$ is large or when $V_R$ is small, the model collapses into the standard two-parameter model. Because of this, our approach was to make some simplifying assumptions to fit the average ROCs in Experiment 2 to illustrate how such an approach could in principle account for the effects of complexity. We leave a full assessment of the model for future studies.

Figure 7 illustrates how the VRDP model could account for the results of Experiment 1b. In the simple condition, familiarity is not very useful for source discriminations, and so the two familiarity strength distributions (i.e., the “male” items and “female” items) are largely overlapping and centered quite close to the 3 and 4 guessing responses. Recollection is expected to be more useful in supporting source memory discriminations, so the recollection distributions are spaced further to the left and right. The recollection distributions are illustrated as having less variance than the familiarity distributions, and as falling slightly above the 6 response criterion, but we could have chosen to represent them as having much greater variance, accompanied by much higher memory strength. In either case, if one assumes that participants make their responses on the basis of recollection and that they respond on the basis of familiarity when recollection fails, then these strength distributions would produce the observed response distributions seen in the simple condition in Figure 6.

To account for data in the complex condition, the VRDP model either has to increase the variance of the recollection distribution or decrease the average strength of the recollection distribution (or both). The functional effect, regardless of the way this is achieved, is that more of the recollection distribution falls below the highest confidence level than in the simple condition. We illustrate a combination of both ways to achieve this effect in the complex condition panel in Figure 7 (i.e., a slight decrease in the strength of the recollection distribution and a slight increase in its variability). In this way, most experiences of recollection would still support high confidence source responses, but a proportion of recollection responses would now be associated with lower levels of confidence. Mixing the recollection and familiarity distributions would produce the empirical response distributions in the complex condition in Figure 6. Note that the decision rule governing responses is the same as that underlying most dual process models whereby recollection is used as a basis for responding, and familiarity is used when recollection fails (e.g., Jacoby, 1991; Mandler, 1980; Yonelinas, 1994). However, the model differs from the DPSD model in the sense that it allows for the possibility that there may be items for which familiarity is more diagnostic of source than is recollection. That is, recollection might support only a 5 response, whereas familiarity could support a 6 response. Whether this occurs in practice is not known, but if the model is correct, it should be possible. Overall, the VRDP model can account for complexity by assuming that the recollection distribution falls largely within the highest confidence bin in the simple condition, but that increasing complexity produces a weaker and more variable recollection distribution. This change in the recollection distribution results in more recollected items receiving lower confidence ratings than in the simple condition, which in turn produces more curved ROCs (or flatter zROCs).

The account that we have provided for the effects of complexity on ROC shape is similar to an account provided for the effects of another variable that changes ROC shape: feature overlap. Elfman, Parks, and Yonelinas (2008) reported similar findings in testing the complementary learning systems (CLS) model’s prediction that increases in feature overlap would lead the threshold nature of recollection to break down (Norman & O’Reilly, 2003). When feature overlap was very high amongst stimuli (stimuli were all pictures of suburban houses, for instance), recollection’s typical threshold nature broke down into a more graded process. The complexity manipulation in the current experiments would seem to have a similar effect on the nature of recollection—the more complex the stimuli, the more graded the recollection process appears to be.

To determine whether the CLS model could also account for the current results, we simulated the simple and complex conditions investigated in the current study with the model to test whether it would accurately predict that increasing complexity would result in more curved source ROCs. The CLS model uses biologically constrained parameters to simulate medial temporal lobe functioning. It consists of a recollection component that is based on the hippocampus, and a familiarity component that is based on the surrounding cortex. The hippocampal component of the model is specialized to rapidly encode associations between various features of a stimulus (e.g., item–source associations). Thus, when a test item is presented to the network, the hippocampal component can retrieve the associated source information. Under standard conditions, the hippocampal component acts like a threshold process in the sense that it accurately retrieves the studied source, or it fails to provide any source information (i.e., recollection fails). We conducted the simulations with the standard hippocampal component of the model, using parameter values based on previous publications (e.g., Elfman et al., 2008). In the simple condition, an item (“A”) was represented as a string of features (e.g., color, shape, size) that was paired with a second string of features that represented a source (“X”). In the complex condition, each item was made up of multiple components (e.g., AB), and each source was made up of multiple components (e.g., XY). The network then had to learn two pairings (e.g., A–X and B–Y) to encode all the features. This piecewise approach was meant to capture the fact that as complexity increases, the number of specific associations that the participant can form for a given item-source pairing will increase.

$R$ is the probability of recollection, $\Phi$ is the normal distribution function, $c_k$ is the criterion for confidence level $k$, $R_f$ is the strength of the recollection distribution, $R'_{f}$ is the strength of the familiarity distribution, and $FA$ is false alarm.
To match overall performance across conditions, the item–
source pairings in the simple condition were encoded twice each. The source decisions were modeled by presenting an item and allowing the hippocampal network to pattern complete the missing source information. The strength distributions and the derived ROCs for the simple and the complex conditions are presented in Figure 8. For specific modeling details, see Elfman et al. (2008; also see Elfman & Yonelinas, 2011).

The strength distributions produced in the simple condition are consistent with results of previous simulations; namely, the model produces high confidence correct source responses (akin to 1s and 6s), but it also shows a large proportion of retrieval failures (i.e., the middle peak). The resulting ROCs exhibit a large linear component, consistent with the empirical results. Most importantly, however, is that the threshold shaped strength distributions took on a more Gaussian shape in the complex condition, and the resulting ROCs were more curvilinear, as was seen in the empirical study. Notably, the model also correctly predicted that increasing complexity results in increased source performance, despite the breakdown in the threshold nature of recollection (evidenced by the fact that it was necessary to present “simple” items to the network twice to achieve matched performance). It is important to emphasize that the CLS model was not designed to produce the specific types of ROCs observed here, but it was designed to capture the neuroanatomical properties of the hippocampus. Thus, the fact that it was able to produce the observed ROC effects attests to the power of this neuro-computational modeling approach.

What Are the Implications of the Current Results for Current Measurement Models of Recollection and Familiarity?

Our results suggest that when estimating recollection and familiarity in source memory tests, one needs to be cautious when interpreting parameter estimates whenever the materials become highly complex. In the current study, when we used simple stimulus materials, the VRDP model produced parameters that were effectively the same as those produced by the standard DPSD

![Figure 7.](image7.png)

![Figure 8.](image8.png)
model. That is, because recollection led to high confidence responses, the full VRDP model effectively collapsed into the standard DPSD model. Given that most laboratory studies use simple materials like these, it would suggest that the parameter estimates derived using the simple model would not be altered if the more complex model was adopted. A previous study comparing these models led to similar conclusions (Sherman et al., 2003; also see Onyper et al., 2010). Thus, we recommend using relatively simple materials when using the DPSD model in studies of source memory. When complex stimuli are of interest, it would be best to keep the source as simple as possible, given the finding that complexity effects were much larger when both item and source were complex. Alternatively, one could adopt a more complex model, such as the VRDP model, and use that to estimate recollection and familiarity. Although there are important issues with identifiability that will need to be addressed with these models, we feel that this approach is very promising.

The effect of complexity on recollection and familiarity in item recognition tests has not been formally assessed, and it will need to be examined in future studies. However, to date, complexity does not appear to affect item recognition in the same way as it was found to affect source recognition. That is, the ROC method and the DPSD model have produced estimates that converge with those from other measurement procedures for complex stimuli such as faces, sentences, and scenes (see Yonelinas, 2002, for a review), which one would not expect if the ROC estimates were biased by complexity. Moreover, in preliminary studies, we have not found any noticeable effects of stimulus complexity on standard item recognition ROCs.

Overall, the data demonstrate that recollection can act simultaneously as a threshold and as a graded process and that at least one critical factor determining this behavior is the complexity of the stimuli. Clearly, understanding memory for a range of stimuli from the simplest to the most complex is important, but if considered separately, they may well lead to different conclusions about the nature of the processes supporting recognition. The effects of complexity shown here along with those of Elfman et al. (2008) indicate that the behavior of recollection at the level of events is a variable feature that is dependent at least in part on the nature of the stimuli being remembered.

References


Received March 3, 2010
Revision received October 8, 2010
Accepted October 30, 2010
Chapter 4: Recollection and Familiarity Exhibit Dissociable Similarity Gradients: A Test of the Complementary Learning Systems Model

**Authors:** Kane W. Elfman, Dr Andrew P. Yonelinas (PI)

**Journal:** Journal of Cognitive Neuroscience

**Date of publication:** May, 2015

**Abstract:** Memory can often be triggered by retrieval cues that are quite different from the originally encoded events, but how different memory processes respond to variations in cue-target similarity is poorly understood. We begin by presenting simulations using a neurocomputational model of recognition memory (i.e. the complementary learning systems model), which proposes that the hippocampus supports recollection of associative information whereas the surrounding cortex supports assessments of item familiarity. The simulations showed that increases in the similarity between retrieval cues and learned items led to relatively linear increases in a cortex-based memory signal, but led to steeper and more thresholded increases in the hippocampal signal. We then tested the predictions of the model by examining the effects of varying cue-target similarity in two recognition memory experiments in which participants studied a list of computer-generated faces and then, at test, gave confidence and remember/know responses to morphed faces. In both experiments, as cue-target similarity was increased, familiarity-based recognition increased in a gradual and relatively linear fashion, whereas recollection showed significantly steeper gradients. The results show that recollection and familiarity exhibit distinct similarity functions in recognition memory that correspond with predicted retrieval dynamics of the hippocampus and cortex, respectively.

**Author contributions:** The research and writing of this paper was conducted by myself under the supervision of principal investigator Dr Andrew Yonelinas. The research included designing and implementing neural network model simulations of the hippocampus and association cortex. The simulations were analyzed on a trial-by-trial level and at the aggregate level. The two behavioral experiments were also designed and implemented by myself. Research assistants were also used to aid data collection during participant sessions, which I supervised. I analyzed the
experimental data using advanced nonlinear regression modeling techniques that are described further in the methods chapter of this dissertation. The original draft for the paper was written entirely by myself and then revised with Prof Yonelinas in preparation for journal submission.
Recollection and Familiarity Exhibit Dissociable Similarity Gradients: A Test of the Complementary Learning Systems Model

Kane W. Elfman and Andrew P. Yonelinas

Abstract

Memory can often be triggered by retrieval cues that are quite different from the originally encoded events, but how different memory processes respond to variations in cue–target similarity is poorly understood. We begin by presenting simulations using a neurocomputational model of recognition memory (i.e., the complementary learning systems model), which proposes that the hippocampus supports recollection of associative information whereas the surrounding cortex supports assessments of item familiarity. The simulations showed that increases in the similarity between retrieval cues and learned items led to relatively linear increases in a cortex-based memory signal but led to steeper and more thresholded increases in the hippocampal signal. We then tested the predictions of the model by examining the effects of varying cue–target similarity in two recognition memory experiments in which participants studied a list of computer-generated faces and then, at test, gave confidence and remember/know responses to morphed faces. In both experiments, as cue–target similarity was increased, familiarity-based recognition increased in a gradual and relatively linear fashion, whereas recollection showed significantly steeper gradients. The results show that recollection and familiarity exhibit distinct similarity functions in recognition memory that correspond with predicted retrieval dynamics of the hippocampus and cortex, respectively.

INTRODUCTION

Our ability to retrieve memories for past events in response to different environmental cues is fundamental to how we make sense of and interact with the world. However, the features constituting a retrieval cue rarely match the encoded objects or events perfectly. For example, we may recognize that we have met a person before even if they have had a haircut since our initial encounter or if the lighting conditions have dramatically changed the appearance of their face. A core question about how memory operates then is how similar does a retrieval cue have to be to the original item before we are able to recognize that we have encountered it previously? Here, we test predictions based on a neurocomputational memory model that indicates that increases in cue–target similarity lead to gradual increases in familiarity but lead to steep increases in recollection once a recollective threshold is exceeded.

Many current models of long-term memory posit two underlying processes that serve fundamentally different yet complementary roles. For example, several recognition models differentiate between recollection, which is the retrieval of qualitative information about a study event, and familiarity, which reflects the global match between a cue and what was previously learned (Eichenbaum, Yonelinas, & Ranganath, 2007; Brown & Aggleton, 2001; Yonelinas, 1994; Jacoby, 1991; Mandler, 1980; Atkinson & Juola, 1974).

Neurocomputational models such as the complementary learning systems (CLS) model posit a similar distinction based on neurophysiological evidence that postulates separate memory roles for the hippocampus and surrounding medial-temporal lobe cortex (MTLc; e.g., Norman & O’Reilly, 2003; O’Reilly & Rudy, 2001; McClelland, McNaughton, & O’Reilly, 1995; for similar ideas, also see Sherry & Schacter, 1987; O’Keefe & Nadel, 1978; Marr, 1971). In these models, the hippocampus is assumed to encode associations between various aspects of a single event, such that when a subset of those features is subsequently presented, it is able to pattern complete and retrieve those missing features. Importantly, the encoded representations are highly pattern-separated (i.e., nonoverlapping), and as a consequence, cues that do not closely match the original item or event will rarely trigger recollection (O’Reilly & McClelland, 1994). In contrast, the MTLc is thought to strengthen, via Hebbian learning, associations that are common across many different events and, therefore, forms more overlapping, generalized representations. In this way, cues that are repeated or similar to other encoded stimuli are processed or identified more readily than novel stimuli, and this can be used as a measure of recency or stimulus familiarity.
If, as these computational models predict, the hippocampus forms more pattern-separated representations of events than the cortex, then we can expect that the two systems should respond very differently to changes in cue–target similarity. That is, we hypothesized that, at low levels of similarity (i.e., when tests cues are very different from their corresponding studied items), the hippocampus will rarely pattern complete the studied item, whereas at high levels of similarity it should do so reliably. Thus, if we were to plot performance as a function of cue–target similarity, we would expect to see a non-linear trend characterized by a steep gradient at the point at which pattern completion becomes viable (for similar ideas, see Yassa & Stark, 2011). In contrast, the MTLc, which is predicted by the CLS model to produce a global match signal (Norman & O’Reilly, 2003), is expected to behave in a more linear manner with discriminability increasing more gradually with cue–target similarity.

The similarity functions produced by the hippocampal and cortical components of the CLS model have not yet been directly examined, and little is known empirically about how recollection and familiarity each responds to variations in cue–target similarity. However, evidence from human fMRI studies provides some relevant evidence. For example, Bakker, Kirwan, Miller, and Stark (2008) found that hippocampal subfield CA3 together with the dentate gyrus (DG) responds very differently to repeated items relative to slightly altered items, indicating a steep similarity gradient. Similarly, Lacy, Yassa, Stark, Muftuler, and Stark (2011) identified noncontinuous transitions in the same subregion as retrieval cues were incrementally varied between studied and nonstudied items. And recently, an examination of neural similarity measures has indicated that, within the hippocampus, subsequent memory is predicted by greater pattern distinctiveness, whereas in the surrounding MTLc, subsequent memory is predicted by greater across-item pattern similarity (LaRocque et al., 2013).

Whether recollection exhibits a steeper similarity gradient than familiarity has not been directly tested; however, there is some indirect evidence to support this hypothesis. For example, in recognition memory studies, false recognition of nonstudied items has been shown to occur very rarely when recognition responses are accompanied by reports of conscious recollection (i.e., “remember” reports) but is quite common for items recognized on the basis of familiarity (i.e., “knowing” reports; see Yonelinas, 2002, for a review). However, under conditions in which the retrieval cue (e.g., the word “sleep”) is semantically related to many of the studied items (e.g., “rest,” “nap,” “tired”), false recognition of the nonstudied retrieval cue often occurs on the basis of recollection (e.g., Norman & Schacter, 1997; Roediger & McDermott, 1995). To the extent that recollection is comparatively more resistant to false recognition when cues are dissimilar to studied items, the evidence suggests that recollection has a steeper, narrower similarity gradient than familiarity.

One recent study examined recollection and familiarity for visual objects using remember/know judgments and source recognition (Kim & Yassa, 2013), in which lures had varying degrees of similarity to studied items. Lures that were highly similar to studied items were more likely to be identified as old and were more often identified on the basis of recollection than familiarity. Interestingly, the number of lures identified as old decreased monotonically with similarity; however, confidence levels were not recorded so it is unclear whether this trend would generalize to subjective confidence reports. In addition, although recollection was reported to have a steeper similarity gradient than familiarity, the gradients were not performance-matched so it is difficult to draw strong conclusions about the shapes of the two gradients relative to one another.

In the current paper, we set out to examine the effects of varying cue–target similarity on recollection and familiarity. We begin by presenting simulations using the CLS model, in which we characterized and contrasted hippocampal and MTLc similarity functions by probing the model with test cues that morphed incrementally from nonstudied to studied items. Our goal was to determine the extent to which the similarity functions of recollection and familiarity should differ based on the CLS model’s predictions and extant evidence suggesting differential involvement of the hippocampal and MTLc structures during recollection and familiarity-based memory performance, respectively. The CLS model was selected because it shares many core assumptions with other memory models, and it has been directly applied to recognition memory (Elfman, Parks, & Yonelinas, 2008; Norman & O’Reilly, 2003).

Following the simulations, the results from two item recognition experiments are presented that examine how recollection and familiarity change as cue–target similarity increases to determine if human memory performance is consistent with the CLS model’s predictions of neural network functioning. Participants studied images of faces, and similarity was varied at test by parametrically morphing between novel and studied faces. In the first experiment, we morphed each face in 10% increments, from 0% to 100% similarity, and participants gave old/new confidence ratings at each step, followed by a remember/know judgment (Tulving, 1985) at the end of the trial. In the second study, we derived estimates of recollection and familiarity using participants’ confidence ratings (Yonelinas, 1994) at different levels of similarity, with each test item shown at only one level of similarity, rather than over a series of morphs.

MODEL SIMULATIONS

The aim of the simulations was to determine the predicted similarity gradients of recollection and familiarity by examining the effects of varying stimulus similarity on
hippocampal and MTLc network signals. Each network was trained on a set of items and then was probed for item memory using retrieval cues that parametrically varied from very different to very similar to studied items. We first examined the similarity gradients of each of the two networks within single trials and then examined the similarity functions of the two networks averaged across items, as is typically done in behavioral studies. Finally, we combined the outputs of the two networks to determine how overall recognition performance would be influenced by stimulus similarly.

Methods
The simulations were implemented using the software package Emergent, version 5 (Aisa, Mingus, & O’Reilly, 2008) that incorporates the Leabra neural network algorithm (O’Reilly & Munakata, 2000). The model’s structure is based on a widely accepted model of hippocampal architecture and a simple approximation of the association cortex (or MTLc; Hasselmo, 1995; O’Reilly & McClelland, 1994; Rolls, 1989). Layers of units represent different anatomical regions, and each unit approximates the behavior of a group of neurons using a continuous sigmoidal activation function with values ranging between 0 and 1. Connection weights were updated using a conditional principal components analysis-based Hebbian learning rule, and competitive inhibition was simulated within layers using the kWTA rule (O’Reilly & Munakata, 2000). Model parameters were based on previous instantiations of the CLS model (Aisa et al., 2008; Elfman et al., 2008; Norman & O’Reilly, 2003; O’Reilly & Munakata, 2000) and are included in Tables A1 and A2.

The hippocampal network comprises entorhinal (ERC) input and output layers, the DG, and layers CA3 and CA1 (see Figure 1, left) and incorporates separate encoding and retrieval modes. In encoding mode, an input pattern (described below) is presented to the network at the ERC input layer. Activation spreads to CA3, both directly and indirectly via the DG as well as to CA1. Because of strong competitive inhibition within each layer, activation throughout the hippocampus is relatively sparse. This is especially true of the DG, which in turn helps to minimize pattern overlap in CA3. CA3 has strong within-layer associations (recurrent collaterals) that help bind together the various features of an event; however, learning occurs throughout the network (with the exception of ERC–CA1 projections which maintain a static mapping; see Norman & O’Reilly, 2003). At retrieval, a cue is presented to the network, and area CA3 attempts to reactivate—with the aid of its strengthened associations—the encoded representation. If successful, activation then spreads to CA1 via weak, diffuse projections. Projections from area CA1 then reinstate the originally encoded pattern at ERC output. Performance was calculated as the net match between the activation pattern at the ERC output layer and the input pattern. We used this measure as it indicates the quality of retrieval—that is, which specific features were reactivated—and not simply the amount of activation. For convenience, we refer to this measure as “recollection,” but we acknowledge that recollection is a psychological construct that is likely supported by the hippocampus but cannot simply be reduced to it.

The MTLc comprises an input layer—identical to the hippocampal input—and an association layer (see Figure 1, right). The input layer projects diffusely to the association layer. When an item is encoded, connections between active units are strengthened whereas other connections are weakened. At the time of test, when a studied item is presented, it is associated with a sharpened representation—that is, a small number of highly active cortical units and a large number of inactive units—compared with nonstudied items. MTLc memory performance, referred to heuristically as “familiarity,” is equated with the sharpness or contrast of an active representation and is measured as the average activation of the 20% most active units, indicating the limit imposed on activation by competitive inhibition (using the kWTA rule; see O’Reilly & Munakata, 2000) and is diagnostic of sharpness, because greater activation of those units is balanced by a net decrease in activation of the remaining units.

The study and test stimuli (i.e., the input patterns) consisted of 36 slots, with four units per slot. A slot represents a feature dimension or stimulus attribute (e.g., category, shape, color) with one active unit per slot. Any two randomly generated patterns overlapped by an average of 25% (referenced as 0 similarity). There were 20 randomly generated, uncorrelated study items for each of the 20 simulated subjects. For the test list, 13 cue patterns were generated for each study item. Each successive cue comprised an incremental change in cue–target similarity, where the target refers to the original study item. The first cue was a randomly generated pattern (0 similarity), and for the next cue, three slots were overwritten to match the target, then three more for the next cue, and so on until the cue and target matched perfectly (a similarity of 1). One progression of test cues—from 0 to 1 similarity—is referred to as a single trial. Figure 2 presents a simplified representation of the stimuli.

Results
Outputs from individual test trials of the hippocampal and MTLc networks for representative studied and new item trials are plotted as a function of similarity in Figure 3A, B. As illustrated in Figure 3A, the hippocampal network produced essentially no recollection signal until a threshold of similarity was reached and then transitioned to a strong recollection state. The specific level of similarity at which a given item transitioned into a recollection state varied across items, and the absolute level of activation reached differed slightly across trials. Retrieval was highly accurate in the sense that, when it occurred, the retrieved representation matched the correct studied
item. Of the 400 studied items, 191 were correctly recollected; of the 209 that were not, 205 were due to non-recollection, whereby no stored pattern was reactivated, and there were only four instances of false recollection, in which nontarget, studied items were retrieved. Correct recollection was operationalized as at least a 67% match between the output and the target pattern and less than 33% erroneous activation; note that because of the threshold behavior of the network, these results were robust to variations in the criteria that was used to define successful recollection.

The similarity function for the MTLc familiarity model is presented in Figure 3B. In contrast to the hippocampus, the MTLc network exhibited relatively linear functions for studied items, such that activation increased gradually as cues became more similar to studied items. For new items, activation remained quite low across the trials. Looking at single trials, the difference in the similarity functions of the hippocampus and MTLc is quite dramatic. Would these differences still be observed when trials are averaged? To assess this, we examined the averaged hippocampal and MTLc similarity functions (Figure 3C).
Averaging across trials, the threshold nature of the recollection function is less obvious because different trials transitioned at different similarity levels. However, the figure shows that the recollection similarity gradient was still steeper than the familiarity gradient. In addition, we found that the gradient was the same even if we increased the number of trials averaged over from 80 per level of similarity to 400. The results suggest that it should be possible to observe differences in the similarity functions of recollection and familiarity even when averaging across trials.

To further characterize the two networks, plots of the frequency distributions of the recollection and familiarity scores—shown here for .25, .5, .75, and 1 cue–target similarity—are presented in Figure 4A, B. These reflect the hypothetical strength distributions of the two networks.
For recollection, at low similarity, there was a large peak at 0, indicating that most trials did not produce retrieval of any learned representation. As similarity increased, the zero-recollection peak diminished whereas a second high-strength recollection peak emerged, indicating accurate recollection. Notably, as the recollection peak grew, its position along the recollection scale did not change, indicating that cue–target similarity did not impact the strength or accuracy of correctly recollected items.

The strength distributions of the cortical model are presented in Figure 4B. In contrast to the hippocampal network, for the cortical network, the mean familiarity scores were Gaussian-shaped, and the variance remained fairly constant as cue–target similarity was varied. As expected, as the test cues became more similar to studied items, the cortical familiarity signal increased (moved to the right).

To determine the effect of similarity on overall recognition (i.e., when both networks were allowed to contribute to performance), we generated predicted receiver operating characteristics (ROCs) by assuming that if hippocampus-based recollection occurs, it will lead to a high-confidence recognition response, whereas if recollection does not occur, recognition is based on familiarity (Yonelinas, 1994, 2001). The predicted ROCs (Figure 4C) were curved and asymmetrical, similar to those observed in human recognition memory studies (for a review, see Yonelinas & Parks, 2007). Additionally, as similarity increased, performance increased and the ROCs moved upward. Importantly, the y intercept (which tracks recollection) first increased slowly and then more quickly as similarity increased, reflecting a nonlinear transition. The point of greatest change in recollection was between similarity measures of .5 and .75. Note that the precise point of steepest transition can vary as a function of various model parameters like learning rate and stimulus dimensions. The important point is that the hippocampal/recollection similarity gradient should be steeper than the cortical/familiarity gradient, but exactly where the difference will be maximal is likely to change with different model parameters.

**Discussion**

The CLS model simulations demonstrated very different similarity functions for the hippocampal and MTLc networks. For individual test trials, the hippocampus produced thresholded recollection functions, in which there was no output activation until a high level of stimulus similarity was reached, at which point the network strongly activated and retrieved the correct study item. In contrast, the activity of the MTLc tracked similarity in a more graded fashion, such that familiarity for old items increased gradually as similarity increased. The differences in the steepness of the recollection and familiarity similarity functions were reduced when averaged across items because of the variability in the location of the recollection threshold for different items, but the average recollection gradient was nonetheless steeper than that produced by familiarity.

Is the difference in the performance of the hippocampal and cortical networks a natural consequence of their neuroanatomical structures, or is it just a particular parameter setting that caused the gradient differences? To answer this question, we systematically searched the model parameter space to see if there was any one parameter or structural property that could account for the observed differences.

![Figure 4](image-url)
in similarity gradients. We could find no evidence that there was any “threshold parameter” or a single structural property of the model that was responsible for the hippocampal/cortical difference we observed. Rather, it appears that the threshold nature of the model is an emergent property of several of the unique aspects of the hippocampal architecture that are not present in the cortex. To explore which model properties are most critical in producing the different types of outputs, we took two approaches. First, we started with the current hippocampal model that produced the threshold output and then systematically reduced or removed parameters that might have been critical for producing the threshold. Second, we started with a cortical model and systematically added parameters or structures to make it more similar to the hippocampal model to see if any of those would lead the model to produce a thresholded output. For both approaches, we examined five core parameters that differentiate the hippocampal and cortical networks, including: (1) recurrent connectivity in CA3, (2) DG detonator synapses activating CA3, (3) lateral inhibition in DG and CA3, (4) projection strength from CA3 to CA1, and (5) learning rate. Each of these parameters had an impact on overall network performance. However, there was no circumstance that we could identify in which decreasing or removing a parameter from the hippocampal model led it to produce a monotonic output like the cortex, nor was there a case in which adding a single parameter to the cortical model led it to produce the threshold pattern observed in the hippocampal model. On the basis of this, we conclude that the threshold and continuous properties of the hippocampal and MTLc, respectively, are emergent properties of the neuroanatomical architecture of these different regions and not simply a difference in any type of a strength or threshold parameter.

On the basis of these simulation results, we predicted that, in humans, comparable similarity functions for estimates of recollection and familiarity should be observed in recognition memory tests. That is, recollection should exhibit a nonlinear function with a steeper gradient than familiarity, familiarity should be comparatively more linear, and this should be evident at both the aggregate and individual trial level.

**HUMAN RECOGNITION MEMORY**

We set out to test whether the contrasting similarity functions observed in the model simulations would be detected in human behavioral tests of item recognition. Participants studied lists of computer-generated face images, and then recognition memory was tested using items that were manipulated to have varying degrees of similarity to studied items. In Experiment 1, each test item was gradually morphed from a generic, nonstudied face to either a studied face or a nonstudied face. Recognition confidence was assessed at each level of stimulus similarity, and then, once the target face was presented, participants made a “remember” response if the face was recollected and a “know” response if the face was recognized as old on the basis of familiarity without recollection. By doing this, it was possible to measure recognition confidence as a function of stimulus similarity, separately for items that were ultimately recollected or familiar. Experiment 2 was similar to Experiment 1, except that each test face was only tested once at a random morph level, and there were no remember/know judgments.

**Experiment 1**

After encoding a list of faces, recognition was tested for a set of faces that were incrementally morphed from a generic, prototypical face to a studied face or a new face (see Figure 5). We expected that, for trials in which participants reported recollecting the face, there would tend to be a large step increase in confidence at some point as the face morphed into an old face. In contrast, for trials in which participants reported only a sense of
familiarity, we expected to see a more graded increase in confidence.

Methods

Participants. Twenty-eight undergraduate psychology students (mean age = 20 years) participated in the experiment for course credit. Four participants were excluded because of poor performance (i.e., $d' < .5$).

Materials. Forty-eight unique faces (see Figure 5) were computer-generated (FaceGen Modeller, version 3.4, 2009; Singular Inversions, Inc., Toronto, Canada). Each face was generated beginning with the same plain prototype that was constructed by centering all available feature parameters. Unique faces were then constructed by pseudorandomly shifting the parameters so that faces would be as distinct as possible, while remaining realistic. The parameters were additionally controlled such that all faces were approximately equidistant to the prototype. The images were 240 × 240 pixels in size and were presented with a black background in the center of the screen.

Design and procedure. The study phase consisted of 24 unique faces presented one at a time. Participants were instructed to try to remember each face for a later test. A face first appeared by itself for 5 sec. Then, with the face still visible, participants were prompted to rate their impression on three attributes using a 4-point scale. The attributes were unpleasant–pleasant, republican–democrat, and shy–outgoing; for example, 1 = very unpleasant, 2 = mildly unpleasant, 3 = mildly pleasant, and 4 = very pleasant. Each scale appeared for 5 sec. There was a 400-msec ISI.

The test phase comprised the 24 studied faces and 24 new faces, presented in random order. Each trial began with the prototype face (the leftmost test cue in Figure 5) and was followed by 10 presentations, incrementally morphing into either a studied face or a new face. Participants were instructed to rate each face on a 9-point scale, ranging from 1 (sure new) to 5 (no idea) to 9 (sure old). Participants were instructed that their first response in each trial should always be “5,” because the first presentation contained no information about whether the face was studied or new. Responses were self-paced, and each image remained on the screen until the participant made a response. At the end of each trial, participants were instructed to rate their memory as “remember” if they remembered studying the face (i.e., if they could retrieve some qualitative information about the event in which the face was initially studied), “know” if the face was only familiar (i.e., the face was studied but they were unable to recollect any qualitative information about the study event), or “new” if they thought it was a new face (Yonelinas, 2001).

Results

Figure 6A presents the normalized mean recognition confidence for old and new faces as cue–target similarity was varied from 0 to 1. Old item performance is plotted separately for items receiving “remember,” “know,” and “new” judgments, whereas new item performance is plotted for the items receiving a “new” response (there were too few false alarms to plot remember or know responses). The solid lines represent sigmoid functions that were fit to the observed data.

An examination of Figure 6A shows that as cue–target similarity increased, recognition confidence increased for studied items that received “remember” and “know” responses. Also, the average confidence of the “remember” responses reached a higher level on average that did the “know” responses. Conversely, for studied items that were not recognized, confidence gradually decreased. A similar pattern can be seen for new items that were correctly recognized as new.

Curve fits. To determine whether the shapes of the similarity functions were different for “remember” and “know” responses, we fit the observed responses to logistic (sigmoid) functions. Participants’ confidence values were converted to $z$ scores to account for individual differences in response bias. Group-level trends were then plotted for each process judgment (“remember,” “know,” and “new”) and fit to a logistic function, implemented with the following four-parameter equation:

$$f(t) = z_{m} + \frac{\Delta z}{1 + e^{-m(t-t_m)/\Delta t}} \quad (1)$$

where $t$ is cue–target similarity, $z_{m}$ represents the lower limit of $z$ confidence, $\Delta z$ is the net change in $z$ confidence, $m$ is the maximum slope, and $t_m$ is the corresponding point along the $x$ axis. The observed data points were weighted by frequency of responses, and curves were fit using least squares minimization. The models were unconstrained so as to be both flexible and theory agnostic. Mean group-level data points and the minimized fits are presented in Figure 6A (left). The logistic function accounted for 91% of the variance in “remember” data, $R^2 = .908, F(4, 236) = 1142.623, p < .001$, and 73% of the variance in “know” data, $R^2 = .729, F(4, 226) = 273.950, p < .001$.

We set out to determine whether “remember” and “know” trends arose from independent distributions and were not simply an arbitrary categorization based on memory strength. Parametric differences were explored using sum of squares reduction tests as follows. To start with, a combined “remember” and “know” eight-parameter (full) model was fit to both response sets simultaneously, represented by the following equation:

$$F(t)_{r,k} = f(t)_{proc=r} + f(t)_{proc=k} \quad (2)$$

Elfman and Yonelinas 883
where \( r \) is “remember” and \( k \) is “know.” \( proc \) is the process (remember or know) for a given data point and \( f(t) \) is the function shown in Equation 1 but with parameters that are process specific. The full model accounted for 88% of total variance, \( R^2 = .879 \), \( F(8, 226) = 273.950 \), \( p < .001 \). A \( t \) test revealed a significant difference in the slope parameters for the two processes, with “remember” (\( M = 5.095, SE = 0.316 \)) having a steeper maximum gradient than “know” (\( M = 2.349, SE = 0.257 \)), \( t(462) = 6.730 \), \( p < .001 \). Likewise, there was a significant difference in maximum confidence, with a higher \( z_{\text{max}} \) estimate for “remember” (\( M = 2.173, SE = 0.069 \)) than “know” (\( M = 1.384, SE = 0.195 \)), \( t(462) = 3.812, p < .001 \). The other parameters, \( z_{\text{min}} \) and \( t_m \), did not differ significantly between processes. To determine whether slope was an independently significant factor in differentiating the two processes, the eight-parameter model was compared with a reduced, seven-parameter model in which the same slope parameter was used for both processes. The reduced model accounted for significantly less variance than the full model, \( F(1, 462) = 28.541, p < .001 \). In summary, the analysis indicates that the shape of the “remember” and

---

**Figure 6.** Recognition memory results from Experiment 1. (A) Normalized recognition confidence measures for items receiving a “remember,” “know,” or “new” response as cue–target similarity was varied from 0 to 1. The solid lines represent sigmoid functions that were fit to the observed data. The panel on the left presents recognition performance for all of the test trials and shows that, as cue–target similarity increased, the confidence ratings for both remember and know trials increased, with remember trials reaching a higher level of confidence and exhibiting a steeper slope than familiarity trials. In contrast, new items and misses led to increasingly confident new responses (i.e., negative values). The panel on the right shows the same results for a subset of “remember” and “know” trials that were matched for final confidence ratings (for “remember,” \( z < 3 \), and for “know,” \( z > 1.7 \)) and indicates that recollection exhibited a steeper similarity gradient than familiarity, even when overall confidence was controlled. (B) Quantifying the steepness of the recollection and familiarity similarity gradients. The left panel presents the average point at which the “remember” and “know” trials departed from zero memory confidence and the average point at which they reached their maximum level of confidence. The figure shows that the “remember” trials that reached a maximum confidence of 4 exhibited a steeper slope than the “know” trials that reached maximum confidence levels of 4, 3, or 2. There were too few “remember” trials that ended at levels of 3 or 2 to plot. The figure shows aggregate rather subject-level measures, so error bars are not included. The right panel presents the results of a “within-subject” analysis comparing mean slope estimates with ±1 SEM, in which the slope was measured for each trial. The results of the aggregate and the single trial analyses indicate that recollection exhibited a steeper similarity gradient than familiarity.
Matching performance curve fits. To further verify that the shapes of the “remember” and “know” similarity functions were different, we conducted analyses in which we controlled for recognition memory confidence. That is, we examined “remember” and “know” trials that were approximately equal in final z confidence, thus matching for memory strength. Selecting “remember” trials with final z confidence less than 3 ($M = 2.085, SD = 0.562$) and “know” trials with z-confidence greater than 1.7 ($M = 2.008, SD = 0.319$) resulted in the most inclusive data set while maintaining a statistically nonsignificant difference in final z confidence, $t(168.526) = 1.416, p > .1$ (equal variances not assumed). The average similarity functions for the “remember” and “know” responses with matched confidence are presented in Figure 6A (right). Separate fits of the logistic function accounted for 91% of the variance in “remember” data, $R^2 = .906, F(4, 226) = 1076.036, p < .001$, and 88% of the variance in “know” data, $R^2 = .875, F(4, 166) = 616.392, p < .001$. The full eight-parameter model accounted for a significant amount of total variance, $R^2 = .906, F(4, 226) = 1076.036, p < .001$. Importantly, a significant difference was evident between the slope ($m$) parameters for “remember” ($M = 5.034, SE = 0.285$) and “know” ($M = 3.463, SE = 0.351$), $t(462) = 3.476, p < .001$, and there were no other significant differences. Also, because the slope was the only parameter to show a significant effect, the model reduction procedure was unnecessary. Thus, when overall level of recognition confidence was held constant, the slope of the recollection gradient was steeper than that of the familiarity gradient.

Additional slope measures. The function-fitting analyses above were limited to group-level data; that is, there were too few trials to reliably fit separate continuous functions for each participant. For additional analyses, we explored simpler metrics that were obtainable at the single-trial level for each participant. Specifically, we tested whether (a) confidence increased at a faster rate in “remember” trials than in “know” trials and (b) “remember” trials exhibited a larger maximum “step” in confidence when matching for performance.

For the first test, each trial was scored on three metrics: $t_0$, the point in cue–target similarity before confidence changed from “unsure”; $t_c$, the point at which confidence ceased to change; and $\Delta c$, the net change in confidence. Figure 6B (left) shows summaries of the metrics aggregated over all trials. The figure compares matched performance (confidence level = 4) “remember” and “know” trials, and “know” trials that ended at confidence levels of 3 and 2. Average slopes ($\Delta c/t_1 - t_0$) were also measured for each trial and are summarized in the right figure. There were insufficient trials in each category of response to run typical repeated-measures analyses, so the comparisons were performed using a linear mixed model design. For “remember” versus “know” trials ending at $c = 4$, participants reached maximum confidence ($t_c$) significantly earlier for “remember” trials ($M = 0.789, SE = 0.185$) than for “know” trials ($M = 0.934, SE = 0.191$), $t(12.904) = 4.910, p < .001$, and there was no significant difference in the points at which confidence first began to change ($t_0$). Average slopes were also greater for “remember” trials ($M = 13.57, SE = 1.14$) than for “know” trials ($M = 9.60, SE = 1.21$), $t(11.541) = 2.890, p = .014$. In contrast, average slopes did not vary significantly for “know” trials across different levels of final confidence $F(2, 11.920) = 0.748, p = .494$. The results indicate that recollection trials generally did exhibit steeper similarity gradients than the familiarity trials.

For the final test, we compared the maximum step size—that is, the biggest confidence shift between any two adjacent cue–target similarity points within a trial—between processes. Using the mixed model design, participants exhibited larger maximum step sizes at matched performance ($c = 4$) for “remember” trials ($M = 2.112, SE = 0.114$) than “know” trials ($M = 1.684, SE = 0.184$), $t(10.430) = 2.534, p = .029$. Additionally, the maximum step size for “familiarity” did not vary significantly as a function of final confidence, $F(2, 11.920) = 0.748, p = .494$. In summary, the single trial analysis further verified that the “remember” trials exhibited steeper similarity gradients than the “knowing” trials.

Discussion

The results validated the prediction of the CLS model in showing that recollection exhibited a steeper similarity gradient than familiarity. As test stimuli were morphed to studied faces, recognition confidence increased more gradually for items recognized on the basis of familiarity than those recognized on the basis of recollection. This pattern was observed when fitting the average similarity gradients using all trials and when excluding trials to control for differences in overall level of performance. In addition, single trial analysis indicated that recollection trials were associated with steeper similarity functions than familiarity trials.

One question that the current results do not answer, however, is whether the same pattern of results would be observed under conditions in which similarity was not incrementally morphed within single trials. That is, perhaps the similarity functions were affected by having each test item morph across contiguous presentations within a single trial. To test the generalizability of the results from Experiment 1, we conducted another experiment in which each test item was only tested once and
cue–target similarity was varied across items. In addition, the results of Experiment 1 utilized the remember/know procedure to separate recollection and familiarity-based trials. To determine whether the results generalize to another measurement procedure, the second experiment includes a sufficiently large number of trials to support an ROC analysis; on the basis of the ROC shape, we are able to estimate recollection and familiarity (Yonelinas, 1994).

**Experiment 2**

Experiment 2 was similar to Experiment 1 except that participants studied a list of faces and then at test were presented with a random mixture of faces, each appearing once, that varied in cue–target similarity between .4, .6, .8, and 1. On the basis of the confidence responses, we plotted ROC curves, which were used to derive estimates of recollection and familiarity for each participant and at each level of cue–target similarity. Similarity gradients for recollection and familiarity were then contrasted to determine if recollection exhibited a steeper gradient than familiarity.

**Methods**

**Participants and materials.** Twenty undergraduate psychology students (mean age = 20 years) participated in the experiment for course credit. One participant was excluded for using only two of the response keys. For the materials, 560 unique faces were created using the same method described in Experiment 1.

**Design and procedure.** Each session comprised 16 study–test blocks. In each study phase, participants studied 10 unique faces that appeared for 5 sec, with a 0.5 sec ISI. To aid encoding, participants were required to guess the ethnicity of each face, selecting from Asian, European, African, and Middle-Eastern. The study list length and presentation durations were selected to avoid floor and ceiling levels of performance. For each test phase, each of the faces from the prior study phase was morphed with a unique, novel face to create one of four possible levels of cue–target similarity of .4, .6, .8, or 1. Each studied face appeared only once at test, at one of those similarity levels—10 faces in total—mixed with 10 novel faces. Participants were instructed to rate each face on a 6-point scale, from 1 (sure new) to 6 (sure old). The test phase was self-paced, and there was a 10-sec rest period between blocks.

**Results**

Average ROCs along with dual process signal detection (DPDS; Yonelinas, 2002) model fits were plotted for each level of cue–target similarity (Figure 7, left). An examination of the ROCs shows that they were in line with what was expected on the basis of the CLS simulations (compare to Figure 4C). That is, the ROCs moved upward as similarity increased. Moreover, the y-intercept increased most noticeably at the middle of the similarity manipulation (between .6 and .8) and less so earlier (.4 to .6) and later (.8 to 1) on the similarity scale. This pattern is consistent with a dramatic increase in recollection in the middle of the similarity manipulation.

To compare the effects of cue–target similarity on recollection and familiarity, we examined the recollection and familiarity estimates obtained from the DPDS model (Figure 7, right). Familiarity exhibited a relatively linear function that increased gradually as stimulus similarity increased, whereas recollection exhibited a steeper gradient. The observed gradients are consistent with the model predictions (see Figure 3C) and converge with the results of Experiment 1 (see Figure 6).

To quantify these differences, the ROCs were simultaneously fit to a single model containing a logistic function for recollection and another logistic function for familiarity. The model fits were constrained to have x and y intercepts of zero, corresponding to an assumption of zero discriminability at zero cue–target similarity. The model was fit to each participant’s data to obtain parameter estimates for within-participant tests. To ensure the parameters for recollection and familiarity were comparable, predicted values were converted to represent proportions of the maximum value (i.e., the predicted value at cue–target similarity of 1). Participants had a greater slope on average for recollection (M = 4.876, SD = 4.069) than for familiarity (M = 1.780, SD = 0.853), t(18) = 3.196, p = .003 (one-tailed), verifying that recollection had a steeper similarity gradient than familiarity. Note that we report one-tailed tests here because the direction of the effect was predicted by the simulations. In contrast, Experiment 1 was more exploratory so we exercised greater caution in our predictions.

In addition, the inflection point—where on the similarity scale the slope reached maximum steepness—occurred at higher levels of similarity for recollection (M = 0.895, SD = 0.457), than for familiarity (M = 0.465, SD = 0.271), t(18) = 3.027, p = .004 (one-tailed), suggesting that recollection occurred over a smaller range of similarity levels than did familiarity. Further supporting the notion that recollection occurs over a narrower range, an examination of Figure 7 indicates that at low levels of cue–target similarity (left side of the figure), familiarity was beginning to show an increased response to more similar items whereas recollection estimates remained close to zero until similarity was much greater. For example, at a similarity of .4, estimates of familiarity (df) were on average 35.7% (SD = 17.9%) of maximum familiarity (at similarity of 1) compared with 14.3% (SD = 14.4%) for recollection, t(18) = 3.610, p = .001 (one-tailed).
Discussion

ROC plots were constructed from recognition confidence responses at varying levels of cue–target similarity, which were then fit to the DPSD model to obtain a range of recollection and familiarity estimates. The results were consistent with the remember/know results of the first experiment and the CLS model simulations, showing relatively steep, nonlinear recollection-based functions in the hippocampus compared with more linear familiarity-based functions in the MTLc. In addition to confirming that recollection had a significantly steeper gradient than familiarity, the analysis showed that familiarity had a greater impact on recognition at low levels of similarity compared with recollection.

The results of the current study are similar to a recent report that examined the effects of varying the cue–target similarity for photos on recognition confidence judgments (Pustina, Gizewski, Forsting, Daum, & Suchan, 2012). The similarity gradients of recollection and familiarity were not quantified as continuous functions, so a direct comparison with the current results is not possible, but their results indicated that familiarity estimates decreased approximately linearly with decreasing cue–target similarity, compared with recollection, which appeared to exhibit a more nonlinear response and were thus in general agreement with the results observed in the current study.

The average ROCs (Figure 7A) also closely matched the simulated ROCs (Figure 4C) from the CLS model. That is, familiarity—indicated by the level of ROC curvature—increased gradually with similarity, consistent with the MTLc signal. In contrast, recollection—indicated by the y intercept—exhibited a nonlinear change, increasing more dramatically at middle similarity intervals, consistent with the hippocampal signals.

GENERAL DISCUSSION

The current study was conducted to test how differences in the similarity between retrieval cues and previously studied items affect the processes of recollection and familiarity in human recognition memory. We first conducted simulations with the CLS model (Norman & O’Reilly, 2003), which instantiates separate hippocampal and MTLc networks. The networks were trained on a list of stimulus patterns, and at test the patterns were varied incrementally from new to old. Over single trials, the hippocampal network produced discrete transitions from no retrieval to accurate, pattern-completed retrieval when stimuli reached a critical threshold of similarity to the original item. In contrast, the MTLc produced more linear, continuous transitions, from low to high familiarity (i.e., pattern sharpness), for studied items. The networks also demonstrated markedly different functions even when performance was averaged over many trials. The hippocampus produced a nonlinear, sigmoidal function with a comparatively steep slope, whereas the MTLc produced a more linear function, consistent with a global-match signal.

Two recognition experiments were conducted to test whether the similarity functions produced in the simulations were predictive of human recollection and familiarity. The first experiment examined performance within single trials by observing responses at multiple points...
of similarity and used remember/know responses as indices of recollection and familiarity. The second experiment tested only a single, random similarity position for each item, and recollection and familiarity were estimated from confidence-based ROCs. The results from the two behavioral experiments were consistent with the CLS model simulations. Namely, the experiments produced relatively steep, nonlinear similarity gradients for recollection that correspond to the predictions of the hippocampal network and more linear, continuous functions for familiarity that are consistent with the MTLc network.

Relating the Current Findings to Previous Research

The current findings support a growing literature that ties the hippocampus to recollection-based recognition memory and the MTLc to familiarity-based recognition (Eichenbaum et al., 2007; Montaldi, Spencer, Roberts, & Mayes, 2006; Yonelinas, Otten, Shaw, & Rugg, 2005; Ranganath et al., 2004). Importantly, although some previous behavioral studies have investigated the effects of different levels of cue-target similarity on memory performance (e.g., Kim & Yassa, 2013; Pustina et al., 2012; Preminger, Blumenfeld, Sagi, & Tsodyks, 2009), they did not estimate similarity functions that contrasted recollection and familiarity. Thus, the current findings provide a critical direct test of the hypothesis that the similarity gradients of recollection and familiarity differ.

The current study bears some similarity to studies of false recognition, in which individuals must discriminate between studied items and related lures. Such studies have typically shown that the probability of nonstudied items being falsely recollected is rare, compared with the probability that they are falsely recognized on the basis of familiarity. However, when new items are high associates of the studied items, both processes can lead to high levels of false recognition (for a review, see Yonelinas, 2002). The present findings are broadly consistent with this research in the sense that when test items were very different from targets (i.e., low similarity), lures rarely led to recollection responses, whereas when test items became very similar to studied items, these items often led to recollection.

A core feature of the CLS model is that the hippocampus performs pattern separation, thereby making similar items less prone to interference. In recent behavioral experiments, Kim and Yassa (2013) showed that individuals will often identify, on the basis of recollection, lure items that are similar to studied items, thus showing that recollection can occur in the absence of pattern separation. Indeed, past research has shown that pattern separation is not without practical limits (Eifman et al., 2008), and the current results support this by showing that recollection often occurred when items differed from the studied targets. Thus, it is important to bear in mind that, although it is helpful to examine recollection through the guise of computational mechanisms such as pattern separation and pattern completion, the relationship between recollection and these mechanisms is a complex one.

In a related paradigm, Preminger et al. (2009) showed that memory attractors (stable neural representations) can be manipulated by gradually morphing images of learned faces from a “source” to a “target,” over a period of weeks. When the morphing procedure was completed, target faces were often misidentified as source items, indicating that the original attractor had “broadened” to accommodate the new target information. Although the authors did not differentiate between recollection and familiarity, the findings suggest that similarity gradients are to some extent malleable. Thus, an interesting challenge for future research would be to determine whether the attractors associated with recollection and familiarity are differentially affected by this gradual remapping procedure.

The current findings are broadly consistent with a number of human fMRI studies in which memory retrieval was found to be associated with discrete activation states in the hippocampus and more continuous signals in surrounding MTLc areas. For example, in tests of item and associative recognition, hippocampal activation is differentially related to accurate, high-confidence responses to studied items but shows no such trend across lower confidence responses (e.g., Daselaar, Fleck, & Cabeza, 2006; Montaldi et al., 2006), whereas activation in the perirhinal cortex and surrounding MTLc structures tracks more linearly with confidence responses (Daselaar et al., 2006; Yonelinas et al., 2005; Ranganath et al., 2004). However, a limitation of this comparison is that imaging studies typically have not investigated how the neural activation associated with recollection and familiarity varies with objective similarity changes. Thus, an interesting challenge for future research would be to determine whether the hippocampus and perirhinal cortex show step-like or graded activation similarity functions as items are gradually morphed from new to old.

Similarity manipulations have been used in a number of rodent studies in which hippocampal neurons were recorded as a surrounding environment was gradually morphed between two prior exposed shapes (e.g., a circle and a square). Although we cannot assert a link between single cell firing patterns and human recognition similarity functions, attractor dynamics that have been identified in the hippocampus—such as an abrupt shift in the spatial firing locations of neurons near the mid-point of the morph (e.g., Colgin et al., 2010; Wills, Lever, Cacucci, Burgess, & O’Keefe, 2005)—are nonetheless consistent with the discrete activation states observed in the current hippocampal model simulations as similarity was varied, and likewise, the steep gradient observed in recollection part-way along the similarity scale.
Could a Single Process Model Account for the Findings?

The CLS model assumes that recognition memory is the result of two neuroanatomically dissociable networks and is therefore theoretically aligned with dual process memory models (e.g., DPSD; see Yonelinas, 2002). The existing evidence for the contribution of two processes in recognition memory is quite extensive (for reviews, see Diana, Reder, Arndt, & Park, 2006; Yonelinas, 2002; but see Parks & Yonelinas, 2007; Wixted, 2007). Nevertheless, it is useful to ask whether the current results might also arise naturally from a single-process account of memory. Although the current experiments were not designed to address this issue, they do present a number of challenges for any such approach. For example, in Experiment 1 we observed distinct similarity gradients for recollection and familiarity-based responses. That is, direct statistical tests indicated that a two-parameter account was preferred over the single parameter account. Importantly, our statistical tests showed that even when accounting for the one less degree of freedom of the single-parameter model, the single-parameter model was rejected when compared with the two-parameter model. It might be argued, based on the initial analysis, that the higher levels of confidence associated with recollection trials compared with familiarity trials complicated the comparison; however, a direct comparison of the similarity gradients when overall performance was controlled for indicated that the two-parameter model was still preferred. In addition, in Experiment 2, direct model contrasts indicated that a model with two slope parameters provided a significantly better fit than a model with only one slope parameter.

Nonetheless, it is important to point out that, although the results verified the a priori predictions of the dual process model, it may be possible to develop alternative single process models that provide a post hoc account of the data. For example, a single memory system that represents both item information and associative information might be able to produce dissociations if one assumes that recognition reflects a mixture of both associative and item information. Although, whether such a model would naturally predict differences in similarity gradients is unclear. Moreover, whether it could naturally account for the specific differences in the shapes of the ROCs that were observed in Experiment 3 is also unknown. Critically, the ROCs did not simply indicate a monotonic increase in discriminability, but rather, the intercepts (i.e., recollection) increased most dramatically around the middle of the similarity scale.

Computational Insights and Predictions

The current simulation work did not set out to explore in detail the performance characteristics of individual layers within the networks. However, it is interesting to note that the layers making up the hippocampus, which include DG/CA3 and CA1, have quite different attributes. By itself, the architecture of CA1 is quite similar to the MTLc model in the sense that it has less lateral inhibition than the DG or CA3 (as it is instantiated in the current model). That is, it supports graded states of activation and so we might expect to see steeper similarity gradients in the DG/CA3 than in CA1. However, results from some related simulation work (Elfman, Aly, & Yonelinas, 2014) suggest that the story about hippocampal subfields may be somewhat more complicated. Although we do not go into such details in this paper, we would suggest that the behavior of CA1 is largely dependent on task demands. For example, in an experiment in which the hippocampus is probed only with related lures, we might expect a linear relationship in CA1 that reflects global match (if it happens that CA3 is always pattern completing), whereas in an experiment with unrelated lures, one might expect more thresholded performance resulting from a marked drop in activation from CA3 for many items. Additional factors such as hypothesized encode and retrieve phases (e.g., Hasselmo, Bodelon, & Wyble, 2002) further complicate the picture, making this an interesting topic for future research.

The behavioral experiments show that recollection and familiarity have different similarity gradients for images of faces. However, the CLS model is agnostic about the types of materials that give rise to these similarity gradients, so a reasonable prediction is that the current effects should generalize across different materials and modalities. Future studies that examine the effects of similarity using other stimulus classes are needed to test this prediction.

Another interesting aspect of the simulations is that the steeper recollection gradient of the hippocampal network is an emergent—or at least, cumulative—property of the entire network architecture. That is, attenuating or eliminating critical architectural features—such as the recurrent CA3 projections or the detonator cells of the DG—did not result in monotonic output gradients that are comparable to the cortical network. Whether this is true of the human hippocampus is a challenging question but could potentially be addressed with animal lesion studies.

Limitations

The current behavioral findings show that at different levels of similarity, recollection, and familiarity produce different retrieval states that are predicted by their respective similarity functions. However, how the two processes arrive at their respective states is another question. That is, when the networks that underlie these processes are presented with a retrieval cue, there is a progression of activation states (i.e., a “trajectory”) that ultimately leads...
to a stable attractor pattern (i.e., a local minimum). In other words, when a partial cue triggers activation in a memory network, if the resulting pattern falls within a basin of attraction, that pattern will “descend” the basin towards a final, fully retrieved (i.e., pattern-completed) memory. Capturing this descent through behavioral observations would likely prove difficult, not least because of the short durations over which retrieval occurs (i.e., individuals can perform effectively with a stimulus–response deadline under 1000 msec; Yonelinas, 2002). Note that there is some evidence of attractor dynamics over brief timescales in animal studies (for a review, see Daelli & Treves, 2010), and in a recent study, when rodents were exposed to a changed environment, a brief period of competitive flickering was observed in hippocampal subfield CA3 as activation quickly shifted from a neuronal ensemble associated with the old environment to a new ensemble (Jezek, Henriksen, Treves, Moser, & Moser, 2011). Examining these short timescale dynamics in humans, both behaviorally and biologically, will be a challenge for future research.

Another question of interest is whether the neocortical and hippocampal networks interact with one another in a way that affects their respective functioning. In the current research, we modeled the two networks separately and produced simulated recognition ROCs by combining only their output measures. However, we ran some additional simulations in which we structurally combined the two networks and found that this had the effect of bolstering hippocampal performance. In particular, at low levels of similarity, the familiarity signal sometimes “nudged” hippocampal activation toward the encoded pattern. Further work on a combined model may yield other interesting predictions.

Conclusions

In this paper, we explored the similarity functions of recollection and familiarity and found the two processes to produce markedly different gradients. The findings were consistent with the predictions of a popular computational model of the hippocampus and MTLc, indicating that recollection, which is dependent on the hippocampus, has steep, nonlinear similarity functions, whereas familiarity, which is related to the MTLc and association cortex, has wider and more linear gradients. The current work represents an important step in validating the predictions of current computational models and characterizing a core aspect of memory performance.

APPENDIX: NETWORK PARAMETERS

The following notes are a selective description of the rules and parameters used in the model simulations. Table A1 shows each layer size (i.e., number of units) and percentage of activity determined by the k-winners-take-all (Norman & O’Reilly, 2003). Table A2 shows the properties of the main projections, including the mean initial weight strengths (Mean), variances of the weight distribution (Var), relative strengths of the projections during encoding (Scale enc) and retrieval (Scale retr), and the proportions of receiving units that each sending unit is connected to (% Con).

Table A1. Layer Sizes and Activity Levels

<table>
<thead>
<tr>
<th>Layer/Area</th>
<th>Units</th>
<th>Activity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hippocampus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EC (in/out)</td>
<td>144</td>
<td>25.0</td>
</tr>
<tr>
<td>DG</td>
<td>1600</td>
<td>1.0</td>
</tr>
<tr>
<td>Area CA3</td>
<td>480</td>
<td>3.8</td>
</tr>
<tr>
<td>Area CA1</td>
<td>384</td>
<td>9.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Neocortex</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower-level cortex (Input)</td>
<td>144</td>
<td>25.0</td>
</tr>
<tr>
<td>Association/MTLc</td>
<td>400</td>
<td>20.0</td>
</tr>
</tbody>
</table>

in/out = input and output layers, respectively.

Table A2. Properties of Modifiable Projections

<table>
<thead>
<tr>
<th>Projection</th>
<th>Mean</th>
<th>Var</th>
<th>Scale</th>
<th>% Con</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hippocampus</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EC to DG, CA3 (perforant pathway)</td>
<td>0.5</td>
<td>0.1</td>
<td>1</td>
<td>25</td>
</tr>
<tr>
<td>DG to CA3 (mossy fiber) (encode/retrieve)</td>
<td>0.9</td>
<td>0.01</td>
<td>15/0</td>
<td>4</td>
</tr>
<tr>
<td>CA3 recurrent</td>
<td>0.5</td>
<td>0.1</td>
<td>2</td>
<td>100</td>
</tr>
<tr>
<td>CA3 to CA1 (Schaffer collaterals)</td>
<td>0.5</td>
<td>0.1</td>
<td>3</td>
<td>100</td>
</tr>
</tbody>
</table>

| Neocortex                  |      |      |       |       |
| Input to association/MTLc | 0.5  | 0.25 | 1     | 25    |

Mean = mean initial weight strength; Var = variance of initial weight distribution; Scale = scaling of this projection relative to other projections; % Con = percentage connectivity.
Acknowledgments

Preparation of this article was supported by NIMH grant MH059352.

Reprint requests should be sent to Andrew P. Yonelinas, Department of Psychology, University of California, Davis, CA 95616, or via e-mail: apyonelinas@ucdavis.edu.

REFERENCES


Chapter 5: A Neurocomputational Account Of Memory and Perception: Thresholded and Graded Signals in the Human Hippocampus

**Authors:** Kane W. Elfman, Dr Mariam Aly, Dr Andrew P. Yonelinas (PI)

**Journal:** Hippocampus

**Date of publication:** December, 2014

**Abstract:** Recent evidence suggests that the hippocampus, a region critical for long-term memory, also supports certain forms of high-level visual perception. A seemingly paradoxical finding is that, unlike the thresholded hippocampal signals associated with memory, the hippocampus produces graded, strength-based signals in perception. The current paper tests a neurocomputational model of the hippocampus, based on the complementary learning systems framework, to determine if the same model can account for both memory and perception, and whether it produces the appropriate thresholded and strength-based signals in these two types of tasks. The simulations showed that the hippocampus, and most prominently the CA1 subfield, produced graded signals when required to discriminate between highly similar stimuli in a perception task, but generated thresholded patterns of activity in recognition memory. A threshold was observed in recognition memory because pattern completion occurred for only some trials and completely failed to occur for others; conversely, in perception, pattern completion almost always occurred because of the high degree of item similarity. These results offer a neurocomputational account of the distinct hippocampal signals associated with perception and memory, and are broadly consistent with proposals that CA1 functions as a comparator of expected versus perceived events. We conclude that the hippocampal computations required for high-level perceptual discrimination are congruous with current neurocomputational models that account for recognition memory, and fit neatly into a broader description of the role of the hippocampus for the processing of complex relational information.

**Author contributions.** The research and writing of this paper was conducted by myself under the supervision of principal investigator Dr Andrew Yonelinas. The research included designing and implementing neural network model simulations of the hippocampus. The original draft for the paper was written entirely by myself and
then revised with assistance from Prof Yonelinas and Dr Mariam Aly in preparation for submission for journal publication.
Neurocomputational Account of Memory and Perception: Thresholded and Graded Signals in the Hippocampus

Kane W. Elfman, Mariam Aly, and Andrew P. Yonelinas*

ABSTRACT: Recent evidence suggests that the hippocampus, a region critical for long-term memory, also supports certain forms of high-level visual perception. A seemingly paradoxical finding is that, unlike the thresholded hippocampal signals associated with memory, the hippocampus produces graded, strength-based signals in perception. This article tests a neurocomputational model of the hippocampus, based on the complementary learning systems framework, to determine if the same model can account for both memory and perception, and whether it produces the appropriate thresholded and strength-based signals in these two types of tasks. The simulations showed that the hippocampus, and most prominently the CA1 subfield, produced graded signals when required to discriminate between highly similar stimuli in a perception task, but generated thresholded patterns of activity in recognition memory. A threshold was observed in recognition memory because pattern completion occurred for only some trials and completely failed to occur for others; conversely, in perception, pattern completion always occurred because of the high degree of item similarity. These results offer a neurocomputational account of the distinct hippocampal signals associated with perception and memory, and are broadly consistent with proposals that CA1 functions as a comparator of expected versus perceived events. We conclude that the hippocampal computations required for high-level perceptual discrimination are congruous with current neurocomputational models that account for recognition memory, and fit neatly into a broader description of the role of the hippocampus for the processing of complex relational information. © 2014 Wiley Periodicals, Inc.

KEY WORDS: computational model; perception; memory; recognition; recollection

INTRODUCTION

Neuropsychological studies have indicated that hippocampal damage leads to profound long-term memory impairments (Scoville and Milner, 1957; Baddeley and Warrington, 1970; Graf and Schacter, 1984; Zola-Morgan et al., 1986; Cohen and Eichenbaum, 1993; Vargha-Khadem et al., 1997; Reed and Squire, 1998; Yonelinas et al., 2002). Neuroimaging and electrophysiological studies have corroborated the role of the hippocampus in long-term memory by establishing a clear link between hippocampal activity and memory encoding and retrieval (Ranganath et al., 2004; Yonelinas et al., 2005; Montaldi et al., 2006; Sederberg et al., 2006; Fell et al., 2011; Axmacher et al., 2010).

Additionally, neuroimaging studies have indicated that memory-related hippocampal activity is characterized by relatively distinct states (Ranganath et al., 2004; Yonelinas et al., 2005; Montaldi et al., 2006; for a review, see Eichenbaum et al., 2007). For example, in recognition tests, previously studied items that are recognized on the basis of recollection are associated with high levels of hippocampal activity, whereas items that are not recollected are associated with very low levels of activity, even if they are successfully recognized on the basis of familiarity (Yonelinas et al., 2005; Montaldi et al., 2006). Similarly thresholded patterns of hippocampal activity have been found in relational memory tasks (e.g., source and associative memory; Eichenbaum et al., 2007), which are largely dependent on recollection (Yonelinas, 2002; Yonelinas et al., 2010).

Cognitive memory models describe recollection as a threshold retrieval process, in that qualitative information about a study event is either retrieved or not (Yonelinas, 2002). For example, in tests of relational memory, thresholded memory retrieval has been evidenced by linear receiver operating characteristics [ROCs]. In addition, in a recent fMRI study, ROCs constructed from hippocampal activity values during a source memory task were found to be linear and therefore consistent with the threshold model (Slutnick and Thakral, 2013). The thresholded response of the hippocampus, and of recollection, can be contrasted with the perirhinal cortex, which responds in a more graded manner and is associated with familiarity; or strength-based, memory (Ranganath et al., 2004; Montaldi et al., 2006; Staresina et al., 2013), which in turn produces curved ROCs (Parks and Yonelinas, 2007).

Challenges to the Traditional View of Hippocampal Function

The neuropsychological and neuroimaging work on the hippocampus has, until recently, focused on its role...
in long-term memory. However, empirical findings in the past decade have challenged the traditional view that the hippocampus is critical for long-term memory but not other cognitive functions (Scoville and Milner, 1957; Baddeley and Warrington, 1970; Graf and Schacter, 1984; Zola-Morgan et al., 1986; Cohen and Eichenbaum, 1993; Reed and Squire, 1998). For example, it is now clear that hippocampal damage can impair working memory and high-level scene perception (e.g., Bussey and Saksida, 2005; Olson et al., 2006; Graham et al., 2010; Lee et al., 2005, 2012; Warren et al., 2012; Aly et al., 2013; Yonelinas, 2013). Moreover, neuroimaging studies have linked perceptual processing with hippocampal activity in healthy adults (e.g., Barense et al., 2010; Lee and Rudebeck, 2010; Lee et al., 2008, 2010; Mundy et al., 2012; Aly et al., 2013).

As a result of such findings, it has been argued that the hippocampus serves a broader role than previously thought, contributing to perception and working memory as well as long-term memory (Cohen and Eichenbaum, 1993; Graham et al., 2010; Saksida and Bussey, 2010; Lee et al., 2012; Olsen et al., 2012; Nadal and Peterson, 2013; Shohamy and Turk-Browne, 2013; Yonelinas, 2013). The mechanisms that would account for the joint roles of the hippocampus across these cognitive domains are still not well established, although it has been suggested that the hippocampus is involved in relational (Cohen and Eichenbaum, 1993; Olsen et al., 2012) or complex conjunctive (Graham et al., 2010; Lee et al., 2012; Saksida and Bussey, 2010) processing in the service of perception, working memory, and long-term memory (also see Nadal and Peterson, 2013; Shohamy and Turk-Browne, 2013; Yonelinas, 2013).

Insights on Hippocampal Function from Neurocomputational Models

Neurocomputational models have furthered our understanding of how memories are encoded and retrieved in the hippocampus by exploring how distinct patterns of neural activity and task performance arise from the network’s architecture (e.g., McNaughton and Morris, 1987; Treves and Rolls, 1994; McClelland et al., 1995; Rolls, 1996; Hasselmo and Wyble, 1997; Rudy and O’Reilly, 1999; Hasselmo et al., 2002; Norman and O’Reilly, 2003; Hasselmo and Eichenbaum, 2005). However, it is currently unclear whether these models can also account for the recent findings linking the hippocampus to perception. In addition, an important challenge for any hippocampal model that attempts to account for both the memory and perception findings is that the nature of the respective signals can be quite different; whereas in memory the contribution of the hippocampus appears to be thresholded, or state-based, there is evidence that in perception its contribution is graded, or strength-based (Aly et al., 2013).

In this article, we first provide an overview of how computational models of the hippocampus account for thresholded memory retrieval. Then, we outline the recent evidence implicating the hippocampus in high-level perception and highlight how the type of signal associated with perception differs from memory, and the challenge this poses for existing models. Finally, we report simulation results from a computational model of the hippocampus to determine whether it can simultaneously account for memory and perception, and specifically whether it can produce the thresholded memory signals and graded perception signals that have been observed in the empirical literature.

Thresholded Hippocampal Signals in Long-Term Memory

Computational models of long-term memory generally agree that the hippocampus is capable of forming distinct representations of events, even when the events are quite similar (McNaughton and Morris, 1987; O’Reilly and McClelland, 1994; McClelland et al., 1995; Rolls, 1996). This capability, sometimes referred to as pattern separation (or orthogonalization), is thought to result from high levels of lateral inhibition, effectively limiting the network to very sparse excitatory activity, especially within the dentate gyrus (DG). This constraint creates a competitive learning environment, so that if two stimuli differ even subtly, the hippocampal neurons that activate to encode those stimuli (via Hebbian learning) are likely to differ dramatically.

A common assumption is that the hippocampus is fundamental to the storage and reinstatement of episodic memories. Due to the singular nature of such events (e.g., remembering where you parked your car today), synaptic weights must be updated rapidly to capture unique associations (e.g., “Toyota,” “Shields Avenue,” and “this morning”) as they occur (Tulving and Markowitsch, 1998; Aggleton and Brown, 1999; Eichenbaum and Cohen, 2001; Leutgeb et al., 2006, 2007). Pattern separation is therefore important as it allows new events to be quickly learned without damaging or confuting representations of other, similar events (e.g., remembering where you parked your car today versus yesterday).

At the time of retrieval, if the network is presented with part of the original stimulus event, activation will sometimes spread through the network and reinstate aspects of the original event that are not currently present; this is referred to as pattern completion. Some computational models and electrophysiological studies suggest that this process is facilitated by recurrent connections within CA3, which acts as an autoassociation network (Marr, 1971; O’Reilly and McClelland, 1994; Rudy and O’Reilly, 1999; but for evidence supporting an alternative hypothesis, see Colgin et al., 2010). However, due to the aforementioned effects of pattern separation, if a retrieval cue differs too greatly from the original stimulus, the network will be unable to reactivate the learned pattern. Thus, the hippocampus is expected to behave in a thresholded manner, pattern completing to produce a high level of activity, or else failing to pattern complete and thus producing much lower activity.

The above computational account of the hippocampus is consistent with theories that propose that the hippocampus is

---

*By “high-level perception”, we refer to perceptual discriminations or judgments about complex scenes, faces, or objects (i.e., at a level of stimulus complexity that is higher than that of simple features such as brightness, orientation, color, etc.).

**Hippocampus**
critical for recollection (Schacter et al., 1996; Aggleton and Brown, 1999; Eldridge et al., 2000; Yonelinas, 2002; Eichenbaum et al., 2007; Yonelinas et al., 2010). Recollection is proposed to be thresholded in the sense that individuals retrieve qualitative information about some events, but for other events recollection fails entirely (Yonelinas, 1994, 2001; Yonelinas et al., 2010). Thus, the thresholded output from computational models of the hippocampus is consistent with the thresholded nature of recollection, and converges with neuroimaging findings that relate thresholded hippocampal signals to encoding and retrieval of recollection-based memories (Ranganath et al., 2004; Yonelinas et al., 2005; Montaldi et al., 2006; see Eichenbaum et al., 2007 for review), as well as patient findings that show that damage to the hippocampus selectively impairs recollection (Yonelinas et al., 2002; see Yonelinas et al., 2010 for review).

It is worthwhile emphasizing that the term “threshold” does not imply that memory retrieval, or pattern completion, is “all-or-none.” That is, successful retrieval may involve reinstating any number of features from a past episode; consequently, hippocampal activity levels related to memory retrieval may vary across different recollected episodes. The term “threshold” is used to indicate that only some studied items are recollected (e.g., items exceeding a recollective threshold, or leading to pattern completion), whereas others fail to be recollected and thus remain indistinguishable from nonstudied items. This definition of “threshold” is consistent with both the hippocampal model instantiated by Norman and O’Reilly (2003), which was adopted for the current simulations, and behavioral models of memory such as the dual-process signal detection model (e.g., Yonelinas, 2001). Such thresholded memory signals give rise to ROCs that approximate a linear function when all but the most stringent response criteria are used (Elfman et al., 2008). Importantly, a thresholded signal does not necessitate that all items that exceed the threshold be associated with the same memory strength.

Strength-Based Hippocampal Signals in Perception

As mentioned above, a growing body of literature indicates that the hippocampus is involved in tasks assessing high-level visual perception (e.g., Lee et al., 2005; Saksida and Bussey, 2005; Barense et al., 2010; Lee and Rudebeck, 2010; Mundy et al., 2012; Warren et al., 2012; see Graham et al., 2010 and Lee et al., 2012 for review). In addition, a recent behavioral study with hippocampal lesion patients and a neuroimaging study with healthy adults have suggested that the hippocampal signal in perception tasks is strength-based, or continuously graded (Aly et al., 2013), rather than reflecting a thresholded signal as in recollection-based memory.

In the behavioral task in Aly et al. (2013), the stimuli were pairs of scenes that were either identical or differed, in that the scenes were slightly contracted or expanded relative to one another. The specific manipulation was a “pinching” or “spherizing,” which kept the sizes of the images the same but contracted (pinched) or expanded (spherized) the scenes, with the largest changes at the center and gradually decreasing changes toward the periphery (also see Aly and Yonelinas, 2012). These changes altered the configural, or relational, information within the scenes (i.e., the relative distance between component parts) without adding or removing any objects.

Patients with selective hippocampal or more extensive MTL damage were briefly presented with these scene pairs, simultaneously, and rated their confidence that the two scenes were the same or different. The patients were significantly impaired relative to healthy controls on this task. Importantly, however, this impairment was specific to one kind of perceptual judgment. Patients were not impaired in their ability to make high-confidence judgments associated with identifying specific details that had changed (i.e., state-based perception; see Aly and Yonelinas, 2012). Rather, they were selectively impaired in graded, lower-confidence judgments associated with a sense of match/mismatch (i.e., strength-based perception, Aly and Yonelinas, 2012). Importantly, this pattern of results was observed even when the analysis was restricted to patients with selective hippocampal damage.

The role of the hippocampus in graded, strength-based perception was corroborated in a follow-up neuroimaging study with healthy adults. Aly et al. found that BOLD activity in the hippocampus linearly tracked graded, lower-confidence responses, but was not differentially sensitive to high-confidence responses associated with a state of identifying specific, detailed differences.

Thus, in contrast to studies of long-term memory, where the hippocampus supports high-confidence responses and shows thresholded activity, in perception tasks, the current evidence suggests that the hippocampus is critical for strength-based responses at lower confidence levels, and shows graded levels of activity. How can these conflicting findings be reconciled?

To our knowledge, no previous computational modeling work has attempted to apply a hippocampal memory model to perception. However, if the existing memory models accurately capture the functions of the hippocampus, then they may also produce the type of strength-based signals that have been observed in perception. In this study, we examined the hippocampal component of Norman and O’Reilly’s (2003) complementary learning systems (CLS) model to test whether (1) this model could be used to make accurate perception judgments, and (2) the same model would produce a graded signal in perception while simultaneously producing a thresholded signal in memory.

**METHOD**

### The Hippocampal Model

The network design is based on the hippocampal component of the CLS model, as implemented by Norman and...
O’Reilly (2003). We used the software package Emergent (Aisa et al., 2008) version 6 that incorporates the Leabra neural network algorithm (O’Reilly and Munakata, 2000). The algorithm includes a continuous rate-code output function, conditional principal component analysis Hebbian learning, and a competitive inhibition function that can be tuned to produce distributed representations with varying levels of sparseness.

The model comprises entorhinal cortex (EC) input and output layers, the DG, CA3, and CA1 layers (see Fig. 1), and has distinct encoding and retrieval modes of operation. Stimuli are presented to the network via the EC input layer, which has projections to DG and CA3 (the perforant pathway), and to CA1. DG, which has very sparse activation, has strong projections to CA3 (the Mossy fibers). During encoding (i.e., in encoding mode), recurrent collaterals in CA3 help bind together the different stimulus features. At the same time, CA1 learns to associate direct input from the EC with input from diffuse CA3 projections (the Schaffer collaterals). During retrieval (i.e., in retrieval mode), the pathway from EC input to CA1, and the mossy fiber pathway (DG-CA3), are inactivated so that any pattern retrieved by CA3 becomes the dominant output signal (Rolls, 1996; Hasselmo et al., 2002).

A subtle departure from past simulation work with the CLS model is that, in the current work, the pathway from EC input to CA1 was diminished in strength during retrieval mode, rather than inactivated completely. This allowed CA1 to act as a “comparator” of the current stimulus from the EC input and the retrieved pattern from CA3 (see Results and Discussion for more). For consistency, this modification was used in both the memory task and the perception task. Importantly, this modification did not affect our ability to replicate previous results with memory stimulations using the CLS model (see Results for more details). Specific model parameters are included in Appendix A.

Perception Task

For the perception task, a single trial comprised two stimulus patterns (items) that were presented to the network sequentially. A single item was composed of 36 feature dimensions, each made up of four units (groups of neurons) with one active unit per dimension. Consequently, any two randomly generated stimulus patterns overlapped by an average of 25%.

The task consisted of 20 trials, with 10 matching and 10 mismatching pairs presented in random order. The pairs were created from 20 randomly generated items. Matching pairs were created by simply duplicating half of the items, and mismatching pairs were created by duplicating the remaining items and then pseudo-randomly changing one or more (depending on the condition) features of the duplicate, ensuring that the same units were never reselected. Simplified illustrations (i.e., fewer features than were actually used) of matching and mismatching item pairs are presented in Figure 2A. The task was run 50 times at each of 36 possible levels of feature mismatch.

The current approach to modeling perceptual differences was to change a small number of features in the mismatch condition. In this way, we were able to vary perceptual similarity in a simple and tractable way, and no changes to the original model’s architecture were required. An alternative approach would have been to make smaller changes to a large number of features, which might appear to be more analogous to the “pinching” or “spherizing” manipulation used in previous empirical studies (Aly and Yonelinas, 2012; Aly et al., 2013, 2014). However, because it is not clear how features in the model are mapped to visual features in real-world stimuli, the current feature change approach can be used as an approximation to the empirical manipulation.

Two simulation approaches were examined; an encode/retrieve model and an always-encode model. For the encode/retrieve model, an item was presented to the network in encoding mode, and then, for the second (matching or mismatching) item, the network was switched to retrieval mode. In taking this approach, we used the network in a way that is comparable to typical long-term memory simulations (Norman and O’Reilly, 2003; Elfman et al., 2008). The always-encode model differed in that both items were presented to the network in encoding mode. We included the latter approach to ensure that...
the results would replicate if one assumes that perception does not involve any kind of (even very short-term) "retrieval."

To measure performance, the overall level of activity was recorded from CA1 and CA3 when the second item in each pair was presented. We also measured activity at the EC output layer, but since this layer receives input from CA1 alone and is in this sense redundant with CA1, we did not analyze the EC data. An additional reason for this approach was that CA1, as it is instantiated in the model, supports comparatively more fine-grained activity levels than EC.

Memory Tasks

Recognition memory was simulated in two ways: using complete retrieval cues and partial retrieval cues. For both simulations, 20 randomly generated items—the same items used in the perception task—comprised the study list. The items were sequentially presented to the network in encoding mode. The network was then switched to retrieval mode for the test phase. The test list comprised the 20 study items and 20 randomly generated lures (see Fig. 2B) presented in random order. For the complete retrieval cue simulations, the entire studied item (or lure) was presented at test. In this respect, the memory task was kept as similar as possible to the perception task (in which complete stimulus patterns, i.e., all feature dimensions, were always presented). For the partial retrieval cue simulations, half the feature dimensions in each test item were left blank. The partial retrieval cue simulation approach is typical of the memory simulations used to probe hippocampal pattern completion (e.g., Norman and O'Reilly, 2003). In addition, it may more accurately reflect episodic item recognition tests, in which participants are required to indicate if the test item was presented in an earlier study context (i.e., the earlier study list, which is represented by the missing features). In either case, as indicated below, the two simulations led to similar conclusions.

RESULTS

Figure 3A shows the average activity of CA1 units for matching item pairs and mismatching pairs in the perception task. The activity distributions were overlapping and Gaussian in shape, and activity strength was on average predictive of match or mismatch condition. Additional simulations showed that, as the number of mismatching features was systematically varied in equal intervals from 0 (complete match) to 36 (complete mismatch), CA1 activity distributions gradually shifted from high to low activity (Fig. B1, Appendix B). That is, on average, the match trials produced higher activity levels than mismatch trials. The initial simulations were conducted using separate encoding and retrieval modes, but the same pattern of results was found using the always-encode model (see below for additional analyses). Together, the results show that the
FIGURE 3. Comparison of hippocampal network performance during the perception and recognition memory tasks. A: Distributions of average activity levels for CA1 in the perception task (shown for matching items and items with 4 and 8 mismatching features out of a possible 36). B: Distributions of CA1 activity for studied (target) versus nonstudied (lure) items in the complete retrieval cue recognition memory task. In perception (A), the match and mismatch items produced overlapping strength distributions. The match items produced slightly higher activity, on average, than mismatch items, but critically, both types of items led to pattern completion. In contrast, in recognition memory (B), some studied items led to pattern completion (the hump on the right) whereas others failed to lead to pattern completion (the smaller hump on the left). Nonstudied items never led to pattern completion. Thus, the result was a thresholded distribution. C. ROCs, plotted using the average CA1 activity in the perception task and the memory tasks. The ROC in the perception task is curvilinear, whereas the memory ROCs are relatively linear. Note that in the memory tasks, overall performance is lower when a partial retrieval cue is used (vs. a complete cue), but in both cases, the resulting ROCs are relatively linear. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Hippocampus produces continuous strength distributions when used to make perceptual discriminations.

In contrast, the recognition memory simulations produced distributions that were consistent with a threshold process. Studied versus nonstudied item recognition distributions for the complete retrieval cue simulations are presented in Figure 3B, and the distributions for the partial retrieval cue simulations are shown in Figure C1 (Appendix C). In both cases, the studied items had a bimodal distribution, producing either strong activity (indicating pattern completion) or weak activity (indicating retrieval failure). In contrast, nonstudied items produced only weak activity (retrieval failure).

To verify that pattern completion corresponded to accurate item retrieval, we compared each EC output pattern with its respective target pattern using a match-mismatch rule (see Norman and O’Reilly, 2003; correct retrieval was assigned to scores greater than 0.67). Using a binary logistic regression, we found that CA1 activity predicted accurate item recognition (i.e., target versus lure) for all trials of the complete retrieval cue test; for the partial retrieval cue test, 99.8% of target trials and 100% of lures were accurately predicted.

ROC curves were then plotted to compare the performance characteristics of the perception and memory tasks, using activity from CA1 (Fig. 3C). The perception ROC plots the proportion of correctly identified matching items (“hit rate”) against incorrectly identified mismatching items (“false alarm rate”); using four mismatching features as the criterion for attributing activity to a “matching” pair was relaxed from the maximum observed activity to zero. Similarly, the recognition ROCs plot correctly identified target items (“hit rate”) against incorrectly identified lure items (“false alarm rate”) as a function of the criteria based on activity.

The perception ROC function was curved and had a y-intercept of 0; this shape is consistent with a continuous, signal detection process (Swets, 1988). By comparison, the memory ROC functions were relatively linear and had a non-zero y-intercept, consistent with a threshold process (Swets, 1988; Slotnick and Thakral, 2013). This latter pattern held for both the partial cue and full cue memory simulations.

To examine the roles of different hippocampal subfields in the perception task, we examined measures of perceptual sensitivity in CA1 and CA3. Sensitivity (d’ ) of the model to matching versus mismatching pairs was calculated as the difference in the mean activity for match (μact_match) and mismatch (μact_mismatch) trials divided by the average standard deviations (σ), as follows:

$$d' = \frac{\mu_{act\_match} - \mu_{act\_mismatch}}{\sqrt{\sigma_{act\_match}^2 + \sigma_{act\_mismatch}^2}}/2$$

1

Using this metric, we compared CA1 and CA3 as the number of mismatching features was increased from 1 to 36 (Fig. 4). The DG layer was not included in this analysis because it was inactive during retrieval mode, and the EC output was excluded because it is redundant as a strength signal in the model (i.e., its activity is determined only by CA1).

For the encode/retrieve model (Fig. 4A), CA1 sensitivity increased gradually as the number of feature differences was increased, and eventually tapered off. CA3 showed a similar trend, but was much less sensitive overall. Using the always-
encode model, CA1 demonstrated a similarly graded increase in sensitivity (Fig. 4B), while CA3 again showed weak sensitivity compared to CA1, and did not improve as the number of feature differences was increased.

The results of the simulations suggest that, whether one adopts the encode/retrieve or the always-encode model, CA1 is particularly sensitive to perceptual changes, and its sensitivity gradually increases as stimulus pairs become more different.

We then explored the network dynamics that produced continuous strength-based perception by examining how items were represented relative to one another as the number of mismatching features increased. To compare perception to previous findings in the memory literature, we examined the encode/retrieve model and plotted the average pattern distance—measured as the mean square error (MSE)—between mismatching representations for CA1 and CA3 (Fig. 5).

The distance plots reveal two interesting transitions that speak to both the continuous perception signal and the thresholded memory signal. First, CA1 shows a steady rise in pattern distance at the low end of mismatch (proportions of 0 to 0.4 features changed). This indicates that the two items initially overlapped completely in CA1, but gradually diverged as the number of feature differences increased. Thus, pattern similarity in CA1 tracks the objective similarity of the two items. Importantly, the shift in overlap corresponded with increasing strength-based sensitivity over the same period (see Fig. 4A, proportions of 0 to 0.4 features changed). In contrast, CA3 pattern distances remained minimal over the same period, indicating that CA3 was more likely to simply reactivate the pattern associated with the first item. Thus, pattern similarity in CA3 does not track objective similarity very well, and is less sensitive than CA1 to very small changes between stimuli (see Fig. 4A, proportions of 0 to 0.4 features changed).

The results of the simulations suggest that, whether one adopts the encode/retrieve or the always-encode model, CA1 is particularly sensitive to perceptual changes, and its sensitivity gradually increases as stimulus pairs become more different.

We then explored the network dynamics that produced continuous strength-based perception by examining how items were represented relative to one another as the number of mismatching features increased. To compare perception to previous findings in the memory literature, we examined the encode/retrieve model and plotted the average pattern distance—measured as the mean square error (MSE)—between mismatching representations for CA1 and CA3 (Fig. 5).

The distance plots reveal two interesting transitions that speak to both the continuous perception signal and the thresholded memory signal. First, CA1 shows a steady rise in pattern distance at the low end of mismatch (proportions of 0 to 0.4 features changed). This indicates that the two items initially overlapped completely in CA1, but gradually diverged as the number of feature differences increased. Thus, pattern similarity in CA1 tracks the objective similarity of the two items. Importantly, the shift in overlap corresponded with increasing strength-based sensitivity over the same period (see Fig. 4A, proportions of 0 to 0.4 features changed). In contrast, CA3 pattern distances remained minimal over the same period, indicating that CA3 was more likely to simply reactivate the pattern associated with the first item. Thus, pattern similarity in CA3 does not track objective similarity very well, and is less sensitive than CA1 to very small changes between stimuli (see Fig. 4A, proportions of 0 to 0.4 features changed).

Figure 5 also indicates that CA3 had two distinct retrieval states, delimited by a steep transition near the center of the similarity scale. On the left side of the transition, distances are small, indicating pattern completion. That is, the second item...
was drawn into the attractor created by the first item. In contrast, on the right side, much larger pattern distances indicate that CA3 was no longer reactivating the first item pattern (c.f. Yassa and Stark, 2011).

Because CA1 receives projections directly from CA3, it showed a similar transition as CA3 shifted from one state to the other. Critically, however, CA1 showed gradually increasing pattern distance with small stimulus differences, and it did so before CA3 started to discriminate similar items (i.e., < 0.4 proportion of features changed). This indicates that CA1 can produce a continuously graded signal that is diagnostic at low levels of mismatch, while CA3 is considerably less effective at discriminating between similar stimuli until a threshold is reached.

Why does CA1 show a more graded response than CA3, particularly given that CA3 is one of the major inputs to CA1? There are several reasons for this. First, because CA3 is subject to greater lateral inhibition than CA1, CA3 produces proportionately sparser activation. Greater inhibition reduces the probability of pattern overlap except at very high levels of input similarity, the result being a rapid transition between high and low overlapping states. In addition, recurrent excitatory projections within CA3 act to bind the elements of a stimulus representation together as a cohesive unit, together forming an attractor. When an input is sufficiently close to the attractor, activation spreads along the strengthened recurrent connections to reproduce the original pattern. In contrast, the lack of recurrent connectivity within CA1, combined with more diffuse activation patterns, supports more graded levels of pattern overlap and activation strength.

An interesting finding of the current simulations is that there was greater activity for matching than mismatching trials in the perception task. Why might such a pattern arise? When presented with the second item in a mismatch trial, CA3 pattern completes to the previous stimulus, and this signal is projected to CA1. Concurrently, CA1 receives information about the current, slightly mismatching stimulus from the EC input layer. Excitation of CA1 is therefore spread out over more units compared with a trial consisting of matching pairs. Although this might lead to the prediction of greater overall activity in CA1 for mismatching trials, competitive inhibition within CA1 counteracts such a result by dampening the activity of units that are less strongly excited (i.e., that are stimulated by EC alone or CA3 alone). In contrast, on matching trials, the same CA1 units are jointly and strongly reactivated by both EC and CA3, and inhibitory dynamics do not dampen the overall level of activity. (A simplified model that demonstrates the effect is included with the hippocampal model project file referenced in Appendix A.)

**DISCUSSION**

The involvement of the hippocampus in memory is well established, and neurocomputational models of the hippocampus have been shown to explain a wide range of extant memory data (e.g., Norman and O’Reilly, 2003; Elfman et al., 2008). Recent evidence suggests the hippocampus also plays an important role in high-level scene perception (e.g., Graham et al., 2010; Lee et al., 2005, 2012; Olsen et al., 2012; Warren et al., 2012; Aly et al., 2013). Whether existing computational models are able to account for these perception findings has not been established. To address this issue, we conducted simulations using a hippocampal model that was adapted from the CLS framework (Norman and O’Reilly, 2003) and found that the same network was able to reproduce findings from both memory and perception paradigms.

Consistent with the patient and neuroimaging work implicating the hippocampus in strength-based perception (Aly et al., 2013), the model produced graded, overlapping levels of activity that discriminated between matching and mismatching pairs in the perception task. The same model showed a thresholded activity function for recognition memory, in agreement with prior work indicating that the hippocampus is involved in recollection in long-term memory (Yonelinas et al., 2002, 2005; Ranganath et al., 2004; Montaldi et al., 2006; see Eichenbaum et al., 2007 and Yonelinas et al., 2010 for review). In addition, the continuous perception signal was most strongly related to mismatch magnitude in the CA1 subfield.

The finding that CA1 was the primary locus of the perceptual signal is consistent with a number of models in which CA1 acts as a “comparator” of perceived versus remembered events (Eichenbaum and Buckingham 1990; Hasselmo and Wyble, 1997; Meeter et al., 2004; Lisman and Grace, 2005; Kumar and Maguire, 2007; Yassa and Stark, 2011; Duncan et al., 2012). The current results extend this work by indicating that CA1 can also serve as a comparator in perceptual tasks. In the case of perception, the comparison is between a stimulus that is currently in the environment (or in the focus of attention), and one that was recently perceived (or recently attended).

Importantly, the current simulations with separate encoding/retrieval modes showed that the hippocampus can produce both continuous signals and thresholded signals without changing any properties of its architecture or basic mode of operation. The continuous output, however, was not dependent on having separate encoding and retrieval modes: a continuous perceptual signal was also observed in simulations in which the network was kept in an encoding mode. Thus, we need not assume that the hippocampus rapidly alternates between encoding and retrieval modes (e.g., Hasselmo et al., 2002) in order to support perceptual discriminations and output a graded signal.

**Why are There Thresholded Signals in Memory and Graded Signals in Perception?**

A critical point of success for the model is that we did not need to modify the way the network functioned—such as adjusting connectivity parameters—between the perception and recognition memory tasks in order for the network to perform...
both tasks successfully, or for it to produce the respective strength-based (continuous) and state-based (thresholded) signals. Rather, each signal emerged naturally as a result of the differing task demands; that is, the demand to differentiate between items with a high proportion of feature differences, as in the memory task, and the demand to differentiate between items with relatively few feature differences, as in the perception task.

The threshold nature of memory retrieval is not entirely attributable to any single subfield or parameter of the hippocampal model. However, one aspect of the model that is critical is the high level of inhibition within the hippocampal layers, particularly the DG, which leads to sparse activity and minimal representational overlap between encoded events. Consequently, only test cues that are very similar to a specific studied item will support successful retrieval. In addition, recurrent connectivity in CA3 produces nonlinear neural attractors that support a high level of pattern completion when a test cue is sufficiently similar to a stored pattern, but cues that do not match any specific studied item very well will effectively fail to lead to pattern completion. It is this “retrieval failure” that produces the thresholded hippocampal output.

In contrast, in our tests of perception, because the cue stimulus (i.e., the second item) is highly similar—or identical in the case of matching item trials—to the initial stimulus in the pair, the cue stimulus invariably leads to pattern completion of the initial item and produces an output that is highly correlated with the cue stimulus. There is variability in how closely the completed pattern matches the cue and this allows for the continuous signal that is useful in discriminating between match and mismatch trials. Importantly, because the two items in a perception trial are so similar to one another, the hippocampus is effectively always operating at a level well above its pattern completion threshold; thus, the perceptual signal reflects the continuous variability beyond the threshold.

The Role of the Hippocampus in Perception

A growing body of work suggests that the hippocampus is critical for perceptual decisions that require representations of detailed relational or spatial information (e.g., Graham et al., 2010; Lee et al., 2005, 2012; Olsen et al., 2012; Warren et al., 2012; Aly et al., 2013; see Lee et al., 2012 and Yonelinas, 2013). In the case of matching item trials—to the initial stimulus in the pair, the cue stimulus invariably leads to pattern completion of the initial item and produces an output that is highly correlated with the cue stimulus. There is variability in how closely the completed pattern matches the cue and this allows for the continuous signal that is useful in discriminating between match and mismatch trials. Importantly, because the two items in a perception trial are so similar to one another, the hippocampus is effectively always operating at a level well above its pattern completion threshold; thus, the perceptual signal reflects the continuous variability beyond the threshold.

The Role of Other Brain Regions in Strength-Based Perception

Is the hippocampus the only brain region that is able to perform computations that support strength-based perception, or can other regions also support these judgments? Aly et al. (2013) reported that selective damage to the hippocampus is associated with significant impairments in strength-based perception, indicating that the hippocampus itself makes an important contribution. However, strength-based perception was not completely eliminated in hippocampal patients, suggesting that other brain regions might also be capable of supporting this kind of perceptual decision. Moreover, an fMRI study provided evidence that regions outside of the hippocampus are related to strength-based perception, including the parahippocampal cortex, lateral occipital complex, and fusiform gyrus (Aly et al., 2013, 2014).

The current simulations showed how the hippocampus can support strength-based perception via Hebbian weight changes, and additional simulations indicated that an adaptation-based mechanism can produce comparable results in the absence of weight-based learning (Appendix D, Fig. D1). Can a model of cortex also support perceptual discriminations via Hebbian weight changes and/or adaptation-based mechanisms? To investigate these questions, we conducted additional simulations with a simple, two-layer model of cortex, based on the CLS model (Norman and O’Reilly, 2003). We found that, using either Hebbian weight-based learning or neural adaptation, the cortex could reliably discriminate between match and mismatch trials (see Fig. E1, Appendix E). These results are consistent with the above-mentioned hypothesis that, although the hippocampus makes an important contribution to strength-based perception, it may not be the only region that can support these judgments.

The findings with the cortical model raise an important question, namely, does the hippocampus play a role in strength-based perception that is distinct from other regions? We have suggested previously that the hippocampus is critical in supporting complex, high-resolution bindings (Yonelinas, 2013). That is, the hippocampus is critical in linking together the complex, qualitative aspects that make up events, and in a way that maintains high-resolution, detailed information (e.g., the specific color of the shirt you wore in a specific location at the specific color of the shirt you wore in a specific location at

Hippocampus
a specific time). Detecting the global changes used in the Aly et al. (2013) study would presumably benefit from such high-resolution, relational representations—representations that contain detailed information about the components of the scenes as well as their precise relations to one another. The hippocampus may contribute to strength-based perception by virtue of its representation of high-resolution, relational information; but other regions may contribute to strength-based perception in other ways. Although the manipulation used in Aly et al., (2013) is “global” in the sense of extending over the entire image, relatively local components are also affected. For example, “pinching” an image may alter the perception of individual objects within the image. Alternatively, global changes over the image may have effects on the perceived depth of the scene. The detection of these kinds of changes (i.e., to individual objects, or to image depth) need not require the hippocampus. Rather, visual or spatial representations in medial temporal, occipito-temporal, or parietal cortical regions may be sufficient. Thus, hippocampal damage reduces performance, but does not entirely eliminate accurate strength-based perception (Aly et al., 2013). If these hypotheses are correct, it should be possible to take the current scene perception task and reduce the extent to which the hippocampus will be critical by decreasing the complexity of relational bindings or by varying the extent to which the required discrimination relies on high-resolution (detailed) versus low-resolution (less precise) representations (see Yonelinas, 2013). Such studies are ongoing.

Limitations and Future Directions

A matter for future consideration concerns the direction of the relationship between hippocampal activity and match/mismatch detection. In the study by Aly et al. (2013), the level of hippocampal activity—inferrerd from the blood-oxygen-level dependent (BOLD) signal—was positively correlated with mismatch confidence. That is, the more that individuals were confident that a pair of items was different, the greater the hippocampal activity. By contrast, the current model showed greater overall activity for matching items than for mismatching items. The relationship between the BOLD signal, neural firing rates, and local field potentials is not yet fully understood (Ekstrom et al., 2009; Ekstrom, 2010; Logothetis, 2003), so the reason BOLD activity increased whereas activity in the model decreased cannot be confidently addressed. One possibility is that increases in the BOLD signal do not reflect increases in activation strength per se, but rather the time it takes the network to settle into a stable pattern. In fact, we found that the model took longer to settle for mismatching pairs, which is at least consistent with this explanation.

Another possible reason for the difference in signals is the inherent limitations in the algorithm used in the current simulations. Specifically, competitive inhibition was simulated as a fairly firm limit on the number of neurons that can be activated above a certain level (O’Reilly and Munakata, 2000). In our simulations, mismatching stimulus cues caused excitation to be spread out over more CA1 units relative to matching cues. With a more flexible inhibition rule it is possible that this would have led to an increase, rather than decrease, in overall CA1 activation.

Finally, another factor that may affect whether perceptual matches lead to an increase or decrease in neural activity is adaptation. Adaptation (or repetition suppression) is frequently reported in fMRI studies that involve repeating stimuli, and this adaptation has been observed in the MTL (Goh, et al., 2004; Howard et al., 2011; Diana et al., 2012; see Krekelberg et al., 2006). In these studies, presenting the same stimulus twice in a row is found to produce a diminished neural response (note, however, that repetition enhancement is also sometimes observed; see Turk-Browne et al., 2007).

In post-hoc simulations, we found that incorporating neural adaptation into the hippocampal model reversed the sign of match/mismatch difference, such that mismatching items were associated with greater activity than matching items (Fig. D1, Appendix D). That is, as the level of mismatch increased, activity also increased in CA1 (but not CA3, consistent with our simulations without adaptation). We also found that the network produced similar results with or without Hebbian learning, that is, using adaptation alone (data not shown). This suggests that activation-based mechanisms (adaptation), in contrast to weight-based mechanisms (Hebbian learning), might be an important component of how the hippocampus performs tasks in which information need not persist over long intervals. Nevertheless, the inclusion of adaptation did not improve the ability of the model to account for graded perceptual signals; CA3 alone was vastly less effective at performing adaptation-based mismatch detection compared with CA1, and CA1 was effective with or without the inclusion of adaptation. Importantly, CA3 did not produce graded signals in the simulations with adaptation. Nonetheless, future research examining the effects of neural adaptation on the model will be useful.

The current simulations lead to some potentially interesting predictions of how different hippocampal subfields might respond to changes in perceptual similarity that could be assessed using high-resolution fMRI. For example, by incrementally manipulating the number of mismatching features between pairs of items in a perceptual matching task, different subregions of the hippocampus should respond in comparatively unique ways. That is, based on the simulations, one would expect CA3 to show a consistent response at low levels of mismatch, indicative of pattern completion, and a fairly steep transition beyond some critical level of mismatch. Conversely, the CA1 subfield should exhibit a more gradual and continuous change in signal. Results that are broadly consistent with this were reported in a recent high-resolution fMRI memory study (Lacy et al., 2011), in which participants viewed a series of images that were either first presentations of an image, low-similarity lures, high-similarity lures, or repetitions of previous items. In CA1, activity was found to vary in a graded fashion with the level of change, whereas in CA3/DG, activity transitioned more steeply for small changes.

Hippocampus
CONCLUSIONS

Several lines of evidence suggest the hippocampus contributes to a range of cognitive functions, but how it performs seemingly disparate tasks under a single neural architecture is a question of great interest. The modeling work presented here provides an important step towards a unified understanding of hippocampal functioning across the domains of memory and visual perception, by showing how the hippocampal network is able to produce contrasting signals—that is, strength-based perception and state-based memory—using a common set of core assumptions. Future research will undoubtedly add to this emerging picture.

APPENDIX A

Basic Network Parameters

The following notes are a selective description of the rules and parameters used in the model simulations. Table A1 shows each layer size (i.e., number of units) and percentage of activity determined by the $k$-winners-take-all (Norman and O’Reilly, 2003). Table A2 shows the properties of the main projections, including the mean initial weight strengths (Mean), variances of the weight distribution (Var), relative strengths of the projections during encoding (Scale enc) and retrieval (Scale retr), and the proportions of receiving units that each sending unit is connected to (% Con). The learning rate was 0.1. The project file is available for download at http://goo.gl/4XkKya.

APPENDIX B

CA1 Activity Distributions Ranging 0–100% Feature Mismatch

We repeated the perception simulations with a broader range of match/mismatch levels, systematically varying the number of mismatching features from 0 (complete match) to 36 (complete mismatch). Activity distributions for CA1 across the varying levels of mismatch are presented in Figure B1. These simulations show that, as the number of mismatching features is systematically increased, CA1 activity distributions gradually shift from high to low activity. Moreover, with very high levels of mismatch, the “match” and “mismatch” distributions are essentially nonoverlapping, suggesting consistent pattern completion for “match” items and a failure of pattern completion for “mismatch” items. The latter finding is analogous to recognition memory simulations, in which pattern completion never occurs for nonstudied (lure) items (compare 24, 30, and 36 feature mismatch items in Fig. B1 to the nonstudied item distribution in Fig. 3).

TABLE A1.

<table>
<thead>
<tr>
<th>Layer/area</th>
<th>Units</th>
<th>Activity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entorhinal Cortex (in/out)</td>
<td>144</td>
<td>25.0</td>
</tr>
<tr>
<td>DG</td>
<td>1600</td>
<td>1.0</td>
</tr>
<tr>
<td>Area CA3</td>
<td>480</td>
<td>3.8</td>
</tr>
<tr>
<td>Area CA1</td>
<td>384</td>
<td>9.4</td>
</tr>
</tbody>
</table>

Note. In/out = input and output layers, respectively.

TABLE A2.

<table>
<thead>
<tr>
<th>Projection</th>
<th>Mean</th>
<th>Var</th>
<th>Scale (Enc)</th>
<th>Scale (Retr)</th>
<th>% Con</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC to DG, CA3</td>
<td>0.5</td>
<td>0.1</td>
<td>1</td>
<td>1</td>
<td>25</td>
</tr>
<tr>
<td>CA3 recurrent</td>
<td>0.5</td>
<td>0.1</td>
<td>2</td>
<td>2</td>
<td>100</td>
</tr>
<tr>
<td>CA3 to CA1</td>
<td>0.5</td>
<td>0.1</td>
<td>0.3</td>
<td>0.3</td>
<td>90</td>
</tr>
<tr>
<td>DG to CA3</td>
<td>0.9</td>
<td>0.01</td>
<td>15</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>EC to CA1</td>
<td>0.5</td>
<td>0.25</td>
<td>1</td>
<td>0.15</td>
<td>8.3</td>
</tr>
<tr>
<td>CA1 to EC</td>
<td>0.5</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>8.3</td>
</tr>
</tbody>
</table>

Note: Mean: mean initial weight strength; Var: variance of initial weight distribution; Scale: scaling of this projection relative to other projections; Enc: encoding mode; Retr: retrieval mode; % Con: percentage connectivity; EC: entorhinal cortex; DG: dentate gyrus.

FIGURE B1. Perception strength distributions of CA1 activity as the number of feature differences is varied from 0 to 36 (of 36), at intervals of 6. CA1 produces continuous strength distributions that gradually move farther apart as the number of feature differences is increased. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
APPENDIX C

Partial Retrieval Cue Memory Simulation

APPENDIX D

Match/Mismatch Sensitivity With Adaptation

The simulations reported in this paper generally showed greater activity for matching versus mismatching item pairs in the perception task. However, Aly et al. (2013) found evidence for the opposite effect (i.e., greater hippocampal activity for mismatching versus matching item pairs). Additional simulations were conducted to test the hypothesis that neural adaptation can account for this discrepancy. Figure D1A shows distributions of average activity in CA1. As expected, mismatching pairs are now associated with an increase in the overall level of activity. Figure D1B compares match/mismatch sensitivity for CA1 and CA3 as mismatch was varied from 0 to 100%. CA3 was comparatively very poor at discriminating between matching and mismatching pairs.

Adaptation was included using the “accommodation” parameters of cell activation from the leabra algorithm. There was a firing-rate threshold of onset of .1 and an update rate of 0.7. Computations were made at the end of each trial.

APPENDIX E

Match/Mismatch Strength in the Cortical Model

Additional simulations were run to test whether a simple two-layer model of the cortex (comprising an input layer and a neocortical, or hidden, layer of 800 units) could perform the perception task. Simulations were run using Hebbian-based learning (Fig. E1A) with a learning rate of 0.004, and separately using the adaptation parameters listed in Appendix D (Fig. E1B). Using four feature differences for the mismatching pairs, both approaches were successful in discriminating between matching and mismatch item pairs, and produced overlapping, Gaussian-like distributions that were comparable to those observed in area CA1 of the hippocampal model.

FIGURE D1. Hippocampal network performance for the match/mismatch perception task with the inclusion of neural adaptation (using the encode/retrieve model). A: Distributions of average activity levels for CA1 for matching items and mismatching items with 4 and 8 features changed. B: Sensitivity (d') of CA1 and CA3 as the number of mismatching features was varied from 1 to 36. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
FIGURE E1. Match/mismatch task performance using a two-layer cortical model. The input layer is identical to the input layer for the hippocampal model; the neocortical layer is composed of 800 units with 20% active at a given time. A: Distributions of average activity levels with four mismatching features, using Hebbian weight-based learning. B: Distributions using adaptation and no weight-based learning. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

REFERENCES


Hippocampus
Yonelinas AP. 2013. The hippocampus supports high-resolution binding in the service of perception, working memory and long-term memory. Behav Brain Res 254:34–44.
Chapter 6: Discussion

The current thesis sets out to further our understanding of human memory and related cognitive processes by applying a combination of computational and behavioral research approaches. This combination was found to be very effective and often provided a more nuanced and useful explanation of various memory and cognitive phenomena than might have been possible with either approach in isolation. In this chapter, summaries of the key experimental findings from the papers making up this thesis are presented. This is followed by a critique of the strengths, weaknesses, successes, and shortcomings of the methodologies used, and how the current findings speak to existing controversies and debates. The dissertation is concluded with a commentary that considers the research as a cohesive body of work. Throughout the discussion, signposts and recommendations for future research are highlighted.

6.1 Summary of Findings

6.1.1 Predicting the Breakdown of the Recollection Threshold

The first paper, *Testing a Neurocomputational Model of Recollection, Familiarity, and Source Recognition*, had three primary aims. The first was to examine whether the complementary learning systems (CLS) model (McClelland, McNaughton, & O’Reilly, 1995; Norman & O’Reilly, 2003), a prominent computational model of the medial temporal lobes, can account for the types of receiver operating characteristics (ROCs) observed in various item and source recognition memory studies. The second aim was to describe the types of ROCs that the model typically produces to confirm that it is generally consistent with extant behavioral ROCs. And the final aim was to test a novel set of predictions obtained from model simulations, which suggest circumstances under which source ROCs will be linear and when they will become more curved. That is, we found that the model makes specific predictions about when the hippocampus should produce thresholded signals that in turn produce thresholded recollection and when the threshold should break down to produce a more continuous signal.
For the initial simulations we examined the types of strength distributions and ROCs produced by the hippocampal and neocortical components of the CLS model. The hippocampal component, which is heuristically and theoretically tied to the construct of recollection produced thresholded strength distributions, which translated to linear ROCs (and U-shaped zROCs). In contrast, the cortical component of the model, which is theoretically linked to familiarity, produced Gaussian strength distributions and clearly curved ROCs (and linear zROCs). The results are consistent with various quantitative dual process models that assume that recollection is quantifiable as a threshold process, and familiarity as a signal detection process (e.g., Kelley & Wixted, 2001; Sherman et al., 2003; Yonelinas, 1994, 1999, 2001). However, the CLS model went beyond such quantitative models by showing how the threshold and signal detection properties of recollection and familiarity could arise from the structural and functional neural properties of the hippocampal and neocortical networks.

To the extent that the simulations produced ROCs that are consistent with cognitive, dual process models, the CLS model was found to be in good agreement with the existing ROC literature. Specifically, the results were consistent in showing that source recognition, which in the behavioral literature has been found to rely more heavily on recollection than familiarity, produced ROCs that were very linear, and this result was specific to the hippocampal network. In contrast, item recognition, which has been shown to rely more heavily on familiarity, produced ROCs that were more curved. This effect was tied to the cortical network.

The final goal was to test a novel prediction of the CLS model concerning what happens in the hippocampus when it is presented with many stimuli that share a large proportion of features. It was shown that under ordinary list task conditions, the hippocampal network operates as a threshold retrieval process, and thus it provides a viable explanation for the linear ROCs (U-shaped zROCs) seen in human source recognition tasks. However, Norman and O’Reilly (2003) found that the hippocampal component does not always exhibit this threshold nature. Specifically, they conducted simulations of an item recognition paradigm in which old and new items shared either a low or high percentage of prototypical features (i.e., the items were all very similar to each other or they were very distinct). They found that
greatly increasing the feature overlap between items significantly hindered the ability of the hippocampal model to assign distinctive representations to each item, making stimulus representations more prone to interference.

In order to verify that this pattern of results is reliably predicted by the model, and is specifically applicable using a source memory paradigm, we conducted several simulations in which we similarly manipulated the degree of feature overlap. The results showed a decline in overall model performance when overlap was high, but more importantly, instead of lures reliably failing to trigger a response from the hippocampus, there were now overlapping old–new item strength distributions, that each showed continuous, Gaussian-like shapes. Having determined that this was a reliable prediction of the model, several behavioral experiments were devised to put the model to the test.

Three experiments were designed to manipulate feature overlap in different ways, including: increasing the number of words from one semantic category (i.e., animal names, Experiment 1); increasing the number of pictures of one type (i.e., house pictures, Experiment 2); and decreasing the opportunity to fully sample the item features of each picture (Experiment 3). The predictions of the model were validated by all of these experiments. That is, highly overlapping stimuli produced source ROCs that were more curved (and more linear in z-space) than minimally overlapping stimuli, even when overall level of performance was controlled. These differences in the shape of the ROCs supported the CLS model’s prediction of a shift from threshold to Gaussian-like strength distributions.

The change in the shape of distributions and increased overlap is best understood in terms of a failure of pattern separation. That is, when feature overlap is low, the patterns of activity in the DG and CA3 associated with each item are highly distinct from one another. Consequently, test cues will either lead to pattern completion of the correct source, or else will fail. However, if similar items being tested are all very similar, the ability of the hippocampus to pattern separate can be exceeded. That is, CA3 no longer forms distinctive representations for each study item. Thus, the hippocampus always pattern completes, and the threshold – which is usually highly diagnostic – is nullified. The resultant target and lure distributions are highly overlapping.
In addition to validating the predictions of the CLS model, the behavioral findings of this study also weigh in on a long-standing debate over the validity of dual versus single process recognition memory models. In particular, it is difficult to explain the current findings in the context of a single process paradigm. These implications are discussed later in this chapter.

Overall, the CLS model was found to provide a viable account for typical recognition memory ROCs, producing thresholded recollection distributions and Gaussian familiarity distributions consistent with various quantitative ROC models. In addition, a novel prediction of the CLS model, that the threshold nature of recollection should begin to break down as feature overlap is increased, was supported in new source recognition experiments. The findings from this paper provide strong support for the CLS model of recognition memory and demonstrate the power of a combined computational–behavioral research approach.

6.1.2 Complexity Affects the Characteristics of Recollection

The second paper, Variations in Recollection: The Effects of Complexity on Source Recognition, had two primary aims. The first aim was to determine whether increasing the complexity of stimuli in a source memory experiment leads ROCs to become more curved (and corresponding zROCs to become more linear). Contingent on the results of this inquiry, the second aim was to determine whether the change in ROC form is due to recollection, familiarity, or both. In addition, a computational model of the effects of complexity on recognition memory was proposed, which was adapted from the hippocampal component of the CLS model. The simulation results were briefly reported in the paper’s discussion section to explain why the hippocampus might respond to changes in complexity in the manner observed in our experiments, and is expanded on here.

For the behavioral experiments, the simplest condition, audio-word, matched the typical source conditions that are found to produce approximately linear ROCs, and the most complex condition, video-sentence, matched the richer source conditions found in Qin et al.’s (2001) study. Experiments 1a and 1b showed that increasing the complexity of studied items or increasing the complexity of source information can lead to more curved ROCs and flatter zROCs. In addition, while
both stimulus and source complexity were found to affect the shape of ROCs, the most dramatically curved ROCs were found when both item and source complexity were increased together. Increasing complexity generally led to an increase in overall performance in Experiment 1a, particularly when both stimulus and source complexity increased together. Experiment 1b addressed the issue of unequal performance by examining the effects of complexity while controlling for levels of overall performance. The results showed the same pattern, with relatively linear ROCs in the simple condition, and more curved ROCs the complex condition.

Experiment 1 showed that complex study events can lead source ROCs to become more curved, but they do not indicate why differences in complexity produce differences in ROC shape. It could be that source memory judgments in complex conditions rely more on familiarity if the richer information allows the participants to encode the item and source in an integrated manner. Previous work has indicated that familiarity can support source and associative recognition performance if the item and source information are unitized or integrated during study (e.g., Diana et al., 2008, 2010; Quamme, Yonelinas, & Norman, 2007; Yonelinas, 1999). Thus, encoding sentences from a video of the speakers may provide a richer study event that more readily allows an integration of the item and source information, thereby increasing the utility of familiarity in source tests. Another possibility is that the increased curvature of the ROC might occur because increasing complexity causes recollection to behave in a more continuous and Gaussian manner (Qin et al., 2001; Yonelinas, 1999).

To test these alternative accounts, we examined the effects of divided attention (Experiment 2a) and amnesia (Experiment 2b) on source discriminations for the simple and complex source conditions from Experiment 1b. In Experiment 2a, we examined divided attention at test because previous work has indicated that recollection is particularly sensitive to the disruptive effects of divided attention (e.g., Gruppuso, Lindsay, & Kelley, 1997; for a review, see Yonelinas, 2002). So, if the differences seen in the ROCs in the previous experiment were due to differences in the nature of recollection, but the contribution of recollection was the same, then dividing attention should reduce performance equally in the simple and complex conditions. In contrast, if the exaggerated curve seen in the complex condition was
due to an increased reliance on familiarity in that condition, then dividing attention should have a smaller effect in the complex condition than in the simple condition. We applied similar logic in Experiment 2b by comparing the performance of hypoxic patients with that of healthy controls on the simple and complex source tasks.

For the divided attention test (Experiment 2a), the results across the simple and complex memory tests did not differ significantly, and there was no significant interaction between test type and attention condition. The similarity in the effects of divided attention on the simple and complex conditions is most consistent with the idea that recollection was supporting source decisions to the same degree in both conditions. Had the increase in curvature found in the complex condition ROCs been due to familiarity, performance in that condition would have been less vulnerable to the effects of divided attention. Thus, the patterns in Experiment 2a suggest that recollection acts in a thresholded, probabilistic manner when stimuli are relatively simple, but becomes more signal detection-like, or continuous, when stimuli are more complex.

For the patient versus control test there were only two patients so statistical analysis was limited. However, the overall performance of the two patients was significantly lower than that of the control group in each condition but there were no differences in performance as a function of condition. The results suggested that the combination of recollection and familiarity that supports source memory in the simple condition is about the same as that which supports source memory in the complex condition. The shapes of the ROCs and zROCs supported the conclusion that the amnesic patients’ deficits were comparable across conditions, suggesting that the tasks were equally dependent on recollection. DPSD model estimates were also consistent with this conclusion, showing no evidence that the amnesics were able to rely more on familiarity in the video-sentence condition than in the audio condition. Thus, the results were consistent with the claim that recollection drives performance in both simple and complex conditions and support the findings of the divided attention experiment.

To determine whether the CLS model could account for the current results, the simple and complex conditions from the current study were simulated to test
whether the model would accurately predict that increasing complexity would result in more curved source ROCs. The simulations were conducted with the hippocampal network in isolation. In the simple condition, an item, $A$, was represented as a string of features (e.g., color, shape, size) that was paired with a second string of features that represented a source, $X$. In the complex condition, each item was made up of multiple components (e.g., $AB$), and each source was made up of multiple components (e.g., $XY$). The network then had to learn two pairings (e.g., $A-X$ and $B-Y$) to encode all the features. This piecewise approach was intended to capture the fact that as complexity increases, the number of specific associations that the participant can form for a given item–source pairing will increase. To match overall performance across conditions, the item–source pairings in the simple condition were encoded twice each. The source decisions were modeled by presenting an item and allowing the hippocampal network to pattern complete the missing source information.

The strength distributions produced in the simple condition were consistent with the results of previous simulations; namely, the model produced “high confidence” correct source responses, but it also showed a large proportion of retrieval failures. The resulting ROCs exhibited a large linear component, consistent with the empirical results. Most importantly, the once threshold-shaped activation strength distributions took on a more Gaussian-like shape in the complex condition, and the resulting ROCs were more curvilinear, as was seen in the empirical study. Notably, the model also correctly predicted that increasing complexity results in improved source performance, despite the breakdown in the recollection threshold (evidenced by the fact that it was necessary to present “simple” items to the network twice to achieve matched performance). The fact that the hippocampal model was able to reproduce the observed behavioral ROC effects, even though the current paradigm was not an \textit{a priori} purpose of its design, attests to the power of the computational modeling method.

In summary, the results from this study suggest that recollection generally leads to linear ROCs, but when materials become highly complex, recollection can produce more curved ROCs. One might argue that the failure to find effects of divided attention and amnesia on ROC shape reflected a lack of statistical power.
However, both of these manipulations had sizable effects on overall recognition performance, and there was no indication in the results that the complexity manipulation influenced familiarity. Future studies using convergent methods to verify these conclusions will be useful, but the current results certainly favor the graded recollection account of complexity over a familiarity account. In addition, the results of modeling complexity using the hippocampal network from the CLS model suggest a compelling explanation of how complexity affects the recollection-based retrieval process at the neurocomputational level.

### 6.1.3 Similarity Gradients of Recollection and Familiarity

In natural settings one can reasonably assume that the same stimulus will never appear, or be perceived, in exactly the same way twice, so the ability of human memory processes to be flexible to differences in stimulus cues and at the same time perform optimally is a critical function. In the third paper, *Recollection and Familiarity Exhibit Dissociable Similarity Gradients: A Test of the Complementary Learning Systems Model*, we set out to explore how memories are retrieved when the cues are less than perfect. This began with simulations using the CLS model, the aim of which was to determine the predicted similarity gradients (i.e., comparing the original stimulus and a current test item) of recollection and familiarity by examining the hippocampal and neocortical networks while varying similarity. Once similarity gradients had been obtained, two behavioral experiments were conducted with the aim of testing the predictions of the model and whether, in practice, recollection and familiarity can be dissociated based on their respective similarity gradients.

For the simulations, each network was trained on a set of items and was then probed using retrieval cues that were parametrically varied from very different to identical to the studied items. The similarity gradients of each of the two networks were first examined across single trials, and then similarity functions were obtained by averaging across items. Finally, the outputs of the two networks were combined, based on the dual processes signal detection formulation (Yonelinas, 1994), to determine how overall recognition performance would be influenced by stimulus similarly. That is, for a given trial the ‘recollection’ value was used if the
hippocampal output exceeded threshold, and failing that the ‘familiarity’ value from the neocortical network was used.

The single trial results indicated thresholded activation functions in the hippocampus, showing essentially no output signal until a threshold of study-test similarity was reached, at which point the network transitioned to a strong output activation state. Importantly, retrieval was highly accurate in the sense that when the threshold was reached, the retrieved representation matched, with few exceptions, the correct studied item. In contrast to the hippocampus, the neocortical network exhibited relatively linear functions, such that activation increased gradually as a cue became more similar to a studied item.

Averaging across trials, the threshold nature of the hippocampus was less obvious because different trials transition at different points. However, the similarity gradient was nonetheless steeper than that of the neocortex, transitioning over a smaller range of similarity levels. The results suggested that it should be possible to observe differences in the similarity functions obtained from behavioral measures of recollection and familiarity, even when averaging across trials.

The combined hippocampal and neocortical ROCs that were constructed from the simulation data were curved and asymmetrical, similar to those observed in typical human recognition memory studies (for review see Yonelinas & Park, 2007). In addition, as expected, as similarity was increased, performance also increased and the ROCs moved upward along the $y$-axis. Based on the simulation results, for the behavioral experiments it was predicted that recollection would exhibit a nonlinear function with a steeper gradient than familiarity, whereas familiarity was expected to be comparatively more linear, increasing gradually for studied items over the course of the trial.

To examine behavior-based similarity functions in humans, in the first experiment participants were required to study lists of computer-generated face images, and were then tested on items that were manipulated over sequential presentations to have increasing similarity to their respective studied versions. Each test item was gradually morphed from a generic, non-studied face to either a studied face or a non-studied face. Recognition confidence was assessed at each level of stimulus similarity, and then, once the target face was presented, participants made a
‘remember’ response if the face was recollected, a ‘know’ response if the face was recognized as old on the basis of familiarity in the absence of recollection, or a ‘new’ response if they believed the face had not been studied. Thus, we could measure recognition confidence as a function of stimulus similarity separately for items that were ultimately recollected or just familiar. Experiment 2 was similar except that each studied face was only tested once at a random morph level and there were no remember/know judgments.

The results for Experiment 1 verified the prediction of the CLS model, showing that recollection exhibits a steeper similarity gradient than familiarity. As test stimuli were morphed to a studied face, recognition confidence increased more gradually for items recognized on the basis of familiarity than those recognized on the basis of recollection. To determine whether the shapes of the similarity functions were different for remember and know responses, the observed responses were fit to sigmoidal functions. The resulting functions indicated that the shapes of the ‘remember’ and ‘know’ gradients were significantly different, even when accounting for differences in final level of performance. To further verify that the shapes were different, the analysis was repeated for ‘remember’ and ‘know’ trials that were approximately equal in final confidence, thus matching for memory strength.

The second experiment was designed to test whether the same pattern of results would be observed under conditions in which similarity was not incrementally morphed within single trials. That is, perhaps the similarity functions were affected by having each test item morph across contiguous presentations within a single trial. In addition, the fact of having multiple observations for single items violates assumptions of independent observations and thus limits the range of statistical tests we could use. To test the generalizability of the results from Experiment 1, in Experiment 2 each test item was only tested once and cue-target similarity was varied across items. In addition the data was fit to a model that incorporated the DPSD framework.

The results were consistent with the remember/know results of the first experiment and the CLS model simulations, showing relatively steep, nonlinear recollection-based functions for the hippocampus compared with more linear
familiarity-based functions for the neocortex. In addition to confirming that recollection had a significantly steeper gradient than familiarity, the analysis showed that familiarity had a greater impact on recognition at low levels of similarity compared with recollection. The average ROCs also closely matched the simulated ROCs from the CLS model. That is, familiarity—indicated by the level of ROC curvature—increased gradually with similarity, consistent with the neocortical signal. In contrast, recollection—indicated by the y-intercept—exhibited a nonlinear change, increasing more dramatically at middle similarity intervals, consistent with the hippocampal model data.

Overall, the similarity functions derived in this study showed that recollection and familiarity are dissociable in a way that had been relatively unexplored in the behavioral literature. Applying a logistic (or sigmoidal) model to the recognition data informed us that the two processes display distinct similarity gradients with quantifiable differences. Importantly, based on cognitive models alone, it was neither an obvious nor fundamental assumption that recollection and familiarity should demonstrate these reliably different similarity gradients. However, the CLS model provided not only the prediction, but also a sound theoretical basis for the current experiments. Thus, the findings join the first two papers in validating a set of a priori predictions of the CLS model with empirical tests.

6.1.4 Memory and Perception Signals in the Hippocampus

The final paper, *A Neurocomputational Account of Memory and Perception: Thresholded and Graded Signals in the Hippocampus*, examined recent evidence suggesting that the hippocampus, in addition to its role in memory, plays a critical role in high-level scene perception (Aly et al., 2013; Graham et al., 2010; Lee et al., 2005; Lee et al., 2012; Olsen, et al., 2012; Warren et al., 2012), and presented simulations that suggest how the hippocampus can account for both types of cognitive functioning. This problem was vexing at the outset because the hippocampal signal that is associated with memory functioning is quite different to that found in response to perception tasks. Specifically, an established literature indicates that the hippocampus produces relatively thresholded signals during
memory-dependent tasks (see Eichenbaum et al., 2007; Yonelinas et al., 2010 for review), but recent evidence indicates that the hippocampus also supports strength-based, or continuous, forms of perception (Aly et al., 2013). The paper began with an overview of how computational models of the hippocampus account for thresholded memory retrieval. Then, an outline of the recent evidence implicating the hippocampus in high-level perception was presented, highlighting how the type of signal associated with perception differs from memory, and the challenge this poses for existing models. Finally, the results of simulations from a computational model of the hippocampus were presented, showing how the network can in fact account for both thresholded and graded signals.

For the perception simulations, study lists were created from randomly generated items. Half of the items were assigned to a ‘match’ group, and the other half to ‘mismatch’. Matching pairs simply comprised two identical item patterns. Mismatching pairs were created by pseudo-randomly changing one or more (depending on the condition) features in the second item of the pair, ensuring the same units were not reselected. The two items making up each trial were presented sequentially to the network.

Two simulation approaches were examined for perception; an encode/retrieve model and an always-encode model. For the encode/retrieve model, an item was presented to the network in encoding mode and the network was switched to retrieval mode for the second (matching or mismatching) item. In taking this approach, we used the network in a way that is comparable to typical recognition simulations (Norman & O’Reilly, 2003). The always-encode model differed in that both items were presented to the network in encoding mode. We included the latter approach in order to ensure that the results would replicate if one assumes that perception does not involve any kind of (even very short-term) ‘retrieval’. Separate simulations were run for 36 levels of feature mismatch (effectively ranging from identical to chance similarity). The assessment of performance in a trial was based on the overall output activation (at CA1) of the second item in the pair. The CA1 layer was used as it supports comparatively more fine-grained activation levels than other layers in the model (in particular, the entorhinal cortex output layer).
For the recognition task, 20 randomly generated items—the same as the match/mismatch simulations—were used for the study list. The items were sequentially presented to the network in encoding mode. The test list comprised the 20 study items and 20 new, randomly generated lures. Half the features in each test item were left blank so as to probe memory retrieval performance (see Figure 2B). The network was switched to retrieval mode for the test phase, and the assessment of performance in a trial was again based on the overall output activation of CA1 at time of retrieval.

Consistent with evidence from patient and neuroimaging work implicating the hippocampus in strength-based perception (Aly et al., 2013), the model produced graded, overlapping levels of activation that discriminated between matching and mismatching pairs on the perception task. In addition, the same model showed a thresholded activation function when subjected to a recognition memory test. Also, the strength-based sensitivity of the hippocampus to matching versus mismatching pairs increased continuously with the number of feature differences. The perception signal was most strongly correlated with mismatch size in the CA1 subfield compared with other areas, and this was the case for separate encoding and retrieval modes as well as the encoding-only mode.

A critical point of success for the model is that we did not need to modify the way the network functioned, such as adjusting parameters, between the perception and recognition tasks in order for the network to perform both tasks successfully, or for it to produce the respective strength- and state-based signals. Rather, each signal emerged naturally as a result of the differing task demands. For memory, discrete activation states occurred because, for some items, the network was able to pattern complete and thus retrieve the correct pattern, while for others, the network failed. And for perception, pairs of items were typically very similar and occurred together, so the network never failed to pattern complete (except at high levels of mismatch). Instead, performance was more subtly strength-based and resulted primarily from the ability of the CA1 subfield to represent differences continuously.

The finding that CA1 was the primary locus of the perceptual signal is consistent with a number of models in which CA1 is hypothesized to act as a “comparator” of perceived versus remembered events (Duncan et al., 2012;
Eichenbaum & Buckingham 1990; Hasselmo & Wyble, 1997; Kumaran & Maguire, 2007; Lisman & Grace, 2005; Meeter, Murre & Talamini, 2004; Yassa & Stark, 2011). This idea is conceptually compatible with our approach as it incorporates the comparison of separate, but co-occurring input and retrieval patterns; that is, from the entorhinal cortex and CA3, respectively. In addition, the level of mismatch is correlated with CA1 activation strength. Importantly, the continuous output associated with perception was not dependent on having separate encoding and retrieval modes. That is, additional simulations in which the model remained in encoding mode also produced a continuous signal at CA1. Thus, we need not assume that the hippocampus rapidly alternates between encoding and retrieval modes (e.g., Hasselmo et al., 2002) in order to support perceptual discriminations and output a graded signal.

In summary, several lines of evidence suggest the hippocampus contributes to a range of cognitive functions, but how it performs these disparate tasks under a single neural architecture is a question of great interest. The modeling work presented in the final paper provided an important step towards a unified understanding of hippocampal functioning across the broad domains of memory and perception, by showing how the network is able to account for contrasting signals—i.e., strength-based perception and state-based memory—using a common set of core assumptions.

6.2 A Critical Review of Methodology and Frameworks

6.2.1 On the Use of Cognitive Models

Cognitive models attempt to quantify psychological constructs—as they are manifest, behaviorally—without directly commenting on how or where they arise in the brain. In this thesis, dual process models of recognition memory—and in particular, the dual process signal detection (DPSD) model—were used extensively as the basis for numerous hypotheses and experimental protocols, and proved to be a powerful theoretical framework. To recap, the DPSD model characterizes recognition memory as the product of two independent memory processes; familiarity, which is a continuous-strength, signal detection process, and
reollection, which is a threshold process whereby qualitative information about a specific event is retrieved, or else retrieval fails (Yonelinas, 1994).

In the first paper of this thesis, we began by asking whether simulations using the complementary learning systems (CLS) computational model can produce ROCs that are consistent with the DPSD cognitive model. This was an important first step because there is extensive empirical evidence that the DPSD model is robust across a range of tests (for a review, see Yonelinas, 2002). Then, when it was shown, via the CLS model, how and when the recollection threshold could be expected to break down, we were able to quantify changes to the recollection threshold in humans with respect to the DPSD model, and in terms of how the findings deviated from the model. The DPSD model was similarly instructive in the second paper—investigating stimulus complexity—and the third paper—analyzing similarity gradients—by helping to frame specific research questions in terms of the effects of different encoding and retrieval conditions on the model’s parameters, and as a paradigm for data analysis. For example, given evidence that ROCs become more curved with greater stimulus complexity, one possible implication ascribed by the DPSD model is that familiarity—which is associated with curved ROCs—plays a greater role in memory performance when stimuli are more complex. Our experiments ultimately showed that the effect was not likely the result of a greater contribution of familiarity and a reduced contribution of recollection; rather, it appeared that recollection was violating the threshold assumption described by the DPSD model and was behaving more like a continuous memory-strength process (the implications of which are discussed below). In the third paper, we used the model to further understand and contrast the processes of recollection and familiarity as the level of similarity between studied items and their associated test cues was varied. And finally, the fourth paper presented recognition memory as a dual process complement to a separate, dual process perception model.

A trend that surfaced throughout these studies is that the rules that characterize the cognitive processes thought to underlie recognition memory—as measured using ROC plots—are not consistent across all experimental paradigms. Arguably, this discovery is particularly important in the grander scheme of seeking to understand memory functioning in realistic settings because, when we leave the
controlled space of a computer-based experiment, many complexities are introduced—stimuli are frequently confusable (i.e., have high feature overlap), and episodes are often very complex—considerations that were directly addressed by the research. Importantly, what may be seen as “violations” of the DPSD or other dual process models should not be interpreted as evidence refuting their validity. Rather, the research shows that there are highly predictable conditions under which the functions used to describe memory will deviate from their standard formulation.

Could another model have done a better job of characterizing the collective research findings, perhaps without violating its own assumptions? As highlighted in the first two papers, there exists a long-running debate over the “true” shape of recognition ROCs and the cognitive processes that give rise to them. This debate has been fueled in part by a diverse array of ROC functions reported in the literature, varying from the very linear to the highly curved (for a review, see Yonelinas & Parks, 2007). Moreover, this debate has proven difficult to resolve due, in large part, to the ease with which a plurality of models can be convincingly fit to many of the data sets (see Koen, Aly, Wang & Yonelinas, 2013). Superficially, the findings reported in this thesis further complicate matters because, as discussed above, they imply that memory processes are malleable, presenting different characteristics depending on the demands of the task at hand. As stated earlier, such findings do not undermine the validity of the DPSD model, but they do highlight boundaries within which the model formulation can be strictly applied.

Fortunately, a number of competing models lose viability when we compare the data with their more nuanced predictions. For example, some early models posit that there is only a single, signal detection recognition process (e.g., Donaldson et al., 1996; Dunn, 2004; Wixted, 2007). These models dictate that the functions underlying item and source memory ROCs are always curved and intercept at 0 on both the x- and y- axes. In contrast, the DPSD model predicts that for source memory tasks performance is heavily reliant on threshold-based recollection, and so the corresponding ROCs are expected to be linear and have a non-zero intercept. Thus, the numerous curved source memory ROCs presented in our experiments ought to provide an opportunity for alternative models to prevail over DPSD. However, an additional prediction of pure signal detection models is that z-score
ROCs are always linear. The numerous zROCs presented in the first two papers, under various source memory conditions, were always curved in a U-shaped direction, indicating the presence of a retrieval threshold. Importantly, even if the signal detection models are modified to include separate Gaussian strength measures for recollection and familiarity (Rotello, et al., 2004; Wixted, 2007), they still fail to produce the types of U-shaped zROCs so often reported in source memory tests.

Other models that, like DPSD, include separate threshold and continuous processes (e.g., DeCarlo, 2002; Kelley & Wixted, 2001; Squire & Zola, 1998) do a better job of fitting the current data. For example, the variable recollection dual process (VRDP) model (Sherman et al., 2003) is based on the DPSD model, but assumes that items that exceed the recollection threshold form a Gaussian strength distribution that is described by separate strength and variance parameters. In contrast, the DPSD model does not make any claims about the form of the above-threshold distribution; only that it supports high confidence responses. The VRDP model technically could account for curved source ROCs by assuming that the apparent breakdown in threshold actually reflects a weaker and more variable recollection distribution that spans multiple confidence bins. A weakness of the VRDP model, however, is that it is comprised of four free memory parameters (as opposed to DPSD’s two) and thus tends to be under-constrained. That is, when the model is fit to ROC data, there are often multiple good solutions, so it is not possible to provide a statistically meaningful assessment of the model. Critically, while complex models such as VRDP can in principle account for variations in the shapes of source memory ROCs, they are unable to make such predictions a priori.

In many situations, the DPSD model has proven very effective in predicting the outcomes of recognition memory experiments (Yonelinas, 2002). And in the current thesis, it very effectively characterized the functions of recollection and familiarity at various levels of cue-target similarity. In the first two papers, where the DPSD model appeared to fall short, the CLS model had predicted, a priori, such atypical memory performance effects. For example, simulations using the CLS model showed how the hippocampus typically produces threshold retrieval functions in response to test cues, and it also showed how the threshold breaks down when pattern separation fails. Thus, the DPSD model is consistent with computational
models that describe separate recollection- and familiarity-like processes, within predictable boundary conditions.

6.2.2 On the Use of Computational Models

Neurocomputational models are typically designed to simulate cognitive functions, or neural activity, within a particular scope. For example, Hasselmo and Wyble (1997) used rate code models of hippocampal neurons to model performance on free recall and cued recall tasks. Norman and O’Reilly (2003) used a similar approach to model recognition-based memory in the hippocampus and cortex. Other focuses have included modeling the mental time-travel aspect of episodic memories in the hippocampus (Hasselmo, 2009), the involvement of the hippocampus in short-term memory (Burgess & Hitch, 2005), and other researchers have devised models to investigate how spatial coding in the hippocampus is used to guide behavior (Eichenbaum, Wiener, Shapiro & Cohen, 1989). In attempting to model the neural bases of cognition and behavior, model complexity can quickly get out of hand. Thus, a specific focus with limited adjustable parameters is often necessary for a model to remain tractable and therefore useful. One could argue then, that a good model will explain a wide range of data while making few core assumptions. This was found to be the case with the complementary learning systems (CLS) model (McClelland et al., 1995; Norman & O’Reilly), the computational framework used for the neural network simulations in this thesis. Using the CLS model, we were able to move beyond some of the quantitative characterizations of recognition and perception offered by cognitive models, and explain how such processes are likely to arise from particular patterns of neural connectivity and activity in the brain. Critically, this approach also led to novel, testable predictions, as highlighted throughout the papers.

Fundamentally, the CLS model—which is comprised of independent hippocampal and neocortical networks—is a dual process model, with a separation of concerns that aligns closely with the dual process signal detection (DPSD) cognitive model. That is, memory retrieval in the hippocampus is similar in quantitative characteristics to threshold-based recollection, and activity in the neocortex is comparable to strength-based familiarity. This equivalency, while not
without limitations, was highly instructive. In particular, in the first paper, the CLS model was first shown to produce hippocampal- and neocortical-based recognition memory ROCs that closely matched human experimental ROCs when biased to favor recollection or familiarity, respectively. Findings such as these support the argument that important aspects of the neural underpinnings of recognition memory can be understood in terms of the principles instantiated by the CLS model; namely, the neural mechanisms of pattern separation and pattern completion, described in the introductory chapter.

The computational approach also helped to move beyond the purely quantitative approach of matching data with archetypal distributions—such as the Gaussian curve—towards a more nuanced perspective in which the functions ascribed to cognitive processes are themselves malleable. Specifically, in the first two papers on stimulus overlap and stimulus complexity, respectively, experimental evidence showing atypical ROCs (i.e., curved source ROCs) presented a challenge—in one respect or another—for all existing cognitive models (as discussed in the previous section). Using the CLS model, a qualitative explanation was put forward, in which the ROC effects were attributed to changes in the degree of pattern separation that occurs within the hippocampus. That is, it is hypothesized that the hippocampus is normally able to maintain distinct representations of events and things. However, when the level of feature overlap is very high, pattern separation becomes severely diminished, resulting in highly overlapping neural representations. This high level of overlap is similar to the typical performance of the neocortex, which utilizes overlapping representations to generalize from new experiences to an existing knowledge structure. Thus, we can begin to see the distinction between the two systems as arising due to mechanistic biases rather than stringent rules.

A limitation of the CLS model—as it was applied in this thesis—is that the hippocampus and neocortex are instantiated as entirely independent networks. However, the perirhinal and parahippocampal cortices of the medial temporal lobe are the primary afferents to the hippocampus. Conversely, some computational models postulate that cortex-based memory formation is highly dependent on the hippocampus (Gluck & Myers, 1993; Rolls, 1989; Schmajuk & DiCarlo, 1992). In
addition, both the subiculum and deep layers of the entorhinal cortex receive direct projections from hippocampal subfield CA1 (Suzuki & Amaral, 2004). Thus, a more realistic and complete model must take into account interactions between the hippocampus and other closely connected networks. Several such undertakings have been reported in the literature, including a rudimentary combining of hippocampal and cortical networks by Norman and O’Reilly (2003), and a more sophisticated approach by Kumaran and McClelland (2012). Along similar lines, though not reported in the paper, when modeling the similarity gradients of recollection and familiarity, a variation of the model was implemented in which the cortical network was set up to project to the hippocampus in parallel with the stimulus input layer, which was simultaneously fed to both networks. This improved overall retrieval performance, as the hippocampus was found to pattern complete at lower levels of cue-target similarity than it had previously. Thus, there is good reason to suggest that the processes of familiarity and recollection interact with one another to mutual benefit. Detecting such interactions is an interesting and challenging problem for future experimental research.

In the fourth paper, it was shown that the CLS model can in principle account for some of the recent findings linking the hippocampus with the performance of global, perceptual discriminations (Aly & Yonelinas, 2012). This was an impressive achievement for the CLS model given that it was originally intended to account for the neural underpinnings of recognition memory, and perception was presumably not a consideration. The finding was all the more striking given that the kind of perception being modeled was a continuous, global match/mismatch signal, in stark contrast to the threshold signal associated with memory retrieval. However, the findings did not preclude the possibility that other areas of the cortex can also assist with global perception. In fact, we found that the cortical network was also effective at producing continuous, strength-based signals. Thus, how the hippocampus makes a unique contribution to perception is an open question. In addition, another important question that has yet to be addressed is how to account for the separate threshold perception signal also identified by Aly and Yonelinas (2012). A more complete computational model of perception will likely
require additional networks, outside the medial temporal lobes, in order to account for both types of signal.

6.3 Concluding Commentary

Thematically, the current thesis was in large part an exploration of boundary conditions in recognition memory and perception. For example, in the first paper we explored the limits of typical recognition memory performance by manipulating stimuli to become very similar, showing that when feature overlap is sufficiently high, the threshold retrieval process ascribed to recollection will break down to reflect a more continuous process. In the second paper, we explored boundaries associated with stimulus complexity, showing that when stimuli are sufficiently rich in information, recollection again undergoes a transformation and becomes continuous. And in the third paper, we investigated another type of boundary condition by effectively asking the question, “How far can we stretch the perceptual difference between a retrieval cue and a learned stimulus before retrieval will fail, and how can this failure be contrasted across measures of recollection and familiarity?” The boundaries associated with recollection and familiarity were in fact shown to have quite different characteristics—whereas familiarity changed gradually with incremental changes to similarity, recollection displayed a more abrupt, nonlinear transition between states of non-retrieval and retrieval, which consistently occurred over a narrow similarity range. The fourth paper was a departure from the previous three in that it presented only simulated explorations, investigating some of the more nuanced computational properties of the hippocampus in the context of recent behavioral and neuroimaging findings regarding perception. In this case, we explored the boundary between memory and perception, and how the hippocampus might be involved in contributing to either process, depending on task demands.

In exploring the above boundary conditions, we often paired techniques of computational modeling and behavioral experimentation. This proved a powerful investigative approach and, in particular, the CLS computational model illuminated unexpected connections between seemingly unrelated effects. For example, using the hippocampal component of the CLS model we can now hypothesize how the
effects of high stimulus similarity and high stimulus complexity are mechanistically related. The CLS model posits that the hippocampus is uniquely suited to storing and retrieving specific, episodic memories because it is able to encode sparse, highly pattern separated representations. This can be contrasted with the medial temporal lobe cortex, and more generally the posterior neocortex, which forms highly overlapping representations that reflect semantic relatedness, and a bias towards pattern completion. Our simulations of item and source memory paradigms showed that when pattern separation breaks down in the hippocampus, representations will tend to overlap, and consequently, retrieval will be less diagnostic of the correct source (i.e., old versus new, or list A versus list B). This can be related to the effect of increasing stimulus complexity—based on the assumption that information must be broken into more manageable, semantic chunks that are common to many stimuli—as the simulations indicated that the hippocampus is similarly overwhelmed by the level of feature overlap between stimuli in this situation and, consequently, pattern separation again collapses. Although more research is needed to validate this hypothesis, it is at least evident from the current research that utilizing behavioral and computational modeling approaches in concert can lead to compelling new theories that solve multiple disparate problems.

This combination of approaches was similarly effective when investigating how recognition memory performs under conditions in which retrieval cues do not match up perfectly with their learned targets. While much is known about the characteristics of recollection and familiarity, one could argue—in the context of the current research—that recognition memory has typically been explored in a relatively static way. That is, past research has demonstrated how individuals perform under ideal retrieval conditions—i.e., when cue and target do match perfectly—and we understand to some extent how recognition is affected when there is a partial mismatch between cue and target—i.e., false memory paradigms. However, a deeper question presents itself when these paradigms are viewed as cross-sections of a similarity continuum, or gradient. This is arguably of great importance when one considers that, in everyday life, familiar objects and contexts are constantly moving in and out of focus, or varying in subtle, or even dramatic, ways from one episode to another. The CLS model was used to predict how the
similarity gradients of recognition memory might differ when separated into the contributions of hippocampus—from which we inferred recollection-based effects—and the cortex—which we used as the basis for predicting familiarity. The dual process signal detection model was then used as the framework for our behavioral experiments. We found that the familiarity and recollection gradients extrapolated from the two behavioral experiments aligned very closely with the simulations, and while we must be careful not to conflate psychological constructs with the computations of isolated neural networks, the data is a compelling cross-validation of the two investigative approaches.

In light of the similarity gradient memory research, the succeeding computational investigation of perception in the hippocampus could well be viewed as a continuation of the same line of inquiry. That is, both papers were investigations of similarity gradients, but in response to different task demands; in the recognition memory study we were interested in the ability to recognize studied items as the objective distance between cue and target is varied, and in the perception simulations we were interested in how the hippocampus might discern pairs of items as the number of mismatching features is varied. Common to both paradigms was a segment in the gradient of rapid transition in the hippocampal model simulations, where pattern completion either began to occur or to fail—depending on which direction one is coming from—as the number of feature differences varied beyond a critical level.

So, given their similarities, in what ways are memory and perception different? This may be more a question of task demands than of fundamental, mechanistic differences. Specifically, in our recognition paradigm, we started at a very low level of cue-target similarity, such that pattern completion always failed. Once similarity was raised to a critical point, pattern completion occurred with regularity and was almost always accurate (i.e., the retrieved pattern most closely matched the target and not any of the other learned items). Because each target was itself quite distinct, the occurrence of pattern completion was highly diagnostic. In the perception paradigm, the focus was on the opposite, high end of the similarity scale, where similarity refers to the difference between two mismatching items, as opposed to a target and a lure in recognition memory. Thus, pattern completion—or
failure—was not sufficient to make an accurate determination, since all items—
matching or matching—were pattern completed. Instead, performance was based on
graded differences in activation strength above the pattern completion threshold.
Finally, there is the obvious difference that recollection-based memory implies the
persistence of episodic information over time (i.e., via offline storage, such as
Hebbian weight changes), whereas perception does not. However, it seems that
memory does not require a marked temporal delay before it will be effective, and
conversely, it might also be true that perception does not need immediacy to
function. For example, the perceptual comparison of a current episode to a retrieved
episode could conceivable be just as pertinent—if not more so—to everyday human
functioning, as the comparison between two contemporaneous scenes. This
possibility is an important topic for future research.

In conclusion, the current thesis set out to advance our understanding of the
computational mechanisms underlying recognition memory and related cognitive
processes. Several novel predictions were generated through computational
modeling techniques that were subsequently tested with a range of behavioral
experiments. Other predictions have yet to be tested experimentally, but nonetheless
provided valuable theoretical insights that can be drawn upon for future research.
Going forward, additional computational and behavioral research will no doubt yield
more interesting and unexpected results, helping in the drive towards the ultimate
goal of gaining a complete understanding of the computational underpinnings of
healthy brain functioning, and the complex relationship between brain and behavior.
References


Appendix A: Authorship Indication Forms
NOTE
This Authorship Indication form is a statement detailing the percentage of the contribution of each author in each associated ‘paper’. This form must be signed by each co-author and the Principal Coordinating Supervisor. This form must be added to the publication of your final thesis as an appendix. Please fill out a separate form for each associated paper to be included in your thesis.

DECLARATION
We hereby declare our contribution to the publication of the ‘paper’ entitled:

Testing a neurocomputational model of recollection, familiarity, and source recognition

First Author
Name: Kane W Elfman
Signature: Kane Elfman
Percentage of contribution: 80 %
Date: 27/9/2014
Brief description of contribution to the ‘paper’ and your central responsibilities/role on project:
Experiment design, data collection, simulations, writing paper.

Second Author
Name: Dr Colleen M Parks
Signature: Colleen Parks
Percentage of contribution: ___ %
Date: 10/1/2014
Brief description of your contribution to the ‘paper’:
I contributed to the design of the behavioral experiments and to writing the paper.

Third Author
Name: Dr Andrew P Yonelinas
Signature: Andrew Yonelinas
Percentage of contribution: ___ %
Date: 09/29/2014
Brief description of your contribution to the ‘paper’:
assisted with study design and writing

Fourth Author
Name: ____________________________
Signature: ____________________________
Percentage of contribution: ___ %
Date: ___ / ___ / _____
Brief description of your contribution to the ‘paper’:

Principal Coordinating Supervisor: Name: Dr Andrew P Yonelinas
Signature: Andrew Yonelinas
Date: 09/29/2014

In the case of more than four authors please attach another sheet with the names, signatures and contribution of the authors.
DEVELOPMENT

Variations in recollection: The effects of complexity on source recognition

First Author
*Dr. Colleen M Parks

Second Author
*Dr. Linda J Murray

Third Author
*Kane W Elfman

Fourth Author
*Dr. Andrew P Yonelinas

I contributed to the design of the research, analysis of the data, and writing the paper.

*contributed with study design and writing

*computational modeling writing of paper.
Swinburne Research

Authorship Indication Form
For PhD (including associated papers) candidates

NOTE

Recollection and familiarity exhibit dissociable similarity gradients: A test of the complementary learning systems model

DECLARATION

Kane W Elfman

Dr Andrew P Yonelinas

Andrew Yonelinas

Designed and ran behavioral experiments, simulations, data analysis, writing.

Assisted with study design and writing

Kane W Elfman

Dr Andrew P Yonelinas

Andrew Yonelinas

Designed and ran behavioral experiments, simulations, data analysis, writing.
NOTE

- A computational account of memory and perception: Thresholded and graded signals in the human hippocampus

DECLARATION

A neurocomputational account of memory and perception: Thresholded and graded signals in the human hippocampus

First Author

Kane W Elfman

Helped design the experiments, come up with predictions, propose ways of analyzing the data, and write the paper.

Second Author

Dr Mariam Aly

Assisted with study design and writing.

Third Author

Dr Andrew P Yonelinas

Assisted with study design and writing.

Fourth Author

Andrew Yonelinas

Helped design the experiments, come up with predictions, propose ways of analyzing the data, and write the paper.
Appendix B: Journal Reprint Licenses
This is a License Agreement between Kane W Elfman ("You") and American Psychological Association ("American Psychological Association") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by American Psychological Association, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number 3487340680044
License date Oct 13, 2014
Licensed content publisher American Psychological Association
Licensed content publication Journal of Experimental Psychology: Learning, Memory, and Cognition
Licensed content title Testing a neurocomputational model of recollection, familiarity, and source recognition.
Licensed content author Elfman, Kane W.; Parks, Colleen M.; Yonelinas, Andrew P.
Licensed content date Jul 1, 2008
Volume number 34
Issue number 4
Pages 752 - [empty string]
Type of Use Doctoral Thesis
Portion Large text excerpt / Full article
Number of pages requested 17
Page numbers to reuse 17
Format Print
Reference number None
Institution name Swinburne University
Billing Type Invoice
Company Kane W Elfman
Billing address 474 Waller Street
San Francisco, California 94117
Australia
Total 0.00 USD

TERMS AND CONDITIONS FOR PERMISSIONS

APA hereby grants you a non-exclusive license to reproduce this content for this purpose, and for no other use, subject to the conditions below:
1. APA warrants that it has, to the best of its knowledge, the rights to license reuse of this content. However, you must ensure that the material you are requesting is original and does not carry the copyright of another entity (as credited in the published version). If any part of the material you have requested indicates that it was reprinted or adapted with permission from another source, then you should also seek permission from that original source to reuse the material.

2. The specified fee shall be paid on or before your use of APA-licensed material. "$0.00" fee means that APA waived the fee.

3. The reproduced material must include the following credit line:

   **For English language content:** Copyright © 2008 by the American Psychological Association. Reproduced [or Adapted] with permission. The official citation that should be used in referencing this material is [list the original bibliographic citation]. The use of this information does not imply endorsement by the publisher.

   **For translated language content:** This material originally appeared in English as [list the full bibliographic citation]. Copyright © 2008 by the American Psychological Association. Translated and reproduced [or Adapted] with permission. The American Psychological Association is not responsible for the accuracy of this translation. This translation cannot be reproduced or distributed further without prior written permission from the APA. The use of this information does not imply endorsement by the publisher.

4. **Language:** This license is for one language only.

5. **Author permission:** You must obtain the author(s) (or, in case of multiple authorship, one author's) permission. A good faith effort must be made to locate the author and obtain the author's permission. Author permission is not required for Course Packs or Electronic Reserve use.

6. **Re-use in a book:** (a) Permission is limited to the life of the current edition; (b) Permission covers components that are published as an auxiliary to the current edition for the following formats: CDs/DVDs and student or instructor manuals; (c) Permission is granted for WorldRights; (d) A complimentary copy of the book shall be sent to the APAs Permissions Office.

7. **All online use:** (a) The following notice must be added to the credit line: No further reproduction or distribution is permitted without written permission from the American Psychological Association; (b) The credit line must appear on the first screen on which the content appears; (c) Full-text articles or book chapters may only be posted on a secure and restricted web site.

8. The foregoing License shall not take effect unless and until APA or CCC receives the payment in accordance with the CCC Billing and Payment Terms and Conditions (including all payment provisions), which are incorporated herein by reference. Failure to fully comply with the terms of this License or the CCC Billing and Payment Terms and Conditions shall automatically render this License null and void.

9. **License Revocation:** APA or CCC may, within 60 days from the date of license, deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will APA or CCC be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund
of the amount(s) paid by you to CCC for denied permissions.

10. APA assumes that any publisher using APA content in both print and electronic will grant APA a license to use their material under the same conditions, subject to STM Permission Guidelines.

11. Other conditions: None

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.
This is a License Agreement between Kane W Elfman ("You") and American Psychological Association ("American Psychological Association") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by American Psychological Association, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

<table>
<thead>
<tr>
<th>License Number</th>
<th>3487351222080</th>
</tr>
</thead>
<tbody>
<tr>
<td>License date</td>
<td>Oct 13, 2014</td>
</tr>
<tr>
<td>Licensed content publisher</td>
<td>American Psychological Association</td>
</tr>
<tr>
<td>Licensed content publication</td>
<td>Journal of Experimental Psychology: Learning, Memory, and Cognition</td>
</tr>
<tr>
<td>Licensed content title</td>
<td>Variations in recollection: The effects of complexity on source recognition.</td>
</tr>
<tr>
<td>Licensed content author</td>
<td>Parks, Colleen M.; Murray, Linda J.; Elfman, Kane; Yonelinas, Andrew P.</td>
</tr>
<tr>
<td>Licensed content date</td>
<td>Jul 1, 2011</td>
</tr>
<tr>
<td>Volume number</td>
<td>37</td>
</tr>
<tr>
<td>Issue number</td>
<td>4</td>
</tr>
<tr>
<td>Pages</td>
<td>861 - [empty string]</td>
</tr>
<tr>
<td>Type of Use</td>
<td>Doctoral Thesis</td>
</tr>
<tr>
<td>Portion</td>
<td>Large text excerpt / Full article</td>
</tr>
<tr>
<td>Number of pages requested</td>
<td>13</td>
</tr>
<tr>
<td>Page numbers to reuse</td>
<td>13</td>
</tr>
<tr>
<td>Format</td>
<td>Print</td>
</tr>
<tr>
<td>Reference number</td>
<td>None</td>
</tr>
<tr>
<td>Institution name</td>
<td>Swinburne University</td>
</tr>
<tr>
<td>Billing Type</td>
<td>Invoice</td>
</tr>
<tr>
<td>Company</td>
<td>Kane W Elfman</td>
</tr>
<tr>
<td>Billing address</td>
<td>474 Waller Street</td>
</tr>
<tr>
<td></td>
<td>San Francisco, California 94117</td>
</tr>
<tr>
<td></td>
<td>Australia</td>
</tr>
<tr>
<td>Total</td>
<td>0.00 USD</td>
</tr>
</tbody>
</table>

TERMS AND CONDITIONS FOR PERMISSIONS

APA hereby grants you a non-exclusive license to reproduce this content for this purpose, and for no other use, subject to the conditions below:
1. APA warrants that it has, to the best of its knowledge, the rights to license reuse of this content. However, you must ensure that the material you are requesting is original and does not carry the copyright of another entity (as credited in the published version). If any part of the material you have requested indicates that it was reprinted or adapted with permission from another source, then you should also seek permission from that original source to reuse the material.

2. The specified fee shall be paid on or before your use of APA-licensed material. "$0.00" fee means that APA waived the fee.

3. The reproduced material must include the following credit line:

For English language content: Copyright © 2011 by the American Psychological Association. Reproduced [or Adapted] with permission. The official citation that should be used in referencing this material is [list the original bibliographic citation]. The use of this information does not imply endorsement by the publisher.

For translated language content: This material originally appeared in English as [list the full bibliographic citation]. Copyright © 2011 by the American Psychological Association. Translated and reproduced [or Adapted] with permission. The American Psychological Association is not responsible for the accuracy of this translation. This translation cannot be reproduced or distributed further without prior written permission from the APA. The use of this information does not imply endorsement by the publisher.

4. Language: This license is for one language only.

5. Author permission: You must obtain the author's(s) or, in case of multiple authorship, one author's permission. A good faith effort must be made to locate the author and obtain the author's permission. Author permission is not required for Course Packs or Electronic Reserve use.

6. Re-use in a book: (a) Permission is limited to the life of the current edition; (b) Permission covers components that are published as an auxiliary to the current edition for the following formats: CDs/DVDs and student or instructor manuals; (c) Permission is granted for WorldRights; (d) A complimentary copy of the book shall be sent to the APA Permissions Office.

7. All online use: (a) The following notice must be added to the credit line: No further reproduction or distribution is permitted without written permission from the American Psychological Association; (b) The credit line must appear on the first screen on which the content appears; (c) Full-text articles or book chapters may only be posted on a secure and restricted web site.

8. The foregoing License shall not take effect unless and until APA or CCC receives the payment in accordance with the CCC Billing and Payment Terms and Conditions (including all payment provisions), which are incorporated herein by reference. Failure to fully comply with the terms of this License or the CCC Billing and Payment Terms and Conditions shall automatically render this License null and void.

9. License Revocation: APA or CCC may, within 60 days from the date of license, deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will APA or CCC be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund.
of the amount(s) paid by you to CCC for denied permissions.

10. APA assumes that any publisher using APA content in both print and electronic will grant APA a license to use their material under the same conditions, subject to STM Permission Guidelines.

11. Other conditions: None

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.
This is a License Agreement between Kane W Elfman ("You") and MIT Press - Journals ("MIT Press - Journals") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by MIT Press - Journals, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

<table>
<thead>
<tr>
<th>License Number</th>
<th>3607741334099</th>
</tr>
</thead>
<tbody>
<tr>
<td>License date</td>
<td>Apr 14, 2015</td>
</tr>
<tr>
<td>Licensed content publisher</td>
<td>MIT Press - Journals</td>
</tr>
<tr>
<td>Licensed content title</td>
<td>Journal of cognitive neuroscience</td>
</tr>
<tr>
<td>Licensed content date</td>
<td>Jan 1, 1989</td>
</tr>
<tr>
<td>Type of Use</td>
<td>Thesis/Dissertation</td>
</tr>
<tr>
<td>Requestor type</td>
<td>Author of requested content</td>
</tr>
<tr>
<td>Format</td>
<td>Print</td>
</tr>
<tr>
<td>Portion</td>
<td>chapter/article</td>
</tr>
<tr>
<td>Number of pages in chapter/article</td>
<td>17</td>
</tr>
<tr>
<td>Title or numeric reference of the portion(s)</td>
<td>Entire paper</td>
</tr>
<tr>
<td>Title of the article or chapter the portion is from</td>
<td>Recollection and Familiarity Exhibit Dissociable Similarity Gradients: A Test of the Complementary Learning Systems Model</td>
</tr>
<tr>
<td>Editor of portion(s)</td>
<td>N/A</td>
</tr>
<tr>
<td>Author of portion(s)</td>
<td>Kane Elfman</td>
</tr>
<tr>
<td>Volume of serial or monograph.</td>
<td>Vol 27</td>
</tr>
<tr>
<td>Page range of the portion</td>
<td>1-17</td>
</tr>
<tr>
<td>Publication date of portion</td>
<td>May 2015</td>
</tr>
<tr>
<td>Rights for</td>
<td>Main product</td>
</tr>
<tr>
<td>Duration of use</td>
<td>Current edition and up to 5 years</td>
</tr>
<tr>
<td>Creation of copies for the disabled</td>
<td>no</td>
</tr>
<tr>
<td>With minor editing privileges</td>
<td>no</td>
</tr>
<tr>
<td>For distribution to</td>
<td>Other territories and/or countries</td>
</tr>
<tr>
<td>Territory/Countries where you intend to distribute new product</td>
<td>United States, Australia</td>
</tr>
</tbody>
</table>
In the following language(s) Original language of publication
With incidental promotional use no
The lifetime unit quantity of new product Up to 499
Made available in the following markets education
The requesting person/organization is: Kane Elfman
Order reference number None
Author/Editor Kane Elfman
The standard identifier Thesis
Title Behavioral and Neurocomputational Investigations of Human Recognition Memory and Cognition
Publisher Swinburne University
Expected publication date Jun 2015
Estimated size (pages) 150
Total (may include CCC user fee) 0.00 USD

TERMS AND CONDITIONS

The following terms are individual to this publisher:

None

Other Terms and Conditions:

None

STANDARD TERMS AND CONDITIONS

1. Description of Service; Defined Terms. This Republication License enables the User to obtain licenses for republication of one or more copyrighted works as described in detail on the relevant Order Confirmation (the “Work(s)”). Copyright Clearance Center, Inc. (“CCC”) grants licenses through the Service on behalf of the rightsholder identified on the Order Confirmation (the “Rightsholder”). “Republication”, as used herein, generally means the inclusion of a Work, in whole or in part, in a new work or works, also as described on the Order Confirmation. “User”, as used herein, means the person or entity making such republication.

2. The terms set forth in the relevant Order Confirmation, and any terms set by the Rightsholder with respect to a particular Work, govern the terms of use of Works in connection with the Service. By using the Service, the person transacting for a republication license on behalf of the User represents and warrants that he/she/it (a) has been duly authorized by the User to accept, and hereby does accept, all such terms and conditions on behalf of User, and (b) shall inform User of all such terms and conditions. In the event such person is a “freelancer” or other third party independent of User and CCC, such party shall be deemed jointly a “User” for purposes of these terms and conditions. In any event, User shall be deemed to have accepted and agreed to all such terms and conditions if User
3. Scope of License; Limitations and Obligations.

3.1 All Works and all rights therein, including copyright rights, remain the sole and exclusive property of the Rightsholder. The license created by the exchange of an Order Confirmation (and/or any invoice) and payment by User of the full amount set forth on that document includes only those rights expressly set forth in the Order Confirmation and in these terms and conditions, and conveys no other rights in the Work(s) to User. All rights not expressly granted are hereby reserved.

3.2 General Payment Terms: You may pay by credit card or through an account with us payable at the end of the month. If you and we agree that you may establish a standing account with CCC, then the following terms apply: Remit Payment to: Copyright Clearance Center, Dept 001, P.O. Box 843006, Boston, MA 02284-3006. Payments Due: Invoices are payable upon their delivery to you (or upon our notice to you that they are available to you for downloading). After 30 days, outstanding amounts will be subject to a service charge of 1-1/2% per month or, if less, the maximum rate allowed by applicable law. Unless otherwise specifically set forth in the Order Confirmation or in a separate written agreement signed by CCC, invoices are due and payable on “net 30” terms. While User may exercise the rights licensed immediately upon issuance of the Order Confirmation, the license is automatically revoked and is null and void, as if it had never been issued, if complete payment for the license is not received on a timely basis either from User directly or through a payment agent, such as a credit card company.

3.3 Unless otherwise provided in the Order Confirmation, any grant of rights to User (i) is “one-time” (including the editions and product family specified in the license), (ii) is non-exclusive and non-transferable and (iii) is subject to any and all limitations and restrictions (such as, but not limited to, limitations on duration of use or circulation) included in the Order Confirmation or invoice and/or in these terms and conditions. Upon completion of the licensed use, User shall either secure a new permission for further use of the Work(s) or immediately cease any new use of the Work(s) and shall render inaccessible (such as by deleting or by removing or severing links or other locators) any further copies of the Work (except for copies printed on paper in accordance with this license and still in User's stock at the end of such period).

3.4 In the event that the material for which a republication license is sought includes third party materials (such as photographs, illustrations, graphs, inserts and similar materials) which are identified in such material as having been used by permission, User is responsible for identifying, and seeking separate licenses (under this Service or otherwise) for, any of such third party materials; without a separate license, such third party materials may not be used.

3.5 Use of proper copyright notice for a Work is required as a condition of any license granted under the Service. Unless otherwise provided in the Order Confirmation, a proper copyright notice will read substantially as follows: “Republiced with permission of [Rightsholder’s name], from [Work's title, author, volume, edition number and year of copyright]; permission conveyed through Copyright Clearance Center, Inc. ” Such notice must be provided in a reasonably legible font size and must be placed either immediately adjacent to the Work as used (for example, as part of a by-line or footnote but not as a separate electronic link) or in the place where substantially all other credits or notices for the new work containing the republished Work are located. Failure to include the required notice results in loss to the Rightsholder and CCC, and the User shall be liable to pay liquidated damages for each such failure equal to twice the use fee specified in the Order Confirmation, in addition to the use fee itself and any other fees and charges specified.

3.6 User may only make alterations to the Work if and as expressly set forth in the Order
Confirmation. No Work may be used in any way that is defamatory, violates the rights of third parties (including such third parties' rights of copyright, privacy, publicity, or other tangible or intangible property), or is otherwise illegal, sexually explicit or obscene. In addition, User may not conjoin a Work with any other material that may result in damage to the reputation of the Rightsholder. User agrees to inform CCC if it becomes aware of any infringement of any rights in a Work and to cooperate with any reasonable request of CCC or the Rightsholder in connection therewith.

4. Indemnity. User hereby indemnifies and agrees to defend the Rightsholder and CCC, and their respective employees and directors, against all claims, liability, damages, costs and expenses, including legal fees and expenses, arising out of any use of a Work beyond the scope of the rights granted herein, or any use of a Work which has been altered in any unauthorized way by User, including claims of defamation or infringement of rights of copyright, publicity, privacy or other tangible or intangible property.

5. Limitation of Liability. UNDER NO CIRCUMSTANCES WILL CCC OR THE RIGHTSHOLDER BE LIABLE FOR ANY DIRECT, INDIRECT, CONSEQUENTIAL OR INCIDENTAL DAMAGES (INCLUDING WITHOUT LIMITATION DAMAGES FOR LOSS OF BUSINESS PROFITS OR INFORMATION, OR FOR BUSINESS INTERRUPTION) ARISING OUT OF THE USE OR INABILITY TO USE A WORK, EVEN IF ONE OF THEM HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. In any event, the total liability of the Rightsholder and CCC (including their respective employees and directors) shall not exceed the total amount actually paid by User for this license. User assumes full liability for the actions and omissions of its principals, employees, agents, affiliates, successors and assigns.

6. Limited Warranties. THE WORK(S) AND RIGHT(S) ARE PROVIDED “AS IS”. CCC HAS THE RIGHT TO GRANT TO USER THE RIGHTS GRANTED IN THE ORDER CONFIRMATION DOCUMENT. CCC AND THE RIGHTSHOLDER DISCLAIM ALL OTHER WARRANTIES RELATING TO THE WORK(S) AND RIGHT(S), EITHER EXPRESS OR IMPLIED, INCLUDING WITHOUT LIMITATION IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. ADDITIONAL RIGHTS MAY BE REQUIRED TO USE ILLUSTRATIONS, GRAPHS, PHOTOGRAPHS, ABSTRACTS, INSERTS OR OTHER PORTIONS OF THE WORK (AS OPPOSED TO THE ENTIRE WORK) IN A MANNER CONTEMPLATED BY USER; USER UNDERSTANDS AND AGREES THAT NEITHER CCC NOR THE RIGHTSHOLDER MAY HAVE SUCH ADDITIONAL RIGHTS TO GRANT.

7. Effect of Breach. Any failure by User to pay any amount when due, or any use by User of a Work beyond the scope of the license set forth in the Order Confirmation and/or these terms and conditions, shall be a material breach of the license created by the Order Confirmation and these terms and conditions. Any breach not cured within 30 days of written notice thereof shall result in immediate termination of such license without further notice. Any unauthorized (but licensable) use of a Work that is terminated immediately upon notice thereof may be liquidated by payment of the Rightsholder's ordinary license price therefor; any unauthorized (and unlicensable) use that is not terminated immediately for any reason (including, for example, because materials containing the Work cannot reasonably be recalled) will be subject to all remedies available at law or in equity, but in no event to a payment of less than three times the Rightsholder's ordinary license price for the most closely analogous licensable use plus Rightsholder's and/or CCC's costs and expenses incurred in collecting such payment.

8. Miscellaneous.

8.1 User acknowledges that CCC may, from time to time, make changes or additions to the Service or to these terms and conditions, and CCC reserves the right to send notice to the User by electronic mail or otherwise for the purposes of notifying User of such changes or
additions; provided that any such changes or additions shall not apply to permissions already secured and paid for.

8.2 Use of User-related information collected through the Service is governed by CCC’s privacy policy, available online here: http://www.copyright.com/content/cc3/en/tools/footer/privacypolicy.html.

8.3 The licensing transaction described in the Order Confirmation is personal to User. Therefore, User may not assign or transfer to any other person (whether a natural person or an organization of any kind) the license created by the Order Confirmation and these terms and conditions or any rights granted hereunder; provided, however, that User may assign such license in its entirety on written notice to CCC in the event of a transfer of all or substantially all of User’s rights in the new material which includes the Work(s) licensed under this Service.

8.4 No amendment or waiver of any terms is binding unless set forth in writing and signed by the parties. The Rightsholder and CCC hereby object to any terms contained in any writing prepared by the User or its principals, employees, agents or affiliates and purporting to govern or otherwise relate to the licensing transaction described in the Order Confirmation, which terms are in any way inconsistent with any terms set forth in the Order Confirmation and/or in these terms and conditions or CCC’s standard operating procedures, whether such writing is prepared prior to, simultaneously with or subsequent to the Order Confirmation, and whether such writing appears on a copy of the Order Confirmation or in a separate instrument.

8.5 The licensing transaction described in the Order Confirmation document shall be governed by and construed under the law of the State of New York, USA, without regard to the principles thereof of conflicts of law. Any case, controversy, suit, action, or proceeding arising out of, in connection with, or related to such licensing transaction shall be brought, at CCC’s sole discretion, in any federal or state court located in the County of New York, State of New York, USA, or in any federal or state court whose geographical jurisdiction covers the location of the Rightsholder set forth in the Order Confirmation. The parties expressly submit to the personal jurisdiction and venue of each such federal or state court. If you have any comments or questions about the Service or Copyright Clearance Center, please contact us at 978-750-8400 or send an e-mail to info@copyright.com.

v 1.1

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.
This is a License Agreement between Kane W Elfman ("You") and John Wiley and Sons ("John Wiley and Sons") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by John Wiley and Sons, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

<table>
<thead>
<tr>
<th>License Number</th>
<th>3487350481094</th>
</tr>
</thead>
<tbody>
<tr>
<td>License date</td>
<td>Oct 13, 2014</td>
</tr>
<tr>
<td>Licensed content publisher</td>
<td>John Wiley and Sons</td>
</tr>
<tr>
<td>Licensed content publication</td>
<td>Hippocampus</td>
</tr>
<tr>
<td>Licensed content title</td>
<td>Neurocomputational account of memory and perception: Thresholded and graded signals in the hippocampus</td>
</tr>
<tr>
<td>Licensed copyright line</td>
<td>Copyright © 2014 Wiley Periodicals, Inc.</td>
</tr>
<tr>
<td>Licensed content author</td>
<td>Kane W. Elfman, Mariam Aly, Andrew P. Yonelinas</td>
</tr>
<tr>
<td>Licensed content date</td>
<td>Aug 27, 2014</td>
</tr>
<tr>
<td>Start page</td>
<td>n/a</td>
</tr>
<tr>
<td>End page</td>
<td>n/a</td>
</tr>
<tr>
<td>Type of use</td>
<td>Dissertation/Thesis</td>
</tr>
<tr>
<td>Requestor type</td>
<td>Author of this Wiley article</td>
</tr>
<tr>
<td>Format</td>
<td>Print</td>
</tr>
<tr>
<td>Portion</td>
<td>Full article</td>
</tr>
<tr>
<td>Will you be translating?</td>
<td>No</td>
</tr>
<tr>
<td>Title of your thesis / dissertation</td>
<td>Behavioral and Neurocomputational Investigations of Human Recognition Memory and Cognition</td>
</tr>
<tr>
<td>Expected completion date</td>
<td>Dec 2014</td>
</tr>
<tr>
<td>Expected size (number of pages)</td>
<td>150</td>
</tr>
<tr>
<td>Total</td>
<td>0.00 USD</td>
</tr>
<tr>
<td>Terms and Conditions</td>
<td></td>
</tr>
</tbody>
</table>

TERMS AND CONDITIONS

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a "Wiley Company") or handled on behalf of a society with which a Wiley Company has exclusive publishing rights in relation to a particular work (collectively "WILEY"). By clicking accept in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the billing and payment terms and conditions established by the Copyright Clearance Center Inc., ("CCC’s Billing and Payment terms and conditions"), at the time that you opened your Rightslink account (these are available at any time at
Terms and Conditions

- The materials you have requested permission to reproduce or reuse (the "Wiley Materials") are protected by copyright.

- You are hereby granted a personal, non-exclusive, non-sub licensable (on a stand-alone basis), non-transferable, worldwide, limited license to reproduce the Wiley Materials for the purpose specified in the licensing process. This license is for a one-time use only and limited to any maximum distribution number specified in the license. The first instance of republication or reuse granted by this licence must be completed within two years of the date of the grant of this licence (although copies prepared before the end date may be distributed thereafter). The Wiley Materials shall not be used in any other manner or for any other purpose, beyond what is granted in the license. Permission is granted subject to an appropriate acknowledgement given to the author, title of the material/book/journal and the publisher. You shall also duplicate the copyright notice that appears in the Wiley publication in your use of the Wiley Material. Permission is also granted on the understanding that nowhere in the text is a previously published source acknowledged for all or part of this Wiley Material. Any third party content is expressly excluded from this permission.

- With respect to the Wiley Materials, all rights are reserved. Except as expressly granted by the terms of the license, no part of the Wiley Materials may be copied, modified, adapted (except for minor reformatting required by the new Publication), translated, reproduced, transferred or distributed, in any form or by any means, and no derivative works may be made based on the Wiley Materials without the prior permission of the respective copyright owner. You may not alter, remove or suppress in any manner any copyright, trademark or other notices displayed by the Wiley Materials. You may not license, rent, sell, loan, lease, pledge, offer as security, transfer or assign the Wiley Materials on a stand-alone basis, or any of the rights granted to you hereunder to any other person.

- The Wiley Materials and all of the intellectual property rights therein shall at all times remain the exclusive property of John Wiley & Sons Inc, the Wiley Companies, or their respective licensors, and your interest therein is only that of having possession of and the right to reproduce the Wiley Materials pursuant to Section 2 herein during the continuance of this Agreement. You agree that you own no right, title or interest in or to the Wiley Materials or any of the intellectual property rights therein. You shall have no rights hereunder other than the license as provided for above in Section 2. No right, license or interest to any trademark, trade name, service mark or other branding ("Marks") of WILEY or its licensors is granted hereunder, and you agree that you shall not assert any such right, license or interest with respect thereto.

- NEITHER WILEY NOR ITS LICENSORS MAKES ANY WARRANTY OR REPRESENTATION OF ANY KIND TO YOU OR ANY THIRD PARTY, EXPRESS, IMPLIED OR STATUTORY, WITH RESPECT TO THE MATERIALS OR THE ACCURACY OF ANY INFORMATION CONTAINED IN THE MATERIALS, INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTY OF MERCHANTABILITY, ACCURACY, SATISFACTORY QUALITY, FITNESS FOR A PARTICULAR PURPOSE, USABILITY, INTEGRATION OR NON-INFRINGEMENT AND ALL SUCH WARRANTIES ARE HEREBY EXCLUDED BY WILEY AND ITS LICENSORS AND WAIVED.
BY YOU

- WILEY shall have the right to terminate this Agreement immediately upon breach of this Agreement by you.

- You shall indemnify, defend and hold harmless WILEY, its Licensees and their respective directors, officers, agents and employees, from and against any actual or threatened claims, demands, causes of action or proceedings arising from any breach of this Agreement by you.

- IN NO EVENT SHALL WILEY OR ITS LICENSORS BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR ENTITY FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, EXEMPLARY OR PUNITIVE DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, PROVISIONING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

- Should any provision of this Agreement be held by a court of competent jurisdiction to be illegal, invalid, or unenforceable, that provision shall be deemed amended to achieve as nearly as possible the same economic effect as the original provision, and the legality, validity and enforceability of the remaining provisions of this Agreement shall not be affected or impaired thereby.

- The failure of either party to enforce any term or condition of this Agreement shall not constitute a waiver of either party's right to enforce each and every term and condition of this Agreement. No breach under this agreement shall be deemed waived or excused by either party unless such waiver or consent is in writing signed by the party granting such waiver or consent. The waiver by or consent of a party to a breach of any provision of this Agreement shall not operate or be construed as a waiver of or consent to any other or subsequent breach by such other party.

- This Agreement may not be assigned (including by operation of law or otherwise) by you without WILEY’s prior written consent.

- Any fee required for this permission shall be non-refundable after thirty (30) days from receipt by the CCC.

- These terms and conditions together with CCC’s Billing and Payment terms and conditions (which are incorporated herein) form the entire agreement between you and WILEY concerning this licensing transaction and (in the absence of fraud) supersedes all prior agreements and representations of the parties, oral or written. This Agreement may not be amended except in writing signed by both parties. This Agreement shall be binding upon and inure to the benefit of the parties’ successors, legal representatives, and authorized assigns.

- In the event of any conflict between your obligations established by these terms and
conditions and those established by CCC’s Billing and Payment terms and conditions, these terms and conditions shall prevail.

- WILEY expressly reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC’s Billing and Payment terms and conditions.

- This Agreement will be void if the Type of Use, Format, Circulation, or Requestor Type was misrepresented during the licensing process.

- This Agreement shall be governed by and construed in accordance with the laws of the State of New York, USA, without regards to such state’s conflict of law rules. Any legal action, suit or proceeding arising out of or relating to these Terms and Conditions or the breach thereof shall be instituted in a court of competent jurisdiction in New York County in the State of New York in the United States of America and each party hereby consents and submits to the personal jurisdiction of such court, waives any objection to venue in such court and consents to service of process by registered or certified mail, return receipt requested, at the last known address of such party.

WILEY OPEN ACCESS TERMS AND CONDITIONS

Wiley Publishes Open Access Articles in fully Open Access Journals and in Subscription journals offering Online Open. Although most of the fully Open Access journals publish open access articles under the terms of the Creative Commons Attribution (CC BY) License only, the subscription journals and a few of the Open Access Journals offer a choice of Creative Commons Licenses: Creative Commons Attribution (CC-BY) license Creative Commons Attribution Non-Commercial (CC-BY-NC) license and Creative Commons Attribution Non-Commercial-NoDerivs (CC-BY-NC-ND) License. The license type is clearly identified on the article.

Copyright in any research article in a journal published as Open Access under a Creative Commons License is retained by the author(s). Authors grant Wiley a license to publish the article and identify itself as the original publisher. Authors also grant any third party the right to use the article freely as long as its integrity is maintained and its original authors, citation details and publisher are identified as follows: [Title of Article/Author/Journal Title and Volume/Issue. Copyright (c) [year] [copyright owner as specified in the Journal]. Links to the final article on Wiley’s website are encouraged where applicable.

The Creative Commons Attribution License

The Creative Commons Attribution License (CC-BY) allows users to copy, distribute and transmit an article, adapt the article and make commercial use of the article. The CC-BY license permits commercial and non-commercial re-use of an open access article, as long as the author is properly attributed.

The Creative Commons Attribution License does not affect the moral rights of authors, including without limitation the right not to have their work subjected to derogatory treatment. It also does not affect any other rights held by authors or third parties in the article, including without limitation the rights of privacy and publicity. Use of the article must not assert or imply, whether implicitly or explicitly, any connection with, endorsement or sponsorship of such use by the author, publisher or any other party associated with the article.
For any reuse or distribution, users must include the copyright notice and make clear to others that the article is made available under a Creative Commons Attribution license, linking to the relevant Creative Commons web page.

To the fullest extent permitted by applicable law, the article is made available as is and without representation or warranties of any kind whether express, implied, statutory or otherwise including, without limitation, warranties of title, merchantability, fitness for a particular purpose, non-infringement, absence of defects, accuracy, or the presence or absence of errors.

**Creative Commons Attribution Non-Commercial License**

The [Creative Commons Attribution Non-Commercial (CC-BY-NC) License](https://creativecommons.org/licenses/by-nc/4.0/) permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes. (see below)

**Creative Commons Attribution-Non-Commercial-NoDerivs License**

The [Creative Commons Attribution Non-Commercial-NoDerivs License](https://creativecommons.org/licenses/by-nc-nd/4.0/) (CC-BY-NC-ND) permits use, distribution and reproduction in any medium, provided the original work is properly cited, is not used for commercial purposes and no modifications or adaptations are made. (see below)

**Use by non-commercial users**

For non-commercial and non-promotional purposes, individual users may access, download, copy, display and redistribute to colleagues Wiley Open Access articles, as well as adapt, translate, text- and data-mine the content subject to the following conditions:

- The authors' moral rights are not compromised. These rights include the right of "paternity" (also known as "attribution" - the right for the author to be identified as such) and "integrity" (the right for the author not to have the work altered in such a way that the author's reputation or integrity may be impugned).

- Where content in the article is identified as belonging to a third party, it is the obligation of the user to ensure that any reuse complies with the copyright policies of the owner of that content.

- If article content is copied, downloaded or otherwise reused for non-commercial research and education purposes, a link to the appropriate bibliographic citation (authors, journal, article title, volume, issue, page numbers, DOI and the link to the definitive published version on Wiley Online Library) should be maintained. Copyright notices and disclaimers must not be deleted.

- Any translations, for which a prior translation agreement with Wiley has not been agreed, must prominently display the statement: "This is an unofficial translation of an article that appeared in a Wiley publication. The publisher has not endorsed this translation."

**Use by commercial "for-profit" organisations**

Use of Wiley Open Access articles for commercial, promotional, or marketing purposes requires further explicit permission from Wiley and will be subject to a fee. Commercial purposes include:
• Copying or downloading of articles, or linking to such articles for further redistribution, sale or licensing;

• Copying, downloading or posting by a site or service that incorporates advertising with such content;

• The inclusion or incorporation of article content in other works or services (other than normal quotations with an appropriate citation) that is then available for sale or licensing, for a fee (for example, a compilation produced for marketing purposes, inclusion in a sales pack)

• Use of article content (other than normal quotations with appropriate citation) by for-profit organisations for promotional purposes

• Linking to article content in e-mails redistributed for promotional, marketing or educational purposes;

• Use for the purposes of monetary reward by means of sale, resale, licence, loan, transfer or other form of commercial exploitation such as marketing products

• Print reprints of Wiley Open Access articles can be purchased from: corporatesales@wiley.com

Further details can be found on Wiley Online Library
http://olabout.wiley.com/WileyCDA/Section/id-410895.html

Other Terms and Conditions:

v1.9

Questions? custmercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.
Appendix C: Ethics Certification
February 12, 2013

Dear Dr. Yonelinas,

On 02/12/2013 the IRB reviewed the following protocol:

<table>
<thead>
<tr>
<th>Type of Review:</th>
<th>Continuing Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title:</td>
<td>Memory and the Brain</td>
</tr>
<tr>
<td>Investigator:</td>
<td>Yonelinas, Andrew, PhD</td>
</tr>
<tr>
<td>IRB ID#:</td>
<td>29272-4</td>
</tr>
<tr>
<td>Funding:</td>
<td>None/Department</td>
</tr>
<tr>
<td>Grant Title &amp; Grant ID:</td>
<td>NA</td>
</tr>
<tr>
<td>IND/IDE:</td>
<td>NA</td>
</tr>
<tr>
<td>Documents Reviewed:</td>
<td>Renewal, Summary of Changes, Modification Form, Description of Study, Consent Form, Previously-Approved Modification, Study Flyer</td>
</tr>
</tbody>
</table>

The IRB approved the protocol from 02/12/2013 to 02/11/2016 inclusive.

<table>
<thead>
<tr>
<th>Risk Determination:</th>
<th>No more than Minimal Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category:</td>
<td>Expedited (5,6,7a)</td>
</tr>
</tbody>
</table>

This Assurance, on file with the Department of Health and Human Services, covers this activity:

FWA No: 00004557
Expiration Date: June 28, 2016
IORG: 000251

Before 12/13/2015 or within 25 business days of study closure, whichever is earlier, you are to submit a completed “FORM: Continuing Review (HRP-212)” and required attachments to request continuing approval or closure. If continuing review approval is not granted before the expiration date of 02/11/2016, this protocol expires on that date.

I Attached are stamped approved consent documents. Use copies of these documents to document consent. In conducting this protocol you are required to follow the requirements listed in the INVESTIGATOR MANUAL (HRP-103).

Sincerely,

[Signature]

IRB Analyst
IRB Administration
UC Davis
916.703.9151