

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 test 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 similarity.

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 *k*WTA 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 coactive 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 *k*WTA 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

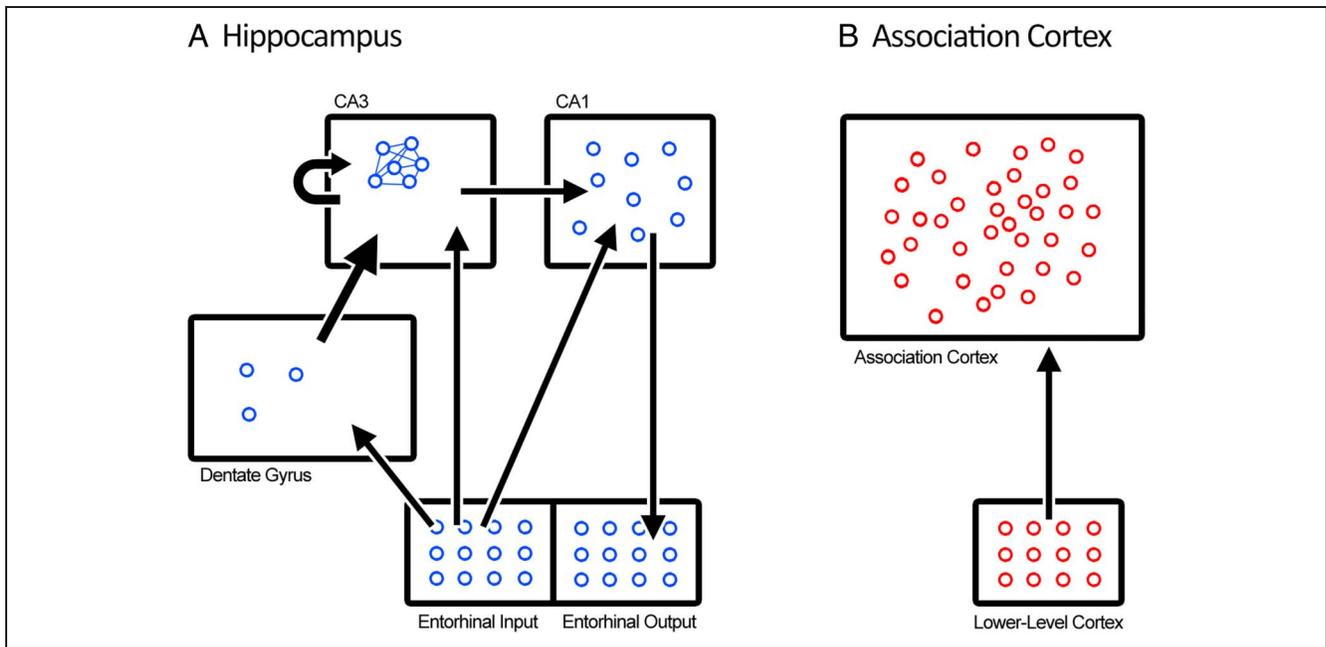


Figure 1. Schematic illustrations of the hippocampus (A) and association cortex (B) of the CLS model. (A) The stimulus item is initially presented to the entorhinal cortex (EC) input layer as a matrix of features (e.g., colors, shapes, configuration, etc.). Activity spreads to the CA subfields, directly and indirectly via the DG. Patterns are learned via Hebbian weight changes between coactive units, including recurrent collaterals within CA3 that help bind the features together. Activity in the DG is relatively sparse because of highly inhibitory interneurons, and this is expected to facilitate the formation of distinct or pattern-separated memory representations. During retrieval, a test cue is presented to the input layer of the EC, which activates CA3 if it matches a stored memory. This in turn activates CA1, which converts the retrieved representation back to its original form at the EC output layer. (B) The association cortex network comprises two layers: the lower-level cortex (which for simplicity is instantiated in an identical manner to the hippocampal EC layer) and the association cortex. The lower-level cortex projects diffusely throughout the association cortex. When an item is encoded, connections between coactive units are strengthened, and the other connections are weakened. In this way, at time of test when a studied item is presented, it will be associated with a sharpened representation (i.e., a small number of highly active cortical units and a large number of inactive units), compared with nonstudied items.

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).

Figure 2. A simplified representation of a study pattern and three corresponding test patterns (cues) that vary in similarity to the study item (target), from chance to an exact match (0 to 1 similarity scores, respectively). Each row of squares represents an item feature with the black square representing a fully active unit and empty squares indicating inactivity.

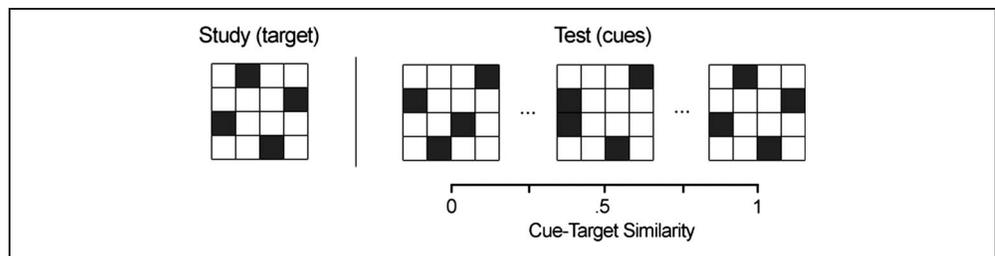
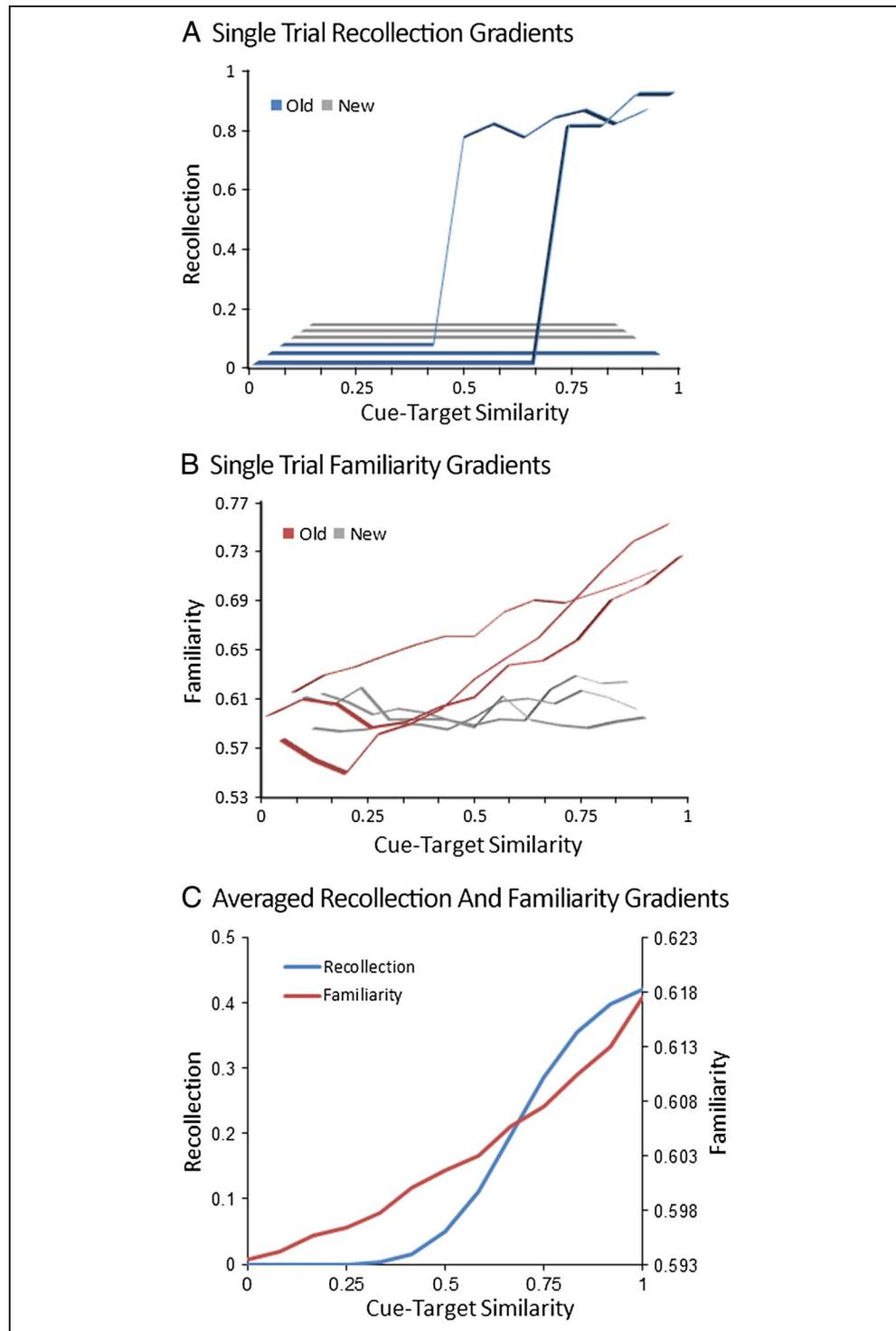


Figure 3. Simulation results showing memory strength for the hippocampus (i.e., recollection) and the association cortex (i.e., familiarity) as cue–target similarity is varied from 0 (chance similarity) to 1 (exact match). Representative similarity gradients of three old and three new items from the hippocampal network (A) and from the association cortex (B), as well as average gradients across trials (C). Increasing cue–target similarity leads to steeply increasing recollection gradients and shallower more gradually increasing familiarity gradients.



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 neuro-anatomical 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

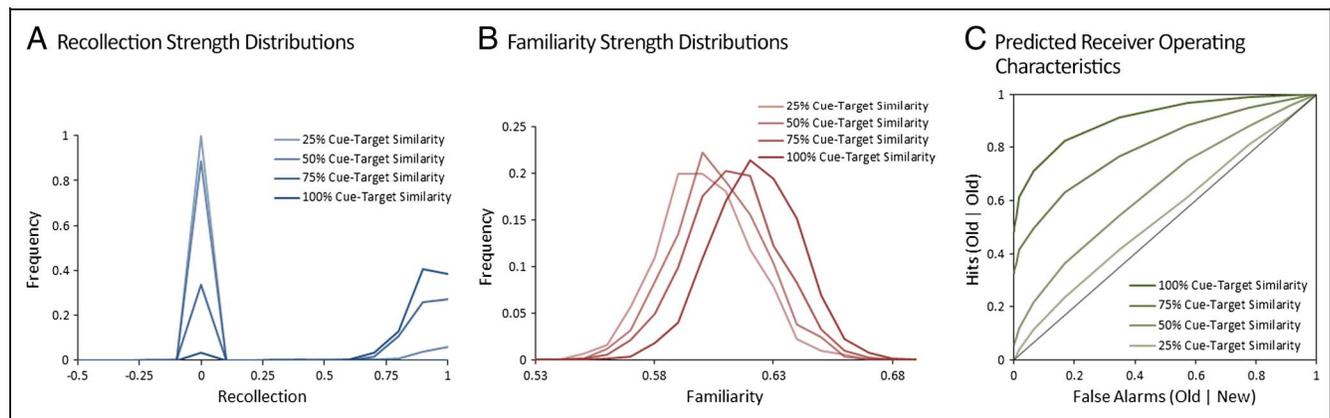


Figure 4. Simulated strength distributions for recollection and familiarity and their resulting ROCs. (A) Histograms of recollection scores at 25%, 50%, 75%, and 100% cue–target similarity, for the hippocampal network. At low levels of cue–target similarity, none of the items led to recollection (i.e., for all items $R = 0$), but as similarity increased, an increasing number of items shifted from the nonrecollected distribution ($R = 0$) to a high level of recollection strength (e.g., $R > .75$). (B) Histograms of familiarity scores for the association cortex network. At low levels of cue–target similarity items are associated with low levels of familiarity and are normally distributed, but as similarity increases the mean of the familiarity distribution increases. (C) Predicted ROCs for each level of target–lure similarity assuming that recognition is based on recollection or on familiarity when recollection fails. The ROCs become more curved as similarity increases, indicating that familiarity is increasing, and the y -intercept increases, indicating that recollection is increasing.

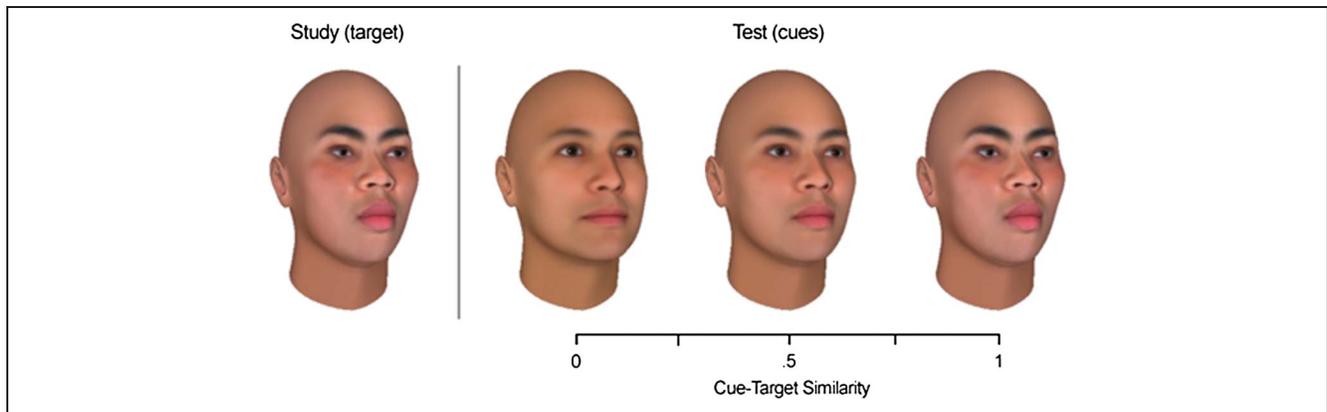


Figure 5. Stimuli for Experiment 1. Participants were presented with a series of distinct faces at study. Test trials always started with the same prototype face (first of the cues, shown here). Over a series of eleven presentations, participants gave old–new confidence judgments—ranging from 1 (*sure new*) to 9 (*sure old*)—as the prototype face was morphed into a studied face (target) or a new face (not shown). At the end of each trial, participants gave a “remember,” “know,” or “new” judgment.

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 hippocampus 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 Modeler, 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 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:

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

where t is cue–target similarity, z_{\min} represents the lower limit of z confidence, Δz 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:

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

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_{max} estimate for “re-

member” ($M = 2.173$, $SE = 0.069$) than “know” ($M = 1.384$, $SE = 0.195$), $t(462) = 3.812$, $p < .001$. The other parameters, z_{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

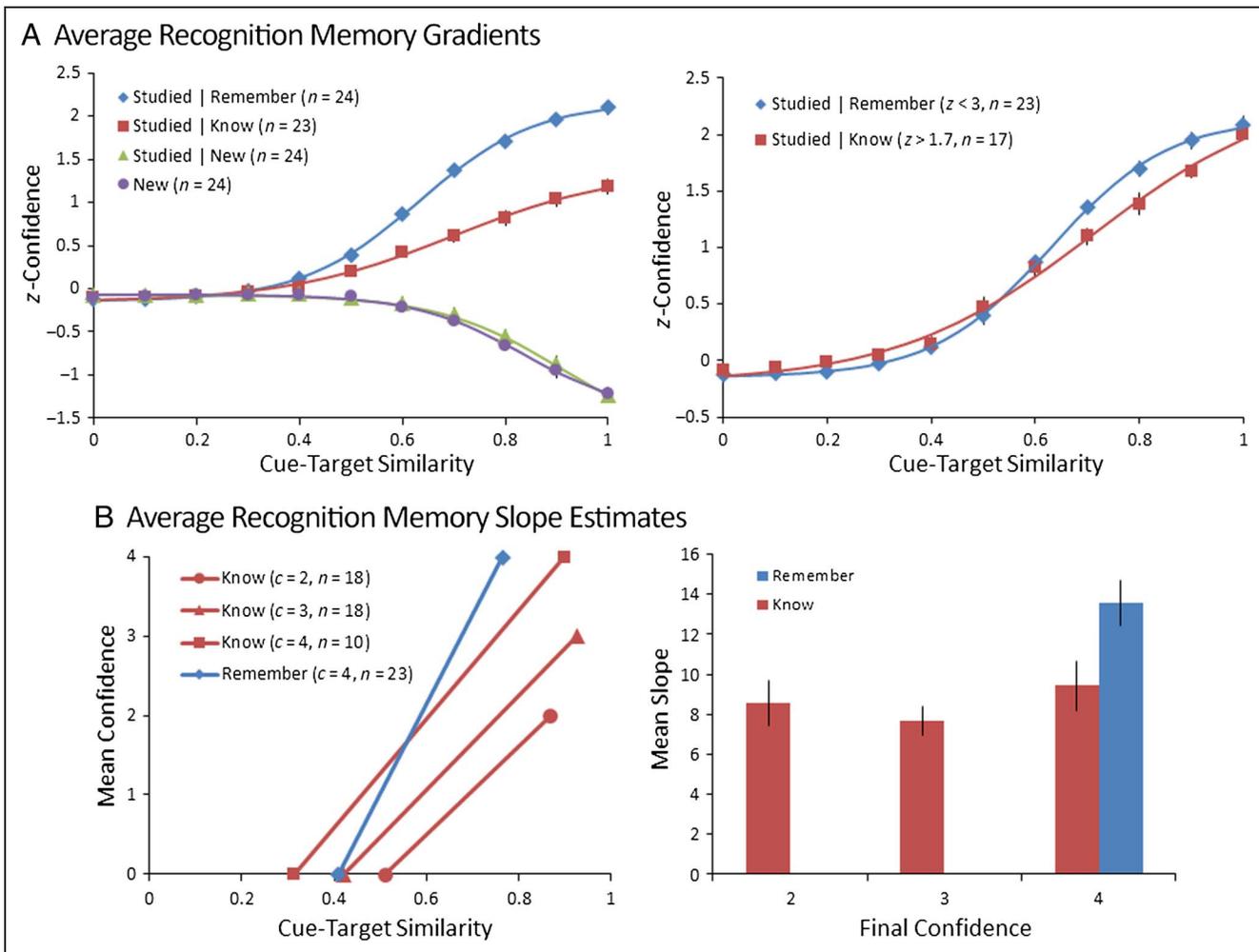


Figure 6. Recognition memory results from Experiment 1. (A) Normalized recognition confidence measures for items receiving a “remember,” “know,” or “new” response as cue–target similarity was varied from 0 to 1. The solid lines represent sigmoid functions that were fit to the observed data. The panel 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 than 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.

“know” functions were statistically different, even when one accounts for differences in the final level of performance. Critically, the recollection gradient was steeper than the familiarity gradient, in line with the predictions of the CLS model.

Matched 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 every 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_1 , the point at which confidence ceased to change; and Δc , 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_1) 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 (DPSD; 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 DPSD model fits (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 (d') 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).

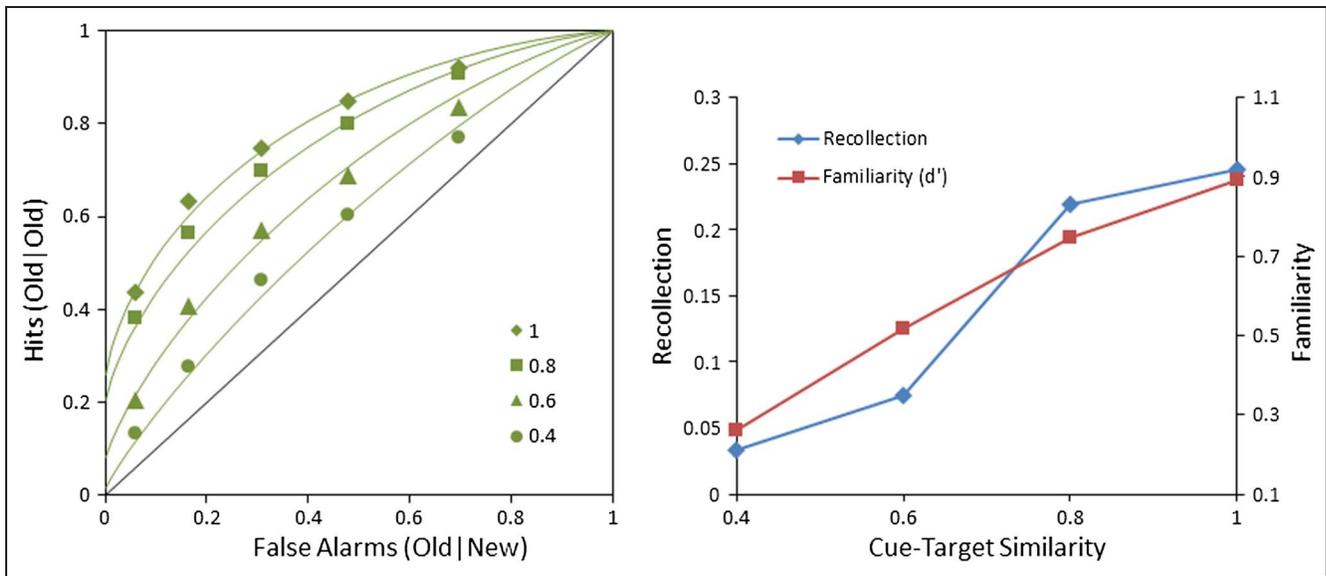


Figure 7. Experiment 2 results showing ROC curves of aggregated subject confidence responses at .4, .6, .8, and 1 cue–target similarity, indicated by the discrete markers, and corresponding DPSD curve fits, indicated by continuous lines (left) and recollection and familiarity estimates based on curve fits at each similarity level (right). Increases in cue–target similarity led to a relatively linear increase in familiarity and a steeper increase in recollection.

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 (Elfman 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 midpoint 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

<i>Layer/Area</i>	<i>Units</i>	<i>Activity (%)</i>
<i>Hippocampus</i>		
EC (in/out)	144	25.0
DG	1600	1.0
Area CA3	480	3.8
Area CA1	384	9.4
<i>Neocortex</i>		
Lower-level cortex (Input)	144	25.0
Association/MTLc	400	20.0

in/out = input and output layers, respectively.

Table A2. Properties of Modifiable Projections

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

- Aisa, B., Mingus, B., & O'Reilly, R. (2008). The emergent neural modeling system. *Neural Networks*, *21*, 1146–1152.
- Atkinson, R. C., & Juola, J. F. (1974). *Search and decision processes in recognition memory*. Oxford: WH Freeman.
- Bakker, A., Kirwan, C. B., Miller, M., & Stark, C. E. (2008). Pattern separation in the human hippocampal CA3 and dentate gyrus. *Science*, *319*, 1640–1642.
- Brown, M. W., & Aggleton, J. P. (2001). Recognition memory: What are the roles of the perirhinal cortex and hippocampus? *Nature Reviews Neuroscience*, *2*, 51–61.
- Colgin, L. L., Leutgeb, S., Jezek, K., Leutgeb, J. K., Moser, E. I., McNaughton, B. L., et al. (2010). Attractor-map versus autoassociation based attractor dynamics in the hippocampal network. *Journal of Neurophysiology*, *104*, 35–50.
- Daelli, V., & Treves, A. (2010). Neural attractor dynamics in object recognition. *Experimental Brain Research*, *203*, 241–248.
- Daselaar, S. M., Fleck, M. S., & Cabeza, R. (2006). Triple dissociation in the medial temporal lobes: Recollection, familiarity, and novelty. *Journal of Neurophysiology*, *96*, 1902–1911.
- Diana, R. A., Reder, L. M., Arndt, J., & Park, H. (2006). Models of recognition: A review of arguments in favor of a dual-process account. *Psychonomic Bulletin & Review*, *13*, 1–21.
- Eichenbaum, H., Yonelinas, A. R., & Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annual Review of Neuroscience*, *30*, 123.
- Elfmán, K. W., Aly, M., & Yonelinas, A. P. (2014). Neurocomputational account of memory and perception: Thresholded and graded signals in the hippocampus. *Hippocampus*. doi:10.1002/hipo.22345.
- Elfmán, K. W., Parks, C. M., & Yonelinas, A. P. (2008). Testing a neurocomputational model of recollection, familiarity, and source recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 752.
- Hasselmo, M. E. (1995). Neuromodulation and cortical function: Modeling the physiological basis of behavior. *Behavioural Brain Research*, *67*, 1–27.
- Hasselmo, M. E., Bodelon, C., & Wyble, B. P. (2002). A proposed function for hippocampal theta rhythm: Separate phases of encoding and retrieval enhance reversal of prior learning. *Neural Computation*, *14*, 793–818.
- Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*, 513–541.
- Jezek, K., Henriksen, E. J., Treves, A., Moser, E. I., & Moser, M. B. (2011). Theta-paced flickering between place-cell maps in the hippocampus. *Nature*, *478*, 246–249.
- Kim, J., & Yassa, M. A. (2013). Assessing recollection and familiarity of similar lures in a behavioral pattern separation task. *Hippocampus*, *23*, 287–294.
- Lacy, J. W., Yassa, M. A., Stark, S. M., Muftuler, L. T., & Stark, C. E. (2011). Distinct pattern separation related transfer functions in human CA3/dentate and CA1 revealed using high-resolution fMRI and variable mnemonic similarity. *Learning & Memory*, *18*, 15–18.
- LaRocque, K. F., Smith, M. E., Carr, V. A., Witthoft, N., Grill-Spector, K., & Wagner, A. D. (2013). Global similarity and pattern separation in the human medial temporal lobe predict subsequent memory. *The Journal of Neuroscience*, *33*, 5466–5474.
- Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, *87*, 252.
- Marr, D. (1971). Simple memory: A theory for archicortex. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, *262*, 23–81.
- McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*, 419.
- Montaldi, D., Spencer, T. J., Roberts, N., & Mayes, A. R. (2006). The neural system that mediates familiarity memory. *Hippocampus*, *16*, 504–520.
- Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychological Review*, *110*, 611.
- Norman, K. A., & Schacter, D. L. (1997). False recognition in younger and older adults: Exploring the characteristics of illusory memories. *Memory & Cognition*, *25*, 838–848.
- O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map* (Vol. 3, pp. 483–484). Oxford, UK: Clarendon Press.
- O'Reilly, R. C., & McClelland, J. L. (1994). Hippocampal conjunctive encoding, storage, and recall: Avoiding a trade-off. *Hippocampus*, *4*, 661–682.
- O'Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience*. Cambridge, MA: MIT Press.
- O'Reilly, R. C., & Rudy, J. W. (2001). Conjunctive representations in learning and memory: Principles of cortical and hippocampal function. *Psychological Review*, *108*, 311.
- Parks, C. M., & Yonelinas, A. P. (2007). Moving beyond pure signal-detection models: Comment on Wixted (2007). *Psychological Review*, *114*, 188–201.
- Preminger, S., Blumenfeld, B., Sagi, D., & Tsodyks, M. (2009). Mapping dynamic memories of gradually changing objects. *Proceedings of the National Academy of Sciences*, *106*, 5371–5376.
- Pustina, D., Gizewski, E., Forsting, M., Daum, I., & Suchan, B. (2012). Human memory manipulated: Dissociating factors contributing to MTL activity, an fMRI study. *Behavioural Brain Research*, *229*, 57–67.
- Ranganath, C., Yonelinas, A. P., Cohen, M. X., Dy, C. J., Tom, S. M., & D'Esposito, M. (2004). Dissociable correlates of recollection and familiarity within the medial temporal lobes. *Neuropsychologia*, *42*, 2–13.
- Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 803.
- Rolls, E. T. (1989). Functions of neuronal networks in the hippocampus and neocortex in memory. In J. H. Byrne & W. O. Berry (Eds.), *Neural models of plasticity: Experimental and theoretical approaches* (pp. 240–265). San Diego, CA: Academic Press.
- Sherry, D. F., & Schacter, D. L. (1987). The evolution of multiple memory systems. *Psychological Review*, *94*, 439.

- Tulving, E. (1985). How many memory systems are there? *American Psychologist*, *40*, 385.
- Wills, T. J., Lever, C., Cacucci, F., Burgess, N., & O'Keefe, J. (2005). Attractor dynamics in the hippocampal representation of the local environment. *Science*, *308*, 873–876.
- Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, *114*, 152.
- Yassa, M. A., & Stark, C. E. (2011). Pattern separation in the hippocampus. *Trends in Neurosciences*, *34*, 515–525.
- Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1341.
- Yonelinas, A. P. (2001). Components of episodic memory: The contribution of recollection and familiarity. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, *356*, 1363–1374.
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*, 441–517.
- Yonelinas, A. P., & Parks, C. M. (2007). Receiver operating characteristics (ROCs) in recognition memory: A review. *Psychological Bulletin*, *133*, 800.
- Yonelinas, A. P., Otten, L. J., Shaw, K. N., & Rugg, M. D. (2005). Separating the brain regions involved in recollection and familiarity in recognition memory. *The Journal of Neuroscience*, *25*, 3002–3008.