

A Single-Trace Dual-Process Model of Episodic Memory: A Novel Computational Account of Familiarity and Recollection

Andrea Greve,^{1,2*} David I. Donaldson,³ and Mark C.W. van Rossum²

ABSTRACT: Dual-process theories of episodic memory state that retrieval is contingent on two independent processes: familiarity (providing a sense of oldness) and recollection (recovering events and their context). A variety of studies have reported distinct neural signatures for familiarity and recollection, supporting dual-process theory. One outstanding question is whether these signatures reflect the activation of distinct memory traces or the operation of different retrieval mechanisms on a single memory trace. We present a computational model that uses a single neuronal network to store memory traces, but two distinct and independent retrieval processes access the memory. The model is capable of performing familiarity and recollection-based discrimination between old and new patterns, demonstrating that dual-process models need not to rely on multiple independent memory traces, but can use a single trace. Importantly, our putative familiarity and recollection processes exhibit distinct characteristics analogous to those found in empirical data; they diverge in capacity and sensitivity to sparse and correlated patterns, exhibit distinct ROC curves, and account for performance on both item and associative recognition tests. The demonstration that a single-trace, dual-process model can account for a range of empirical findings highlights the importance of distinguishing between neuronal processes and the neuronal representations on which they operate. © 2009 Wiley-Liss, Inc.

KEY WORDS: episodic memory model; Hopfield network; recognition; recollection; familiarity

INTRODUCTION

Episodic memory supports the encoding, storage, and retrieval of personally experienced events, each embedded in the context of its acquisition such as when and where the event was experienced. Memory theories have focused heavily on the retrieval mechanisms supporting episodic memory, in particular by asking whether single- or dual-process models best characterize episodic retrieval. Single-process models assume that episodic memories are retrieved by one processes operating on a single representation, whereas dual-process models typically assume two distinct retrieval processes—usually based on two representations. Some dual-process models explicitly assume that retrieval processes operate on distinct memory representations (Atkinson and Juola, 1973, 1974),

whereas others implicitly allude to this view, for example by suggesting that retrieval processes operate on distinct types of information (i.e., perceptual or conceptual) (Mandler, 1980; Jacoby and Dallas, 1981). Of course not all dual-process models aim to describe the underlying neural mechanisms; some are purely cognitive and say very little about the memory representations accessed by the retrieval operations (Jacoby, 1991; Yonelinas, 1994, 2001).

Here, we present a novel computational model of episodic memory that operates on the basis of a single memory representation and is capable of supporting true (one-shot) episodic learning, but which nonetheless uses two distinct retrieval processes. Before introducing our model, we briefly review current single- and dual-process accounts. Single-process models propose that episodic recognition is based on a global index of memory strength, typically called familiarity, which operates as a signal-detection process (Green and Swets, 1966). Retrieval from episodic memory is therefore characterized by d' (the distance between the old and new familiarity distributions), which can be used to identify studied (old) items as more familiar than unstudied (new) items. Empirical data, however, indicate that a single retrieval parameter cannot fully account for all episodic memory findings (Clark and Gronlund, 1996; Yonelinas, 2002).

To account for empirical findings, a variety of dual-process models have been put forward, all of which propose that recognition memory is supported by two distinct retrieval processes—familiarity and recollection—which differ in their speed of operation and specificity of retrieved information (for review see Yonelinas, 2002; Rugg and Yonelinas, 2003; Diana et al., 2007; Eichenbaum et al., 2007). In general, dual-process models typically regard familiarity as a fast acting process which produces a purely quantitative, “strength-like” signal, without retrieval of details from the study episode. In contrast, recollection is typically described as a relatively slow, more controlled process that yields extensive qualitative information about the study episode, including contextual information regarding when and where an item was studied (often, but not always, governed by a threshold process). This distinction is supported by empirical findings. For example, evidence that familiarity is available earlier than recollection comes from studies forcing subjects to make speeded recognition responses, which has small effects on familiarity but

¹Wales Institute of Cognitive Neuroscience, School of Psychology, Cardiff University, Park Place, Cardiff, United Kingdom; ²Institute for Adaptive and Neuronal Computation, School of Informatics, University of Edinburgh, Edinburgh, Scotland, United Kingdom; ³Department of Psychology, University of Stirling, Stirling, Scotland, United Kingdom

Grant sponsors: EPSRC/MRC, Wales Institute of Cognitive Neuroscience.

*Correspondence to: Andrea Greve, Wales Institute of Cognitive Neuroscience, School of Psychology, Cardiff University, Park Place, Cardiff, CF10 3AT, United Kingdom. E-mail: greve@cardiff.ac.uk

Accepted for publication 24 February 2009

DOI 10.1002/hipo.20606

Published online in Wiley InterScience (www.interscience.wiley.com).

results in a robust reduction on recollection (Yonelinas and Jacoby, 1994; Gronlund et al., 1997; Hintzman and Caulton, 1997). The nature of familiarity and recollection memory often leads to the assumption that familiarity memory has a higher capacity than recollection memory, although this does not appear to have been verified experimentally. However, empirical findings do reveal that elaborative encoding such as meaningful processing compared with perceptual processing is beneficial to recollection, whereas familiarity is more susceptible to perceptual changes and fluency manipulations (Mandler, 1980; Gardiner and Java, 1990). Furthermore, since only recollection provides qualitative information about a study event, associative recognition tests that require retrieval of detailed or contextual information (e.g., spatial location, sensory modality, initial pairing with another item) engage predominantly recollection-based retrieval (Hockley and Consoli, 1999; Yonelinas, 1999).

Further evidence that familiarity and recollection are marked by distinct processes comes from receiver operating characteristics curves (ROCs) which plot hits against false alarms as a function of confidence. ROC analyses reveal overall asymmetric shaped curves composed of two independent processes: a symmetrical, curvilinear component associated with a signal detection like familiarity process and an asymmetrical, linear component linked to a threshold like recollection process (Yonelinas, 1994, 1997, 2002; Fortin et al., 2004).

Although behavioral characterizations of retrieval clearly support dual-process models, there is to date little consensus concerning which anatomical brain structures support familiarity and recollection. One predominant view is that activity in the perirhinal cortex and hippocampus are linked to familiarity and recollection processes, respectively. However, this view is currently contentious, as some studies fail to find supporting evidence (Hamann and Squire, 1997; Manns et al., 2003; Wais et al., 2006). Although there is no doubt that the medial temporal lobe (MTL) plays a critical role in episodic memory, it remains unresolved whether the hippocampus and other MTL areas make distinct contributions to retrieval, and if so, what these contributions are. Nonetheless, currently the dominant view is that recollection and familiarity are associated with distinct (if overlapping) brain regions.

The finding that familiarity and recollection are associated with activity in two distinct sets of brain regions can, however, be interpreted in more than one way. It is possible that these two sets of brain regions store two distinct memory traces for the same event. According to this view, information supporting later recollection and familiarity is encoded into memory separately, and during retrieval the recollection and familiarity processes access information stored in these distinct traces, which is reflected in the activation of the corresponding brain areas. Alternatively, familiarity and recollection could both operate on a single memory trace, with the subsequent retrieval output of the two processes leading to activation in different brain areas. One important consequence of this second interpretation is that it does not necessitate the storage of two distinct memory traces. By this view, activity in different brain areas reflects the outcome of familiarity and recollection processes that originate

from a single memory trace. To our knowledge there is currently no evidence that directly discriminates between these competing viewpoints.

In this article, we use a computational model to explore this issue. Most biologically inspired computational models concentrate on the role that the hippocampus plays in recollection (Marr, 1971; McNaughton and Morris, 1987; Treves and Rolls, 1994; Burgess and O'Keefe, 1996; McClelland and Goddard, 1996; Hasselmo and Wyble, 1997). Although these models differ in their specific details, they are in overall agreement as to how the hippocampus supports episodic retrieval: it forms a compact code of mostly nonoverlapping representations in a recurrently connected network. The recurrent network binds different features into a single episode, so that a partial cue can evoke retrieval of the complete episode.

In contrast to hippocampal models of recollection, only a few models attempt to simulate familiarity-based retrieval (Sohal and Hasselmo, 2000; Bogacz and Brown, 2003; Norman and O'Reilly, 2003; Meeter et al., 2005; Norman et al., 2005). Although some models propose a specialized network for familiarity discrimination, others advocate that familiarity emerges from a feature extraction process (Norman and O'Reilly, 2003).

Existing computational models of recollection and familiarity have provided important insights into the neuronal basis for these memory processes, producing biologically plausible accounts of episodic retrieval. To our surprise, however, few attempts have been made to formalize dual-process theories of episodic memory in a single computational model. Yet, in practice, it is difficult to fully evaluate the performance of existing familiarity-only or recollection-only models by comparing them with experimental findings, because assuming that the dual-process view is correct, both familiarity and recollection typically support performance. Thus, there remains the need for models that account for episodic performance as the combined outcome of familiarity and recollection processing.

One such attempt is the complementary learning system (CLS), originally proposed by McClelland and Goddard (1996), which implements the idea that the neocortex and hippocampus perform familiarity- and recollection-based retrieval, respectively (Norman and O'Reilly, 2003). The hippocampus model uses distinct pattern separated representations to enable rapid storage of items. This is achieved by recruiting five different network layers which are linked by different feed-forward, feed-back, and recurrent connections. The model uses Hebbian learning to store a sparse representation of the input. At retrieval, the model performs actually cued recall, by generating a complete version of a studied pattern in response to a partial input cue. The level of corresponding activity between the presented input pattern and the retrieved output pattern is used to assess retrieval success.

The neocortex in this model, by comparison, encodes regularities present in the environment and gradually forms, through slow incremental learning, an internal model of the outside world. Learning in the neocortex is based on a feature extraction process which produces sharpened representations for

repeatedly presented items and overlapping representations for similar items. Consequently, many neurons will show a weak activation in response to novel stimuli, whereas familiar (repeatedly presented) stimuli evoke a strong activation in relatively few neurons. In general, the MTL model describes familiarity as a by-product of a feature extraction process which is a function of the repeated exposure of a stimulus. Conceptually, we view this formulation of familiarity to diverge from with the idea of episodic retrieval; a defining characteristic of episodic memory is the ability to recognize events that have occurred just once (Tulving, 1972). And of course, if events are repeated, other forms of memory such as semantic retrieval may also support performance (a possible confound in the interpretation of such data).

Here, we consider familiarity and recollection memory in one-shot learning memory tasks. We propose a single model, delineating two separate retrieval processes that act on a single representation, but provide two distinct outputs (i.e., assessments of memory). The model follows the dual-process tradition: recognition memory is still supported by two distinct retrieval processes even though they operate within a single network. Modeling familiarity and recollection within a single network has the advantage that no distinct representations and encoding processes are required. Thus, the network proposed here is a simple and parsimonious model of episodic memory retrieval. Moreover, our model emphasizes that a single representation may support multiple independent processes.

In the following sections, we first describe how the familiarity and recollection processes are implemented within a single network and show how these processes can discriminate novel from previously studied patterns. By using experiments that parallel empirical studies, we assess the overall performance of the model in relation to recent findings in item and associative recognition tests. Crucially, we demonstrate that our model has characteristics matching those reported in empirical investigations, such as the time-course of retrieval, the capacity of the retrieval processes, and the shape of ROC curves. Finally, using simulations that parallel empirical studies we compare the performance of the model with recent findings in item and associative recognition.

METHODS

Modeling Framework

Our model uses a standard Hopfield type network with $N = 1,000$ binary valued units. The units are recurrently connected in an “all-to-all” fashion, but without self-connections. During the training phase (memory encoding), a set of patterns is presented to the network. Each pattern \mathbf{x}^μ is an N -dimensional vector (henceforth, μ labels the patterns). The patterns are stored by training the connection weights between neurons via Hebbian learning. The strength of the connection (w_{ij}) between neuron i and j is determined by the outer product of

the patterns, summed across the total number (p) of stored patterns,

$$w_{ij} = \frac{1}{N} \sum_{\mu=1}^p \mathbf{x}_i^\mu \mathbf{x}_j^\mu (1 - \delta_{ij}) \quad (1)$$

where the Kronecker delta, δ_{ij} is one if $i = j$ and 0 otherwise, ensuring that $w_{ii} = 0$.

During the test phase (memory retrieval) both learned and new patterns are presented to the network and a variety of readouts are measured. As is common in these models, the weights are kept fixed during the testing phase.

In the standard Hopfield network typically half of the units are simultaneously active; in most brain systems however the mean activity is low, and only a few neurons are active at any one time (Johnston and Amaral, 1998). For instance, within the dentate gyrus of the MTL, which is thought to play a role in memory processing, activity is known to be sparse (Barnes et al., 1990). We incorporate such sparseness in the model by decreasing the probability of a unit being in the active (+1) state. The memory patterns (\mathbf{x}^μ) are generated by setting each unit randomly and independently to a value of 1 or -1 , with a probability of $(1 + a)/2$ and $(1 - a)/2$, respectively. This can be formally expressed with:

$$p(\mathbf{x}^\mu) = \prod_{i=1}^n p(x_i^\mu) \quad \text{with} \\ p(x_i^\mu) = \frac{1+a}{2} \delta(x_i^\mu - 1) + \frac{1-a}{2} \delta(x_i^\mu + 1) \quad (2)$$

The parameter a indicates the average activity of the patterns. In the simulation presented in this article, sparse patterns are created by randomly generating patterns with an average of 1% active and 99% silent units, corresponding to $a = -0.98$. Note, however, that the actual number of active and silent units across patterns varies around the mean level of sparseness. When storing sparse patterns, the more general covariance rule provides a higher storage capacity (Tsodyks and Feigel'man, 1988):

$$w_{ij} = \frac{1}{N} \sum_{\mu=1}^p (\mathbf{x}_i^\mu - a)(\mathbf{x}_j^\mu - a) - c \quad (3)$$

Global inhibition is implemented in the model by subtracting a constant (c) from the weight matrix. This constant represents a term that keeps the retrieval sparseness of the network identical to the input sparseness and its value was determined from simulations. Equating input and output sparseness is essential for maintaining storage capacity at low values of a , but is known to have no effect on the familiarity measure used here (Bogacz and Brown, 2002).

Familiarity Process

A familiarity process is initiated with the presentation of a test pattern to the network. At time $t = 0$ familiarity is assessed. The output value is dependent on the energy function of the attractor network as a whole. This characterization of familiarity has been previously used by Bogacz et al. (2001) and is defined by:

$$E^{\mu} = -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \mathbf{x}_i^{\mu} \mathbf{x}_j^{\mu} w_{ij} \quad (4)$$

This energy value is low for previously stored patterns and high for new patterns. It is worth noting that, for this type of familiarity discrimination, the Hopfield network is engaged very differently from its typical application. The network does not first undergo a relaxation process which would allow the pattern to settle into an attractor state. Instead, the energy (familiarity) is evaluated at the very first time step. It has been shown that the covariance learning rule [Eq. (3)] is also optimal for the familiarity detection; it gives the highest signal-to-noise ratio (SNR) of the energy in the limit that N approaches infinity (Greve et al., in press).

Recollection Process

The presentation of a test pattern to the network also initiates the recollection process. At time $t = 0$ the activity of the network state vector \mathbf{s} is set to the input pattern \mathbf{x}^{μ} . Next, the state of the network is asynchronously updated according to:

$$s_i(t+1) = \text{sign}\left(\sum_{j=1}^n w_{ij}s_j(t)\right) \quad (5)$$

until an attractor state is reached. A single update of all neurons in the network is referred to as a cycle, and the model settles over multiple cycles. Dependent on how closely the final attractor state resembles the initial pattern, the test pattern will either be classified as an old (previously stored) or new pattern.

Our recollection measure calculates the amount of change the state undergoes when it is presented initially to the network compared with when it is relaxed into a final attractor state. If a previously studied pattern is presented to the network, the stored attractor state should be identical to the presented test pattern. Therefore, the pattern will undergo no (or few) changes when settling into its corresponding attractor state. However, when tested with a new pattern, the network will settle into an attractor state that is quite different from the initial test pattern. A distance measure captures this change by calculating the normalized dot product:

$$d(\vec{\mathbf{a}}, \vec{\mathbf{b}}) = \frac{1}{2} \left(1 - \frac{\vec{\mathbf{a}} \cdot \vec{\mathbf{b}}}{|\vec{\mathbf{a}}| |\vec{\mathbf{b}}|} \right) \quad \text{with} \quad (6)$$

$$\vec{\mathbf{a}} = \mathbf{s}(t=0), \quad \vec{\mathbf{b}} = \lim_{t \rightarrow \infty} \mathbf{s}(t)$$

If the test pattern has been learned previously, the distance measure (d) should be close to zero. By contrast, an unlearned pattern results in a distance larger than zero.

Thresholds and SNRs

It is necessary to set appropriate decision thresholds for the energy and distance measure to differentiate between studied patterns and new patterns. The threshold is established by first generating patterns with given properties (i.e., level of sparseness, number of patterns, and degree of overlap). Those patterns are presented to the network in order to establish the distribution of the old and new items. The decision threshold for both the recollection and familiarity detection is set such that the total error (false positive plus false negatives) is minimal. Once a decision threshold is determined it is applied and kept constant during subsequent tests of the network. Although the same principle is used to set the decision thresholds for the familiarity and recollection process, the values are calculated separately for the two processes and can therefore be considered distinct.

The quality of the separation between responses to old (forming one distribution, d_1) and new patterns (forming a second distribution, d_2) can be quantified with the SNR:

$$\text{SNR} = 2 \frac{|\mu_{d1} - \mu_{d2}|}{\text{sd}_{d1} + \text{sd}_{d2}} \quad (7)$$

where μ_{d1} and μ_{d2} are the mean of d_1 and d_2 , and sd_{d1} and sd_{d2} are the variance of d_1 and d_2 , respectively. The SNR for both familiarity and recollection is used to characterize performance of the network.

ROC Curves

ROC curves provide more information about discrimination performance than the SNR. ROC curves are constructed by plotting hit rates (correctly identified old patterns) relative to false-alarm rates (incorrectly classified new patterns) by varying the decision threshold. The proportion of hits and false alarms can be expressed by the discriminability measure d' , which is defined in terms of z scores (a z transformation, defined by the inverse of the normal distribution, converts hit and false-alarm rates into standard deviation (SD) units):

$$d' = z(\text{hits}) - z(\text{false alarms}) \quad (8)$$

Points on the ROC curve are generated by computing d' while varying the decision threshold. When discrimination performance is at chance ($d' = 0$), the ROC curve is the major diagonal (where hit and false-alarm rates are equal). When performance increases, the ROC curve shifts toward the upper left corner, where discrimination is perfect (where hit rate = 1 and false-alarm rate = 0).

The z -transformed ROC curves are obtained by plotting the z -transformed hit rate versus the z -transformed false-alarm rate. The shape of the resulting z -ROC is informative about the underlying distribution of old and new signals. Those z -ROC

curves which have a slope of one (i.e., are symmetrical around the minor diagonal), are defined by:

$$d'(1 - \text{hit}, 1 - \text{false alarm}) = d'(\text{hit}, \text{false alarm}) \quad (9)$$

For example, z -ROC curves have a slope equal to one if hits and false alarms are described by two Gaussian distributions with identical variances. Asymmetrical z -ROC curves with a slope unequal to one could originate from old and new distributions of unequal variance. However, threshold theories account for asymmetrical ROC curves by assuming that the decision space is characterized by discrete states; while some states reflect decisions purely based on hits, others show a proportional contribution of hits and false alarms. Thus, by this view, the asymmetry of ROC curves allows information to be derived about the contribution of recollection and familiarity to retrieval. In practice, the slope of the z -ROC curve is calculated by using a least-square fit of the data points.

Correlated and Mixed Patterns

We assessed the performance of our network with different types of stimuli. Our simulation with correlated patterns used patterns that were constructed by duplicating the values of 80% of the entries from a studied pattern and combining this with 20% randomly generated new entries (according to the given sparseness level). A second simulation tested associative recognition analogous to a recent empirical investigation (Greve et al., 2007). Item pairs were created by assigning the first half of a pattern to represent one item and the second half to represent another item. We measured how well the network discriminated studied item pairs against two classes of recombined lures (old–old and old–new) and genuinely novel pairs (new–new). The two classes of recombined lures were formed by combining the first half of one vector (50% of the units) with the second half of another vector (50% of the units).

RESULTS

We present a model of episodic memory retrieval that simulates both familiarity and recollection. The aim of this model is to investigate whether two distinct retrieval processes could differentially access a single memory trace. The model is based on the architecture of a Hopfield network, in which binary units are connected through trainable weights (Hopfield, 1982).

The network was designed to mimic human performance on an associative recognition memory task. In such tasks, one presents lists of stimulus pairs (such as words) in the study phase. Memory for these pairs is subsequently tested with lists containing intermixed studied and new word pairs. The participants are asked to discriminate studied from unstudied

pairs, causing familiarity-based and/or recollection-based retrieval.

Familiarity and Recollection Processes

The familiarity process implemented in the current network computes an energy value as soon as a test pattern is presented to the network. This account of familiarity is illustrated in Figure 1, and was originally proposed by Bogacz et al. (2001). If the network has been previously trained with the test pattern, the energy measure will have a low value, whereas untrained or new patterns yield high energy values. The energy is obtained in the very first time step, so that no dynamical process is engaged (as noted earlier, this is not how Hopfield networks are typically used), consistent with the rapid discrimination performance associated with familiarity. The continuum of low to high energy values signals more or less familiarity. A suitable decision threshold is used to discriminate between old and new patterns.

In contrast to familiarity, the recollection process does engage the dynamical evolution of the network once a test pattern is presented. As time passes, a Hopfield network will always settle into a memory state (although this may be a spurious memory state). Interestingly, there is no established method to distinguish true from false memories, that is, to determine whether the retrieved state is a true memory state or whether the retrieved state is irrelevant to the cue. Consequently, Robins and McCallum (2004) proposed a procedure that allows the discrimination between learned and spurious states in a Hopfield Network. Their approach is based on the “energy ratio,” which is the energy of the three lowest energy units of a pattern divided by the energy of the three highest energy units. This energy ratio is able to discriminate whether a given attractor state is learned or spurious. However, as new patterns can converge to attractor states which are either learned or spurious, this measure does not translate into a suitable recollection measure that can reliably discriminate previously learned from unlearned patterns. An alternative method is to design the network such that unlearned stimuli always evolve into the zero state in which all units are silent; however, this works best in the limit of extreme sparse codes (Buhmann et al., 1989).

Our model implements a recollection process that takes advantage of the different settling dynamics between studied and new patterns. The recollection measure computes the change that the network state undergoes from when it is presented initially with the test pattern compared to when it is relaxed into a final attractor state (illustrated in Fig. 1). The basic idea of this approach is that if a previously studied pattern is presented to the network at test, the final attractor state should be (almost) identical to the presented test pattern. Thus, for studied patterns, the state will undergo relatively few changes when settling into the learned attractor state. By contrast, if a new pattern is presented at test, the network will settle into an attractor state that is likely to deviate substantially from the initial state. In practice the extent or degree of deviation can be characterized by the number of changes in the net-

work state. An appropriate decision threshold for these changes can then be used to discriminate recollected from not recollected patterns. Norman and O'Reilly (2003) adopted a similar

approach by using the amount of match and mismatch between a test probe and network output to evaluate "recall" performance.

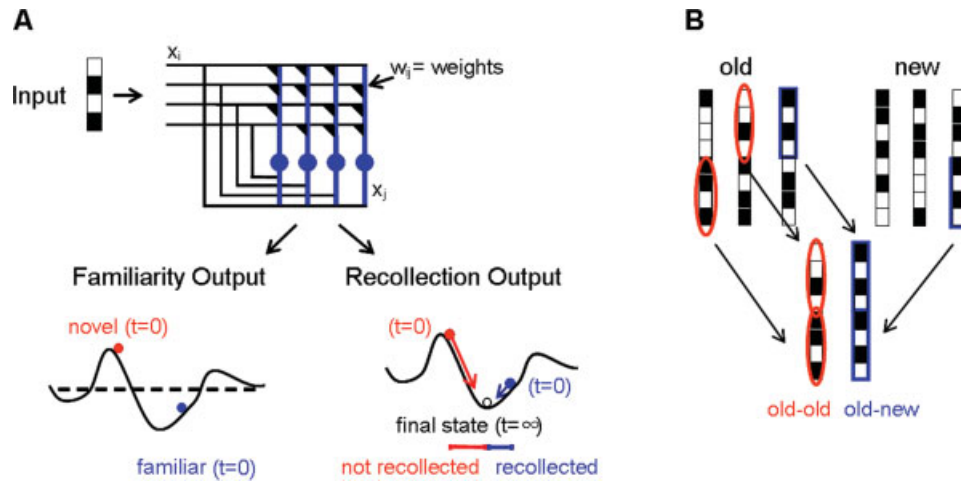


FIGURE 1. A: Overview of the network. The input pattern (simplified in the figure to four units for illustration) provides activity x_i to the network. The connecting weights (w_{ij}) recurrently link all the neurons in the network together. The training phase changes the connection weights (w_{ij}) so that learned patterns are associated with themselves. When the network is working perfectly and a partial input of a trained pattern is presented in the test phase, the network recovers the trained representation as output (x_j). The patterns are classified as familiar or novel on their energy level at time $t = 0$. In the familiarity example, the left pattern has a high energy and is judged as novel, whereas the right pattern has a low energy and is judged as familiar. Whether a pattern is recollected is determined by the amount of change a pattern undergoes between $t = 0$ and $t = \infty$. In the recollection example, the left

pattern covers a long distance in phase space and is not recollected, whereas the right pattern covers a short distance and is recollected. B: A schematic overview of the creation of mixed test patterns used in some simulations. The network has learned a set of training patterns (old) and is presented with test patterns constituting either unlearned (new) or partially learned but recombined (mixed) patterns. Two different classes of mixed patterns were generated dependent on the type of the first and second pattern. One class mixes two previously studied patterns (old-old) (light red marking). A second class combines half-a-studied pattern with half-a-new pattern (old-new) (dark blue marking). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

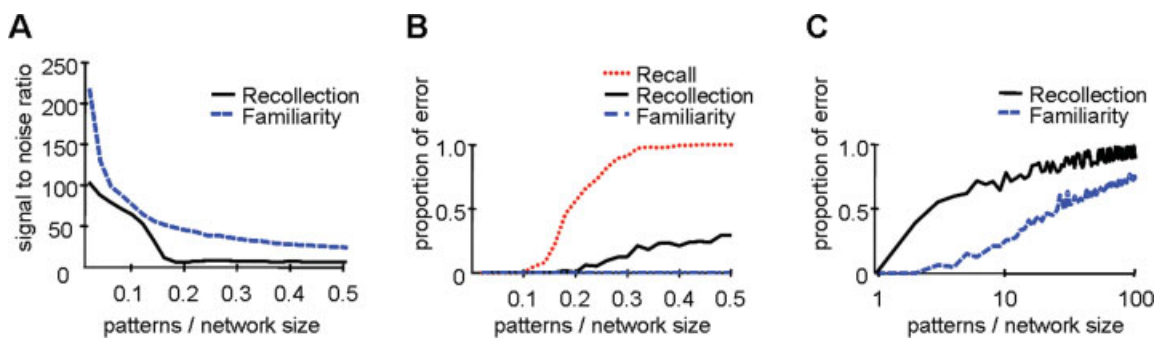


FIGURE 2. Familiarity and recollection performance. A: The signal to noise ratio (SNR) for the familiarity (blue dashed line) and recollection (black solid line) in a network of 1,000 units trained and tested with an increasing number of nonsparse, uncorrelated patterns. The x -axis indicates the number of patterns relative to the number of network units. The SNR indicates successful discrimination for both processes, but familiarity is superior to recollection. Overall, performance decreases as the number of stored patterns increases. B: The fraction of misclassified patterns (proportion of errors including false positives and false negatives) is shown when the decision threshold is set to produce minimum error for familiarity (blue dashed line) and

recollection (black solid line), along with recall (red dotted line). The recall measure is the typical measure used to assess Hopfield network capacity. Even when recall capacity is reached (i.e., when error plateaus) familiarity and recollection continue to provide appropriate discrimination, as indicated by their greatly reduced error rates. C: The proportion of errors obtained for familiarity (blue dashed line) and recollection (black solid line) for very many patterns (note the change in the scale of the x -axis compared with panel B). Error rates are greater, and build up more quickly for recollection than familiarity. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Evaluation of Familiarity and Recollection

A direct quantitative comparison between neuronal network models and psychological experiments is extremely difficult. First, the experiments vary widely in the precise experimental conditions; absolute performance levels are extremely variable. Second, and more fundamentally, it is not known how the network parameters should be exactly chosen to model human cognition. Thus, here we first evaluate the performance of the model by studying how the capacity of familiarity and recollection changes in response to different model parameters such as network size, sparseness, or correlated patterns. We use capacity as means of assessing memory performance, whereby lower capacity corresponds to higher error rates in empirical experiments. This approach allows us to encompass different experimental observations that were obtained under varying conditions. Subsequently, to explicitly contrast the model with observed data, we conducted tests to gauge how the network's performance corresponds to empirically obtained data that show a differential contribution of familiarity and recollection, for example to plurality discrimination and associative recognition.

To determine how well the model discriminates old from new items the network was trained with randomly generated input patterns (using the standard covariance learning rule, *c.f.* Methods section). In addition, we presented an increasing number of stimuli to test how familiarity and recollection performance changes with growing "memory load." In accordance with the empirical literature, the term "memory load" in the network model refers to number of patterns stored. The test session was modeled by presenting previously trained patterns, together with randomly generated new patterns. We use the SNR to assess the performance of familiarity and recollection (Fig. 2A). A larger SNR corresponds to better discrimination. An SNR of one corresponds to 30% false positives and false negatives, which although poor, still indicates above chance performance.

Figure 2A illustrates the SNR values, plotted as a function of the number of patterns to be learned. Both familiarity and recollection can successfully discriminate "old" from "new" patterns, and the SNR is much larger than one when only a few patterns are stored. However, overall discrimination ability decreases with increasing memory load. In particular, the SNR for recollection declines notably at the point where the number of stored patterns reaches about 15% of the number of units. This corresponds to the storage capacity of Hopfield networks with Hebbian learning of nonsparse patterns (Hertz et al., 1991). Above this capacity limit, the attractor deforms and the activity is caught in spurious attractors. This is in agreement with the empirical work that dissociates familiarity from recollection by showing that increased memory (*i.e.*, list length) interferes primarily with recollection but leaves familiarity relatively unaffected (Yonelinas, 1994, Yonelinas and Jacoby, 1994). Despite the noticeable drop in SNR for recollection beyond the capacity limit, the recollection measure is still able to distinguish old from new patterns. Even when the number of patterns exceeds recall capacity, recollection is still able to reasonably discriminate between old and new patterns (Fig. 2B).

Although a comparison of recall and recollection is not the primary aim of the current article, the greater capacity for recollection compared with recall is noteworthy and highlights the importance of distinguishing between the two. Both processes depend on the final attractor state. However, the outcome of the attractor state is evaluated differently in each case. Accurate recall requires the attractor state to be identical to the learned patterns. In contrast, correct recollection can occur even when only parts of the attractor states are identical to the learned pattern, as happens when the attractor of a particular item is altered by learning other patterns. Although recall of that particular item would fail, recollection can still be successful. Hence, recollection can occur even when there is a reduced correspondence between the stored attractor state and test pattern. In memory experiments, this would correspond to the case where subjects correctly identify studied patterns but make errors on free or cued recall as reported by Tulving and Pearlstone (1966).

There is no direct empirical demonstration in the literature that familiarity and recollection differ in the number of events they can store and retrieve. For example, although Standing (1973) demonstrated huge capacity for recognition memory *per se*, this ability is not allied to familiarity or recollection. Similarly, Brown and Aggleton (2001) imply a high capacity for familiarity, but make no direct comparison with the capacity of recollection. Nonetheless, within the memory literature the nature of familiarity and recollection often leads to the implicit assumption that we are familiar with more events than we can recollect. Our model turns this assumption into an explicit prediction; the simulations reveal that the capacity for familiarity discrimination is much greater than either the recollection or recall capacity (Figs. 2A,B). Familiarity discrimination has a capacity of order N^2 , which is proportional to the number of synapses rather than the number of neurons (as derived by Bogacz et al., 2001). The difference in capacity for familiarity and recollection discrimination is further illustrated in Figure 2C, a rescaled and extended version of the data in 2B, which shows a greater proportion of errors arise for recollection compared with familiarity, revealing an overall larger capacity for familiarity. Recollection errors increase rapidly, and ultimately approach complete failure of memory, confirming that memory load predominantly interferes with recollection. For very many patterns, errors also occur for familiarity, showing that successful familiarity discrimination has a large, but finite capacity (errors for familiarity are negligible when storing a few patterns, Fig. 2B).

The difference in storage capacity between familiarity and recollection is also reflected in their underlying old/new distributions. Figure 3 illustrates such changes in the joint distribution of familiarity and recollection values. For both familiarity and recollection, the old/new distributions show a clear separation when using 100 patterns (Fig. 3A). However, if the number of patterns is increased to 250 (Fig. 3B), the recollection measure shows a decreased separation between the old and new distributions, whereas the familiarity measure shows no significant change.

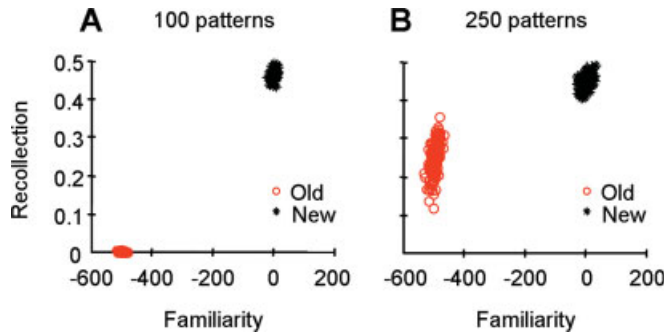


FIGURE 3. The joint distribution of familiarity (i.e., energy values) and recollection (i.e., distance values) associated with 100 (A) or 250 (B) different patterns. The old distribution (black circles) and new distribution (blue stars) are well separated when the number of patterns is low (A). By contrast, for an increasing number of patterns (B) the recollection measure shows an increasingly overlapping distribution for old (black circles) and new (blue stars) patterns. Nevertheless, the familiarity distributions remain virtually unchanged, maintaining a similar distance, reflecting its higher capacity. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

These results demonstrate that, despite operating within a single memory network, both our familiarity and recollection processes can discriminate between previously studied and new patterns, but they do so differently. Although an increased memory load interferes predominantly with recollection, familiarity reveals an overall higher storage capacity. However, both familiarity and recollection discrimination have a higher capacity than recall.

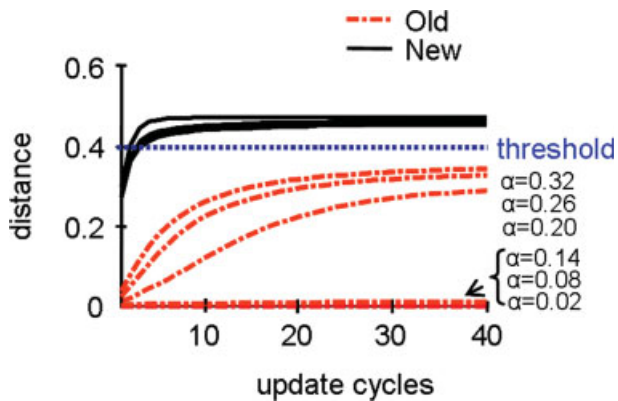


FIGURE 4. The evolution of the recollection measure occurs over multiple update cycles as old and new patterns settle into their final attractor state. A single update of all neurons in the network defines one cycle, that is, 1,000 updates. Different numbers of study and test patterns are used (α indicates the number of test patterns by network size ratio). The distance measure can discriminate between old and new patterns already during the dynamic settling process. Old patterns (red dashed lines) are associated with systematically lower distances compared with new patterns (black solid lines), even when a recall limit is reached. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Constrained Recollection

It is evident from numerous experiments and personal experiences that the recollection process is fallible and not every recollection attempt necessarily results in a retrieval output. However, as noted in the Methods section, during recollection the state of the network will always converge to an attractor state, no matter what the input was.

To allow memory failure to occur in our network, the retrieval mechanism is designed to terminate as soon as a state changed significantly above decision threshold during the settling process. The retrieval dynamics associated with recollection are illustrated in Figure 4, which plots the recollection value (i.e., distance) as a function of the number of network update cycles. Given this termination constraint, an attractor state only provides a retrieval output when test patterns are recollection as previously studied, whereas new patterns can be rejected without reaching a final attractor state.

New patterns reach the critical recollection threshold of 0.4 (which discriminates between old and new patterns) early in the settling process (Fig. 4). Within the first five cycles, new patterns exceed the recollection decision threshold, even though some of the old patterns have not yet settled in their final attractor state. In particular under high memory load ($\alpha > 0.2$), new patterns are identified long before the final retrieval state of an old pattern is reached.

Sparse Patterns

Next, we examine how performance is affected by sparse patterns (representations with only a small fraction of active neurons). This simulation is not only motivated by the theoretical knowledge that sparseness affects the capacity of a network, but also by empirical findings that the MTL, a brain area closely linked to memory processing, operates on nonoverlapping,

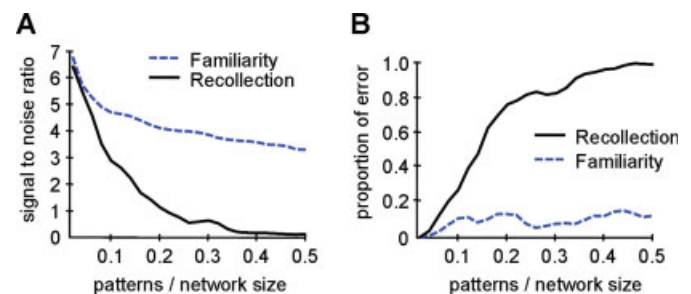


FIGURE 5. A: The signal to noise ratio for the familiarity (blue dashed line) and recollection (black solid line) when the network is trained with sparse patterns. The SNR is much smaller than for nonsparse patterns (Fig. 2), yet both familiarity and recollection allow old and new items to be discriminated, whereas familiarity is superior. B: The amount of misclassification (false positives and false negatives) when the decision threshold for discriminating old or new patterns is set to the point of minimum error (i.e., the point where the old and new distributions intersect). For sparse patterns both recollection and familiarity produce more errors than for nonsparse patterns, but recollection remains inferior. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

sparse representations (Quian Quiroga et al., 2005; Waydo et al., 2006; Quian Quiroga et al., 2008); sparse coding is presumably more biologically realistic. We assume that the sparse inputs presented to the network result from a preceding process, accomplished by other areas of the brain that produce sparse outputs. If the results reveal preserved old/new discrimination under sparse coding, it would demonstrate that our model is in line with behavioral findings and can operate under sparse coding. The network was trained and tested with the covariance rule, using patterns containing 1% active and 99% inactive units, whilst the learning rule was modified to optimally store sparse patterns (see Methods section).

Familiarity- and recollection-based retrieval can successfully discriminate between sparsely coded old and new patterns (Fig. 5), but performance declines compared with the nonsparse case (c.f. Fig. 2). It is known that familiarity performance decreases for sparse patterns, although this decrease is dependent on the precise implementation of sparseness (Bogacz and Brown, 2002). Even though sparse patterns are generally associated with decreased SNR compared with nonsparse patterns, the overall pattern of higher SNR for familiarity compared with recollection is preserved for sparse pattern (although the relative difference is reduced).

In comparison with nonsparse patterns, the familiarity and recollection measures for sparse patterns reveal a greater overlap between the old and new distributions. In particular, an increase in the number of stored patterns influences the recollection discrimination (i.e., it results in a greater overlap in the distribution for old and new patterns), but the familiarity-based distributions for old and new patterns are not significantly altered. In summary, although sparse patterns decrease the discrimination performance of both retrieval processes, the pattern of results is very similar to that of simulations with nonsparse patterns.

Receiver Operating Characteristics

In our model, the shape of the old and new distribution of the familiarity and recollection processes differ. The recollection process reveals a bi-modal distribution for old patterns but a Gaussian distribution for new patterns (Fig. 6, top row). The bi-modal distribution for old patterns corresponds to two possible classes of attractor states for stored patterns. The left peak corresponds to attractor states that are exactly the studied patterns, so that the patterns presented at test are zero distance from their final attractor state. The second class of attractors, belonging to the right peak, are distorted versions of the studied pattern. At test, patterns associated with this second class of attractors yield a distribution of distances different from zero when settling into their final attractor state. Thus, in our model, the recollection process acts effectively as a threshold process.

By contrast, the familiarity process is based on old and new patterns that have Gaussian distributions of similar width, and is well described as a signal detection process. The difference between familiarity and recollection is more pronounced when

few patterns are studied, but gradually decreases when more and more patterns are studied and tested (Fig. 6).

In empirical investigations, the actual distributions underlying familiarity and recollection discrimination are not directly accessible. However, ROCs can probe the underlying distributions by exploring the relationship between false positive (false alarms) and true positive (hits) rates, as a function of different decision thresholds. The shape and symmetry of ROC curves is indicative of the process underlying a discrimination performance, allowing the contribution of familiarity and recollection to be identified. In agreement with the simulations presented so far, familiarity-based ROC curves reveal high discrimination performance, as familiarity distributions for old and new patterns barely overlap (Fig. 6). By contrast, the recollection process evokes old and new distributions that overlap to a greater extent, leading to poorer discriminability.

Empirically, recognition performance has been reported to predominantly exhibit asymmetric ROC curves. The data are interpreted as reflecting components of two independent processes: a symmetric, curvilinear component associated with familiarity and an asymmetrical, more linear component reflecting recollection. Distinct forms of ROC curves associated with different episodic retrieval processes have also been reported in nonhuman species. For example, a study by Fortin et al. (2004) found asymmetrical and curvilinear components in ROC curves in rats, which were interpreted to reflect recollection and familiarity processes. Moreover, selective damage to the hippocampus caused the ROC curves to become entirely symmetrical and curvilinear, supporting the view that recollection is selectively supported by the hippocampus. Equivalent patterns of ROC curves have also been observed in human neuropsychological studies (Yonelinas et al., 2002; Aggleton et al., 2005; but see Wais et al., 2006).

To investigate the ROC curves in our model, we trained the network with sparse patterns. Although familiarity reveals ROC curves with unit slope (1.06 for 10%, 1.05 for 20% and 1.04 for 30% load), the recollection process exhibits slopes that were different from 1 (1.16 for 10%, 1.22 for 20%, 1.17 for 30% load). The difference between familiarity and recollection ROC curves was statistically significant, as confirmed by *t*-tests based on 100 independent simulations ($P < 0.001$ for all conditions). These results demonstrate that the model leads to differently shaped ROC curves for the familiarity and recollection processes, reflecting qualitatively distinct processes of retrieval.

Recognition of Highly Similar Lures and Overlapping Pattern

The remaining simulations explicitly evaluate the network's performance against empirical findings. Here, we compare performance on tasks known with differentially engaged familiarity and recollection processes, for instance by contrasting recognition memory for studied patterns with highly similar lures (Roediger and McDermott, 1995; Curran, 2000; Nessler et al., 2001). In such tasks, subjects may be presented with a list of singular and plural words (e.g., table, cup) and then tested

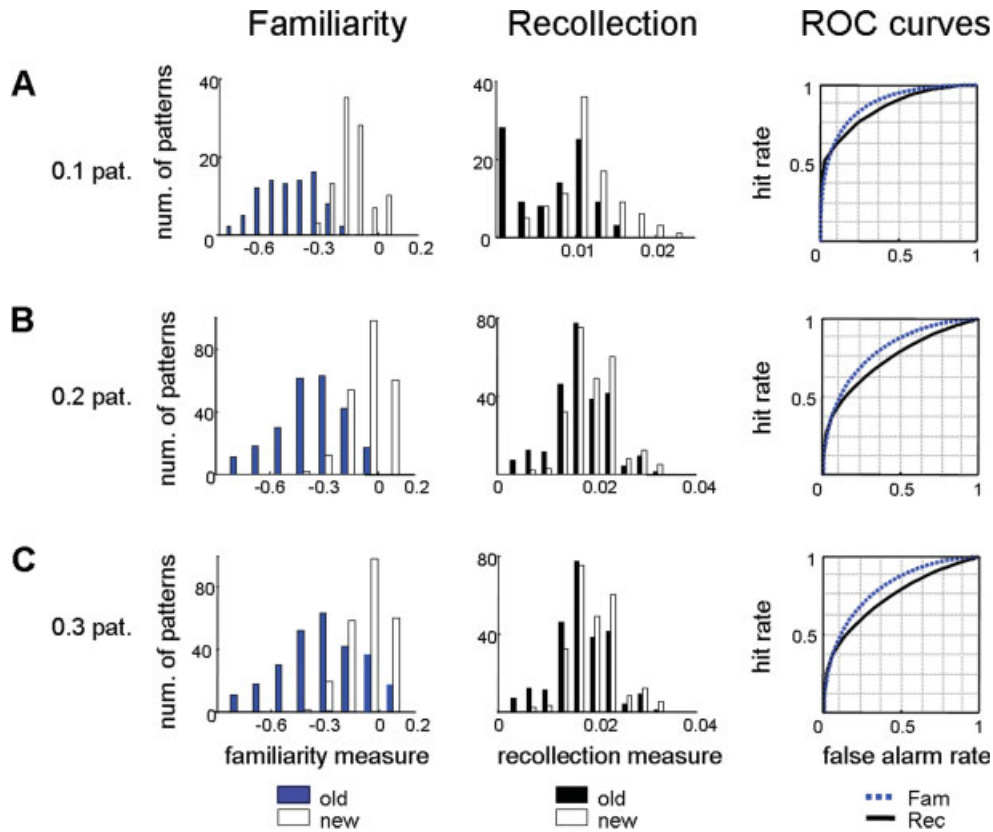


FIGURE 6. The distribution of the familiarity and recollection measures for old and new items, together with their corresponding ROC curves. The ROC curves are simulated with a network of 500 units with sparse representation (1% active units). The different rows (A, B, and C) correspond to the proportion of stored patterns per network size (0.1, 0.2, and 0.3, respectively). The under-

lying distribution of old and new items changes for recollection but not familiarity, this is reflected in changes in the shape of the resulting ROC curves. Overall, the familiarity-based ROC curves (blue dashed line) are more symmetric than the recollection-based ROC curves (black solid line). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

with either studied words (table) or plurality reversed lures (cups) (Curran, 2000). The resulting false-alarm rates are based on the high overlap between studied patterns and lures that yield a feeling of familiarity, whereas correct recognition is thought to be primarily associated with recollection of detailed studied information. Supporting evidence comes from neuropsychological studies of patients with hippocampal damage. Hippocampal patients are considered to suffer from impaired recollection and base their recognition response on familiarity. The findings show that yes/no recognition is impaired in hippocampal patients when targets and lures are highly similar, suggesting that familiarity can not distinguish between these two classes of items. These findings are also supported by the CLS (Norman and O'Reilly, 2003). Overall, the empirical data suggest that successful discrimination between old and highly similar lures is accomplished primarily by recollection relative to familiarity.

In the spirit of the empirical investigations, we tested our model with a mixture of studied patterns and similar lures, assembled from correlated patterns that are similar but not identical to studied patterns. Similarity between old and new

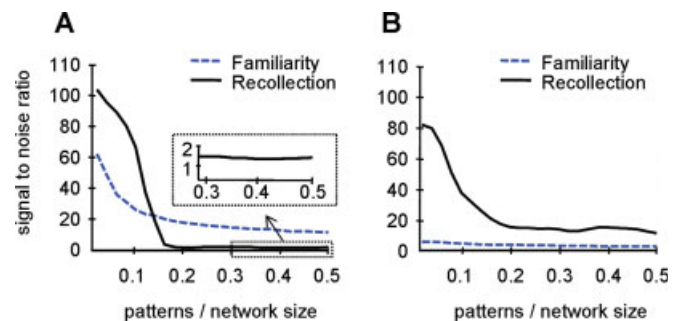


FIGURE 7. The signal-to-noise ratio for familiarity and recollection for overlapping patterns using both nonsparse (A) and sparse (B) patterns. In both cases, the test patterns are overlapping 80% with the study patterns. For nonsparse patterns, the SNR for recollection plateaus just above 1, as shown in the inset to panel A. Familiarity is highly sensitive to the overlap between patterns and declines below the SNR for recollection (in contrast to the results found for nonoverlapping patterns, Fig. 2). (B) For sparse patterns recollection is more robust for overlapping patterns. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

patterns is modeled by constructing a part of the new pattern as identical to (or overlapping with) the activity of an old pattern, whereas the remaining part of the new pattern is randomly generated (see Methods section). It should be noted that such overlapping patterns are correlated patterns, albeit with a particular correlation structure.

For nonsparse patterns, recollection demonstrates comparable SNRs for overlapping and nonoverlapping patterns. By contrast, the SNR for familiarity is reduced for overlapping patterns (compare Fig. 7A with Fig. 2A). Hence, the SNR for familiarity is smaller than the SNR for recollection when tested with overlapping patterns, whereas the opposite pattern of results was found for nonoverlapping patterns (Fig. 2). This reversal in the SNR between nonoverlapping and overlapping patterns only occurs when the magnitude of studied and tested patterns remains below the critical recall capacity (0.15 times the number of network units for nonsparse patterns). For overlapping patterns, the SNR for recollection decreases dramatically and becomes smaller than the SNR for familiarity when the number of patterns exceeds the critical recall capacity. For sparse patterns recollection generally has a superior SNR compared with familiarity.

The reduced performance of the familiarity measure when correlated patterns are introduced was noted by Bogacz et al. (2001). To avoid this problem, they designed a familiarity detector based on long-term depression that also performed well for correlated patterns. Here, because familiarity and recollection are both implemented in the same network, no such change in the familiarity measure can be introduced. Nonetheless, with regards to earlier reported experimental findings, the simulations are consistent with empirical evidence that successful discrimination of overlapping patterns does primarily depend on recollection, whereas familiarity-based discrimination is significantly reduced.

Item and Associative Recognition

A different type of test thought to differentially engage familiarity and recollection is associative recognition, which involves discrimination of specific item configurations. Typically, pairs of items are encoded and memory tested on whether a specific pair of items was previously studied, as opposed to whether the individual items are old or new. Dual-process theories propose that associative recognition should rely primarily on recollection (Donaldson and Rugg, 1998; Yonelinas et al., 1999; Yonelinas, 2002). This is based on the logic that all test items are equally familiar and hence recollection of the item configuration is required to achieve accurate performance. Some dual-process theories assume that familiarity can support associative recognition when individual items are unitized into a single representation (i.e., are bound together and become consciously perceived and remembered as a single entity) (Jäger et al., 2006; Rhodes and Donaldson, 2007, 2008). Recent debates, however, center around a more general contribution of familiarity to associative retrieval. For example, memory for perceptual associations between arbitrary visual items has been

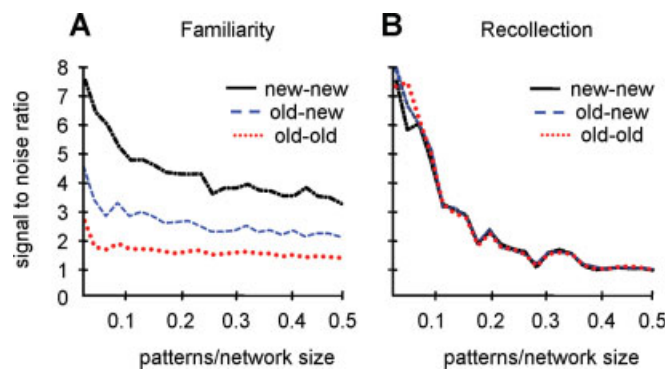


FIGURE 8. The signal-to-noise ratio for mixed patterns for familiarity (A) and recollection (B). This simulation tested whether the network was able to discriminate between previously studied pairs and mixed pairs (Fig. 1B). The mixed pairs were produced by either combining two different learned patterns (old–old: red dotted line), a learned with a new pattern (old–new: blue dashed line) or completely new patterns (new–new: black solid line). When mixed pairs are introduced, familiarity produces a differential response to the three classes of items, whereas recollection is unaffected. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

reported to depend on both familiarity as well as recollection processes (Speer and Curran, 2007). Similarly, Mayes et al. (2007) distinguish within-domain associations which involve very similar kinds of items that are not unitized (e.g., two faces or two building) from between-domains associations which are formed between different kinds of items (e.g., face, building). Familiarity is argued to support within-domain associations but not between-domain associations.

We tested the ability of our familiarity and recollection processes to perform associative recognition. To date, we know very little about how within-domain stimuli might be coded in the brain and we therefore simulate randomly generated patterns which are likely to reflect between-domain stimuli. The network was trained with patterns that represent item pairs; the first and second halves of the patterns reflect the constituent items of that pair (Fig. 1B). The test phase recombines previously studied pairs to produce “mixed” pairs, which were presented along side pairs that were genuinely identical to the studied patterns. There were two types of mixed pairs: one type represents a re-configuration of two items that were previously studied in different pairs (old–old). The second type merges a previously studied item with a new item (old–new). Finally, the network was also tested with a set of inputs representing completely new pairs.

Figure 8 shows the SNR associated with familiarity and recollection during associative recognition (using sparse patterns). The SNR is plotted for each type of mixed pair, relative to genuinely old pairs. For familiarity, the SNR ratio is lowest for recombined old–old pairs, and highest for new–new pairs. This finding indicates that the familiarity process is not good at identifying whether two items were studied as a whole or in separate pairs. By contrast, the recollection process shows no differences in the SNR between the old–old, old–new, or new–

new items (Fig. 8B). Thus, the recollection process is able to discriminate previously studied pairs from those that were studied in a different configuration and new pairs. Again, our model is in agreement with empirical data as our simulations shows that successful associative recognition is predominantly supported by recollection.

DISCUSSION

This article addresses a fundamental question in episodic memory research: how is the dual process distinction between familiarity and recollection implemented? We asked whether familiarity and recollection reflect retrieval from separate memory traces, or a single memory trace that is accessed by separate retrieval processes. To date, the dominant view has been that recollection and familiarity are two distinct processes that produce different outputs by operating on separate memory representations. Here, we provide an alternative view: we presented a single-trace dual-process computational model of episodic memory, demonstrating that two distinct retrieval processes can operate on a single memory representation, yet still generate different retrieval outputs.

Description of the Familiarity and Recollection Processes

In brief, we introduced a memory model based on a Hopfield network. Intuitively, the model performs familiarity discrimination by evaluating whether the activation pattern of the test item agrees with the learned connection strength between the network nodes. If an item was previously encoded, it will be congruent with the network connections and will therefore be judged as familiar. Although this article is primarily concerned with elaborating how successful our model is in simulating familiarity, for completeness we also propose a read out mechanisms by which a familiarity process could be implemented in a biological system (see Appendix). In contrast to familiarity, which is assessed immediately without recovery of stored content, during recollection the network relaxes into an attractor state (network nodes change their activation state until a stable representation is reached) and retrieves stored information. If a previously trained pattern is presented at test, only a few nodes will change, whereas untrained patterns are likely to evoke major changes in activity. The recollection decision is linked to this retrieval dynamic, in that successful recollection occurs only under short relaxation times.

We note that nothing in our model necessitates that familiarity- and recollection-based retrieval must operate on a single network. Information could equally be stored in two separate networks, one dedicated to familiarity and a second devoted to recollection. In this respect, our model performs exactly the same computation as a standard dual-process model. However, by combining the two processes our model not only

leads to additional predictions, it also provides a simpler and more parsimonious account of episodic memory.

Characteristics of the Model Versus Empirical Findings

The model agrees on a number of points with empirical data. Our model suggests, in accordance with the Bogacz network, that familiarity has a high storage capacity, which parallels the empirical assumption that we are familiar with more events than we can actually recollect.

In our model familiarity and recollection operate on different time scales: familiarity of a stimulus is assessed immediately, whereas recollection normally requires additional settling time. This is in accordance with a large number of findings that suggest familiarity typically acts faster than recollection (Atkinson and Juola, 1973; Mandler, 1980; Hintzman and Curran, 1994; Yonelinas and Jacoby, 1994; Benjamin and Craik, 2001). We note, however, that our model allows recollection to be fast under some circumstances, because recollection times are dependent on the strength of the memory trace. Hence, the better learned or more distinct memories are, the more rapidly recollection will occur. Examples of this behavior can be found in a number of memory studies such as Perfect et al. (1996), Wilding (2000), and Vilberg et al. (2006). The fact that discrimination time becomes longer with increasing numbers of study and test patterns (Fig. 4) has also been shown empirically. For instance, Sternberg (1966, 1969) demonstrated that reaction times in list learning experiments increase when the number of studied items increase.

Familiarity and recollection have been further dissociated using manipulations of stimulus material. When we increased the pattern overlap between old and new items, discrimination performance was better for old and overlapping new patterns when retrieval was based on recollection compared with familiarity. Recollection was not, however, always superior to familiarity. When storing many patterns, recollection performance diminished below that of familiarity. Overall, these results suggest that recollection is required for the appropriate discrimination of previously studied from highly similar (correlated or overlapping) but novel patterns. This is in good agreement with behavioral and neuropsychological studies which suggest that targets and highly similar lures can not be discriminated on the basis of familiarity. Nonetheless, the model also clearly demonstrates that when the storage capacity is reached, recollection is unable to support performance—it is unclear whether equivalent capacity limits exist in humans (and decay of the synaptic weights may prevent this limit being reached).

Other experiments show that different memory tasks engage familiarity and recollection differentially. For instance, familiarity can support recognition of single items, whereas recollection is needed for successful associative recognition of previously unrelated (nonunitized) item pairs (Yonelinas, 1997). Simulations of associative recognition reported here show that recollection, in contrast to familiarity, can successfully discriminate previously studied from re-configured items. Thus, our model

incorporates the empirical evidence that, for previously unrecalled (nonunitized) items, recollection but not familiarity contributes predominantly to successful associative recognition.

Further empirical evidence in favor of distinct familiarity and recollection processes comes from reports that the underlying distributions associated with old and new items differ for familiarity and recollection. For instance, empirical studies have reported symmetric ROCs curves in relation to familiarity, whereas asymmetric ROC curves are typical for recollection. These findings are in agreement with the ROC curves and the underlying old/new distributions in our model. The simulations revealed two Gaussian familiarity distributions for old and new patterns, resulting in symmetric ROC curves. By contrast, the recollection measure for old patterns displays a bi-modal distribution, whereas new patterns follow a Gaussian distribution, resulting in asymmetric ROC curves for recollection.

Finally, the functional relationship between familiarity and recollection can in theory be characterized by different models. Although a model that assumes the two processes act separately (independence model) predicts that recollected items can either be familiar or unfamiliar, other models propose that none of the recollected items are familiar (exclusivity model) or all of the recollected items are familiar (redundancy model). The familiarity process implemented in our network is neither conditional on, nor restricted by, successful recollection and is therefore generally compatible with the independence model. This is in agreement with the assumption of most dual-process models and empirical findings in which recollection and familiarity operate independently under different experimental conditions (although there are conditions where the independence assumption may not hold, see Yonelinas and Jacoby, 1995; Kelley and Jacoby, 2000; Yonelinas, 2001, 2002). However, it is important to note that although familiarity and recollection are independent processes in our model, they are correlated in practice (as recollected items are likely to be more familiar, c.f. Fig. 3).

Comparison of the Single-Trace Dual-Process Model With Other Memory Models

Our model is distinct from previously proposed models such as the CLS. The CLS approach portrays familiarity and recollection as two different memory representations with separate encoding and retrieval mechanisms. By contrast, our model emphasizes the fact that the existence of two episodic retrieval processes does not necessitate two distinct memory representations; rather, the two processes could describe distinct ways of accessing information stored in a single memory representation. By demonstrating that our single-trace dual-process model can account for a range of empirical findings, this work highlights the importance of distinguishing between neuronal processes and the neuronal representations on which they operate (as discussed latter).

A second important difference between our model and the CLS model is how familiarity is implemented. The CLS model implements familiarity as a general feature extraction process which detects commonalities across different presentations. Fea-

ture extraction clearly plays an important role in certain types of remembering, most notably in relation to general knowledge or semantic memory (a fact that appears to have provided the original inspiration for the so-called familiarity process in the CLS model, see McClelland and Goddard, 1996). Whether familiarity is best described by feature extraction or an assessment of the energy profile of the network remains to be seen in future research.

Reconciling the Model With Neuroanatomical Data

The proposal that familiarity and recollection describe distinct retrieval processes that operate on the same underlying memory representation raises the question of how a single trace model can be reconciled with neuroanatomical findings. Although there is no doubt that damage to the MTL structure results in memory impairments, the underlying cause of these deficits and the role of distinct MTL structures in different kinds of memory are hotly debated at present. Some researchers argue that the MTL areas subserve distinct mnemonic processes. For example, the seminal work from Aggleton and Brown (1999) proposes two independent anatomical networks supporting memory. The hippocampal-diencephalic system is thought to be critical for recollection, whereas the nonhippocampal MTL regions, including perirhinal cortex and medial dorsal thalamic nuclei, are linked to familiarity. Supporting evidence comes from animal electrophysiological and lesion studies.

Dissociations between recollection- and familiarity-based processes have also been reported in human amnesia but are nonetheless highly controversial. Although some patients show preserved recognition memory after hippocampal damage, others demonstrate impaired recognition across a number of different tests. Evidence from functional Magnetic Resonance Imaging (fMRI) studies are equally contentious. Although some studies suggest that during encoding distinct MTL subregions support familiarity- and recollection-based processes (Davachi et al., 2003) others have failed to replicate those findings (Kirchhoff et al., 2000; Stark and Okado, 2003). Furthermore, to a large extent neuroimaging data has been interpreted on the basis that dual-process theories propose distributed patterns of brain activity should exist for familiarity and recollection. By contrast, our dual-process model implies that this need not be the case and suggests an alternative basis for interpreting these findings.

It is also important to acknowledge that non-dual-process models exist, for example alternative accounts for the division of labor within the MTL are given by the relational memory theory (Eichenbaum et al., 1994; Eichenbaum, 2004) and cognitive map theory (O'Keefe and Nadel, 1978). Some researchers even challenge the view that there is a dedicated memory system in our brain altogether. This is based on findings that the MTL is not specialized for memory but is also involved in perceptual processes (Buckley et al., 2001; Lee et al., 2005). This view suggests that memory actually reflects the operation of a hierarchically organized network of perceptual representations which are distributed throughout the brain, whereby the

MTL receives inputs from anatomically separated streams that process information about objects and scenes (Graham et al., 2008).

Taken together, there is a considerable amount of debate in the literature regarding the neuroanatomical basis of familiarity and recollection. This makes it difficult to test the fit of our model with existing neuroimaging evidence. Nevertheless, a number of authors consider the perirhinal cortex to be critical for item familiarity while the hippocampus is thought to support recollection. In principle these findings do not contradict our model, as long as it is assumed that they reflect activation linked to (contingent upon or downstream from) retrieval processes or retrieval outputs. However, these findings would be difficult to reconcile with our model if it could be demonstrated that the activation directly indexed the site where information is stored. Naively, a model that stores information in a single distributed memory trace is expected to reveal a single specific pattern of activity, regardless of whether familiarity or recollection occurs. However, our model proposes that stored information is accessed differently by familiarity and recollection, leading to distinct retrieval outputs, each of which would be linked to its own distinct neuronal circuit. From this perspective, the neuroanatomical distinction between familiarity and recollection could reflect the downstream consequence of familiarity and recollection having occurred, rather than the activation of separate memory traces per se. This, however, raises the question: where are episodic representations stored in the brain? Although our model makes no explicit predictions in this regard it is compatible with the view that memory representations are nothing more than perceptual (and associative) representations, linked to initial perception, and distributed throughout the brain. Nevertheless, we view this question as an essentially empirical issue; future research is needed to advance our understanding.

At present, empirical findings consistent with the view that distinct brain networks support familiarity and recollection fail to specify whether dissociations originate from distinct memory traces or distinct retrieval operations. We believe that in this regard, the neurally inspired models are underspecified at present. Regardless, we note that if future research does provide clear evidence that episodic memory is stored in two independent traces our model could accommodate such findings. Nonetheless, as constructed, our model parsimoniously demonstrates that a single representation is in theory sufficient to allow multiple retrieval processes to exist. As such, our model contrasts not only with single-process theories, which assume that one retrieval process is sufficient to account for empirical findings, but also with those dual-process models that presume the storage of two distinct memory traces. As mentioned earlier, there is little evidence whether (and if so, how) familiarity- and recollection-based representations differ. We also don't know whether (or how) the acquisition and retention of such representations might diverge. Equally, if separate memory representations are stored for an individual episode and memory representations change over time (e.g., through decay or interference), how do such changes have simultaneous effects on both representations? Proponents of the separate representation view

have not yet attempted to explain how distinct traces could be linked such that they continue to refer to the same event as time passes, without becoming unconnected or differentially changed (i.e., for the representations to drift apart). By comparison, the model presented in this article avoids these problems; it does not require the separate (repeated) storage of identical information.

CONCLUSION

This article introduces a computational dual-process account which models familiarity and recollection with two distinct retrieval processes. We demonstrate that our model mimics performance obtained in empirical studies such as plurality discrimination and associative recognition. More interestingly, however, our model questions, and departs from, several assumptions that have become associated with existing dual-process theories. In our model, familiarity and recollection do not rely on distinct memory representations but operate on the same trace. This makes the model not only parsimonious in principle but also avoids the practical problem that familiarity- and recollection might untie over time if they activate distinct traces. Another deviation is that familiarity and recollection are designed to function independently (and reveal different properties in our model) but the resulting outcomes of the two processes are nonetheless highly correlated with one another. One way of characterizing the relationship is that the familiarity measure initially produced by a probe predicts, at least to some extent, whether the probe will ultimately lead to recollection. Furthermore, our model challenges the view that recollection always has to result in perfect discriminability in an "all-or-none" fashion. The curvilinear ROCs obtained for recollection in our model suggests that recollection should be viewed as a "some-or-none" process, which allows only parts of an event to be recollected. Overall, there is no doubt that future research is needed to test the validity of our single-trace dual-process model; parsimony may not be sufficient justification alone. Regardless, our model challenges some of the central properties ascribed to dual-process theories and highlights the often neglected distinction between neuronal processes and the neuronal representations on which they operate.

Acknowledgments

The authors thank Jim Bednar, Jesus Cortes, David Sterratt, and Alessandro Treves for helpful discussions.

REFERENCES

- Aggleton JP, Brown MW. 1999. Episodic memory, amnesia and the hippocampal-anterior thalamic axis. *Behav Brain Sci* 22:425–498.
- Aggleton JP, Vann SD, Denby C, Dix S, Mayes AR, Roberts N, Yonelinas AP. 2005. Sparing of the familiarity component of recog-

- inition memory in a patient with hippocampal pathology. *Neuropsychologia* 43:1810–1823.
- Atkinson RC, Juola JF. 1973. Factors influencing speed accuracy of word recognition. In: Kornblum S, editor. *Fourth International Symposium on Attention and Performance*. New York: Academic Press. pp 583–612.
- Atkinson RC, Juola JF. 1974. Search and decision processes in recognition memory. In: Krantz DH, Atkinson RC, editors. *Contemporary Developments in Mathematical Psychology: I. Learning, Memory and Thinking*. Oxford, England: W. H. Freeman. pp 243–293.
- Barnes CA, McNaughton BL, Mizumori SJ, Leonard BW, Lin LH. 1990. Comparison of spatial and temporal characteristics of neuronal activity in sequential stages of Hippocampal processing. *Prog Brain Res* 83:287–300.
- Benjamin AS, Craik FIM. 2001. Parallel effects of aging and time pressure on memory for source: Evidence from the spacing effect. *Mem Cognit* 29:691–697.
- Bogacz R, Brown M. 2002. The restricted influence of sparseness of coding on the capacity of familiarity discrimination networks. *Network: Comput Neural Sys* 13:457–485.
- Bogacz R, Brown MW. 2003. Comparison of computational models of familiarity discrimination in the perirhinal cortex. *Hippocampus* 13:494–524.
- Bogacz R, Brown MW, Giraud-Carrier C. 2001. Model of familiarity discrimination in the perirhinal cortex. *J Comput Neurosci* 10:5–23.
- Brown MW, Aggleton JP. 2001. Recognition memory: What are the roles of the perirhinal cortex and hippocampus? *Nat Rev Neurosci* 2:51–61.
- Buckley MJ, Booth MC, Rolls ET, Gaffan D. 2001. Selective perceptual impairments after perirhinal cortex ablation. *J Neurosci* 21:9824–9836.
- Buhmann J, Divko R, Schulten K. 1989. Associative memory with high information content. *Phys Rev A* 39:2689–2692.
- Burgess N, O'Keefe J. 1996. Neural computations underlying the firing of place cells and their role in navigation. *Hippocampus* 6:749–762.
- Clark SE, Gronlund SD. 1996. Global matching models of recognition memory: How the models match the data. *Psychon Bull Rev* 3:37–60.
- Curran T. 2000. Brain potentials of recollection and familiarity. *Mem Cogn* 28:923–938.
- Davachi L, Mitchell JP, Wagner AD. 2003. Multiple routes to memory: Distinct medial temporal lobe processes build item and source memories. *Proc Natl Acad Sci USA* 100:2157–2162.
- Diana RA, Yonelinas AP, Ranganath C. 2007. Imaging recollection and familiarity in the medial temporal lobe: A three-component model. *Trends Cogn Sci* 11:379–386.
- Donaldson DI, Rugg MD. 1998. Recognition memory for new associations: Electrophysiological evidence for the role of recollection. *Neuropsychologia* 36:377–395.
- Eichenbaum H. 2004. Cognitive processes and neural representations that underlie declarative memory. *Neuron* 44:109–120.
- Eichenbaum H, Otto T, Cohen NJ. 1994. Two functional components of the hippocampal memory system. *Behav Brain Sci* 17:449–518.
- Eichenbaum H, Yonelinas AP, Ranganath C. 2007. The medial temporal lobe and recognition memory. *Annu Rev Neurosci* 30:123–152.
- Fortin NJ, Wright SP, Eichenbaum H. 2004. Recollection-like memory retrieval in rats is dependent on the hippocampus. *Nature* 431:188–191.
- Gardiner JM, Java RI. 1990. Recollective experience in word and non-word recognition. *Mem Cognit* 18:23–30.
- Graham KS, Lee ACH, Barense MD. 2008. Impairments in visual discrimination in amnesia: Implications for theories of the role of medial temporal lobe regions in human memory. *Eur J Cogn Psychol* 20:655–696.
- Green DM, Swets JA. 1966. *Signal Detection Theory and Psychophysics*. New York: Wiley.
- Greve A, van Rossum MCW, Donaldson DI. 2007. Investigating the functional interaction between semantic and episodic memory: Convergent behavioral and electrophysiological evidence for the role of familiarity. *Neuroimage* 34:801–814.
- Greve A, Sterratt DC, Donaldson DI, Willshaw DJ, van Rossum MC. 2009. Optimal learning rules for familiarity detection. *Biol Cybern* 100:11–19.
- Gronlund SG, Edwards MB, Ohrt DD. 1997. Comparison of the retrieval of item versus spatial position information. *J Exp Psychol Learn Mem Cogn* 23:1261–1274.
- Hamann SB, Squire LR. 1997. Intact priming for novel perceptual representations in amnesia. *J Cogn Neurosci* 9:699–713.
- Hasselmo ME, Wyble BP. 1997. Free recall and recognition in a network model of the hippocampus: Simulating effects of scopolamine on human memory function. *Behav Brain Res* 89:1–34.
- Hertz J, Krogh A, Palmer RG. 1991. *Introduction to the theory of neural computations*. Redwood City, CA: Addison-Wesley.
- Hintzman DL, Curran T. 1994. Retrieval dynamics of recognition and frequency judgments—Evidence for separate processes of familiarity and recall. *J Mem Lang* 33:1–18.
- Hintzman DL, Caulton DA. 1997. Recognition memory and modality judgements: A comparison of retrieval dynamics. *J Mem Lang* 37:1–23.
- Hockley WE, Consoli A. 1999. Familiarity and recollection in item and associative recognition. *Mem Cogn* 25:1415–1434.
- Hopfield JJ. 1982. Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci USA Biol Sci* 79:2554–2558.
- Jacoby LL. 1991. A process dissociation framework: Separating automatic from intentional uses of memory. *J Mem Lang* 30:513–541.
- Jacoby LL, Dallas M. 1981. On the relationship between autobiographical memory and perceptual learning. *J Exp Psychol: General* 110:306–340.
- Jäger T, Mecklinger A, Kipp KH. 2006. Intra- and inter-item associations doubly dissociate the electrophysiological correlates of familiarity and recollection. *Neuron* 52:535–545.
- Johnston D, Amaral D. 1998. *Hippocampus*. In: Shepherd G, editor. *The Synaptic Organisation of the Brain*, 4th ed. Oxford: Oxford University Press, p 417–458.
- Kelley CM, Jacoby LL. 2000. Recollection and familiarity: Process dissociation. In: Tulving E, Craik FIM, editors. *The Oxford Handbook of Memory*. New York: OUP. pp 215–228.
- Kirchhoff BA, Wagner AD, Maril A, Stern CE. 2000. Prefrontal-temporal circuitry for episodic encoding and subsequent memory. *J Neurosci* 20:6173–6180.
- Lee ACH, Barense MD, Graham KS. 2005. The contribution of the human medial temporal lobe to perception: Bridging the gap between animal and human studies. *Q J Exp Psychol* 58:300–325.
- Mandler G. 1980. Recognizing: The judgement of previous occurrence. *Psychol Rev* 87:252–271.
- Manns JR, Hopkins RO, Reed JM, Kitchener EG, Squire LR. 2003. Recognition memory and the human hippocampus. *Neuron* 37:171–180.
- Marr D. 1971. Simple memory: A theory for archicortex. *Phil Trans R Soc London* 262:23–81.
- Mayes A, Montaldi D, Migo E. 2007. Associative memory and the medial temporal lobes. *Trends Cogn Sci* 11:126–135.
- McClelland JL, McNaughton BL, O'Reilly RC. 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. *Psychol Rev* 102:419–457.
- McNaughton N, Morris RGM. 1987. Chlordiazepoxide, an anxiolytic benzodiazepine, impairs place navigation in rats. *Behav Brain Res* 24:39–46.
- Meeter M, Myers CE, Gluck MA. 2005. Integrating incremental learning and episodic memory models of the hippocampal region. *Psychol Rev* 112:560–585.
- Nessler D, Mecklinger A, Penney TB. 2001. Event related brain potentials and illusory memories: The effects of differential encoding. *Cogn Brain Res* 10:283–301.

- Norman KA, O'Reilly RC. 2003. Modeling hippocampal and neocortical contributions to recognition memory: A complementary learning systems approach. *Psychol Rev* 110:611–646.
- Norman KA, Newman EL, Perotte AJ. 2005. Methods for reducing interference in the complementary learning system model: Oscillating inhibition and autonomous memory rehearsal. *Neural Netw* 18:1212–1228.
- O'Keefe J, Nadel L. 1978. *The Hippocampus as a Cognitive Map*. Oxford, UK: Oxford University Press.
- Perfect TJ, Mayes AR, Downes JJ, Vaneijk R. 1996. Does context discriminate recollection from familiarity in recognition memory? *Quarterly journal of experimental psychology section a-Human Experimental Psychology* 49:797–813.
- Quian Quiroga R, Reddy R, Kreiman G, Koch C, Fried I. 2005. Invariant visual representation by single neurons in the human brain. *Nature* 435:1102–1107.
- Quian Quiroga R, Kreiman G, Koch C, Fried I. 2008. Sparse but not 'Grandmother-cell' coding in the medial temporal lobe. *Trends Cogn Sci* 12:87–91.
- Rhodes SM, Donaldson DI. 2007. Electrophysiological evidence for the influence of unitization on the processes engaged during episodic retrieval: Enhancing familiarity based remembering. *Neuropsychologia* 45:412–424.
- Rhodes SM, Donaldson DI. 2008. Electrophysiological evidence for the effect of interactive imagery on episodic memory: Encouraging familiarity for non-unitized stimuli during associative recognition. *Neuroimage* 39:873–884.
- Robins AV, McCallum SJR. 2004. A robust method for distinguishing between learned and spurious attractors. *Neural Netw* 17:313–326.
- Roediger IHL, McDermott KB. 1995. Creating false memories: Remembering words not presented in lists. *J Exp Psychol Learn Mem Cogn* 21:803–814.
- Rugg MD, Yonelinas AP. 2003. Human recognition memory: A cognitive neuroscience perspective. *Trends Cogn Sci* 7:313–319.
- Sohal V, Hasselmo M. 2000. A model for experience-dependent changes in the responses of inferotemporal neurons. *Network* 11:169–190.
- Speer NK, Curran T. 2007. ERP correlates of familiarity and recollection processes in visual associative recognition. *Brain Res* 1174:97–109.
- Standing L. 1973. Learning 10,000 pictures. *Q J Exp Psychol* 25:207–222.
- Stark CE, Okado Y. 2003. Making memories without trying: Medial temporal lobe activity associated with incidental memory formation during recognition. *J Neurosci* 23:6748–6753.
- Sternberg S. 1966. High-speed scanning in human memory. *Science* 153:652–654.
- Sternberg S. 1969. Discovery of processing stages—Extensions of Donders method. *Acta Psychol* 30:276–315.
- Treves A, Rolls ET. 1994. Computational analysis of the role of the hippocampus in memory. *Hippocampus* 4:374–391.
- Tsodyks MV, Feigelman MV. 1988. The enhanced storage capacity in neuronal networks with low activity level. *Europhys Lett* 6:101–105.
- Tulving E. 1972. Episodic and semantic memory. In: Tulving E, Donaldson W, editors. *Organisation of Memory*. New York: Academic Press, p 381–403.
- Tulving E, Pearlstone Z. 1966. Availability versus accessibility of information in memory for words. *J Verbal Learn Verbal Behav* 5:381–391.
- Vilberg KL, Moosavi RF, Rugg MD. 2006. The relationship between electrophysiological correlates of recollection and amount of information retrieved. *Brain Research* 24:674–684.
- Wais PE, Wixted JT, Hopkins RO, Squire LR. 2006. The Hippocampus supports both the recollection and the familiarity components of recognition memory. *Neuron* 49:459–466.
- Waydo S, Kraskov A, Quian Quiroga R, Fried I, Koch C. 2006. Sparse representation in the human medial temporal lobe. *J Neurosci* 26:10232–10234.
- Wilding EL. 2000. In what way does the parietal ERP old/new effect index recollection? *Int J Psychophysiol* 35:81–87.
- Yakovlev V, Amit DJ, Romani S, Hochstein S. 2008. Universal memory mechanism for familiarity recognition and identification. 28:239–248.
- Yonelinas AP. 1994. Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *J Exp Psychol Learn Mem Cogn* 20:1341–1354.
- Yonelinas AP. 1997. Recognition memory ROCs for item and associative information: The contribution of recollection and familiarity. *Mem Cogn* 25:747–763.
- Yonelinas AP. 1999. Recognition memory ROCs and the dual-process signal-detection model: Comment on Glanzer, Kim, Hilford, and Adams 1999. *J Exp Psychol Learn Mem Cogn* 25:514–521.
- Yonelinas AP. 1999. The contribution of recollection and familiarity to recognition and source-memory judgements: A formal dual-process model and an analysis of receiver operating characteristics. *J Exp Psychol Mem Learn Cogn* 25:513–541.
- Yonelinas AP. 2001. Consciousness, control, and confidence: The 3 Cs of recognition memory. *J Exp Psychol: General* 130:361–379.
- Yonelinas AP. 2002. The nature of recollection and familiarity: A review of 30 years of research. *J Mem Lang* 46:441–517.
- Yonelinas AP, Jacoby LL. 1994. Dissociations of processes in recognition memory: Effects of interference and of response speed. *Can J Exp Psychol* 48:516–535.
- Yonelinas AP, Jacoby LL. 1995. The relation between remembering and knowing as bases for recognition—Effects of size congruency. *J Mem Lang* 34:622–643.
- Yonelinas AP, Kroll NEA, Dobbins IG, Soltani M. 1999. Recognition memory for faces: When familiarity supports associative recognition judgments. *Psychon Bull Rev* 6:654–661.
- Yonelinas AP, Kroll NE, Quamme JR, Lazzara M, Knight RT. 2002. Recollection and familiarity deficits in amnesia: Convergence of remember-know, process dissociation, and receiver operating characteristic data. *Nat Neurosci* 12:323–339.

APPENDIX: READING OUT A FAMILIARITY SIGNAL

In this article, we have used the energy of the network to measure familiarity, but we did not implement a read-out scheme. Instead we followed an “ideal observer” approach, focusing on the fundamental performance limitations of the network. Although in computer simulations the energy is straightforward to calculate, a biological plausible way to read out the energy in our network is more challenging (the network of Bogacz and coworkers does read out energy, but does not do recall). One not very elegant solution would be to create a read-out network of N nodes, each receiving as inputs $h_i = \sum_j w_{ij} s_j$ and s_i but instead of thresholding the input as above, each neuron would compute the product $s_i h_i$. An output neuron could then straightforwardly compute the energy $E = \sum_i s_i h_i$. However, this solution would require the duplication of weights and a multiplication operation. Here, we show how a related familiarity signal can easily be extracted from the network.

A recent modeling study showed that the average activation of a network can be used to determine whether the pattern has been stored in the synaptic connections (Yakovlev et al., 2008).

In that study, the synaptic weights were so weak that the network had no attractor states (i.e., the activity died out when the stimulus was removed). But interestingly even when the weights are strong and the network has attractors, as is the case here, the average activation can still be used to distinguish between old and novel stimuli. This can be done as follows: instead of initializing the network with the test pattern (as we did earlier and as is common in these networks), the test pattern is provided transiently to the network for 100 ms. The activity of each node is given by:

$$s_i(t+1) = \text{sign}\left(\sum_{j=1}^n w_{ij}s_j(t) + x_i^v g(t)\right) \quad (10)$$

Which is as before, except that now the pattern x^v is given as an extra input to each node, gated by the function $g(t)$, which takes either the value 0 or 1. Initially, the input is absent and the network has relaxed to a steady state. At time 0, the pattern x^v is presented ($g(t) = 1$). If the pattern is familiar, the network evolves toward the corresponding attractor state. Consequently, the average activation develops differently depending on whether the pattern is novel or familiar, Figure A1. The duration of the stimulus presentation can be fairly brief. In Figure A1, the time-scale for updating all units once was set to 10 ms, so that in 100 ms each unit is updated 10 times. At the end of the simulation period, we measure the average activation across all nodes, this average activation is used as the familiarity signal and is characterized by its SNR (defined by the square of equation 7 in the main text). In passing, we note that many sensory neurons have transient responses so that such a transient input is quite natural, while it is straightforward to create a neuron that reads the average activity.

Although we find the average activation can work as a familiarity signal, the familiarity discrimination is not quite as good as one based on the energy. In simulations, we find that the SNR of this familiarity signal behaves as $\text{SNR} \sim 0.06 N^2/M$, versus $\text{SNR} \sim 0.5 N^2/M$ for the energy, which means that the

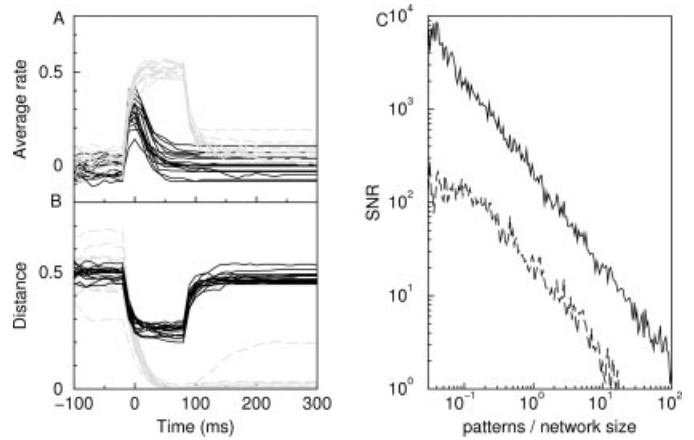


FIGURE A1. **A:** The average activity across the network when novel (solid black) or familiar (dashed gray) stimuli are presented from $t = 0$ until $t = 100$. The average rate at $t = 100$ ms is used as an alternative familiarity signal (Network with 400 nodes storing 60 patterns, responses to 15 familiar and 15 novel stimuli shown). **B:** The distance between the network state and the input pattern for novel (solid black) and familiar stimuli (dashed gray). The network quickly falls into the attractor state if the stimulus is familiar. **C:** Comparison of the signal-to-noise ratio using the energy (solid curve) and the average activity (dashed curve) as a function of the network load. Although the energy has clearly a better SNR, both decrease as $\text{SNR} \sim N^2/M$ and can thus reliably store the familiarity of many stimuli.

capacity is roughly 10-fold less (see Fig. A1). In other words, about three times as many neurons are required to store the same number of familiarity items. However, crucially, the quadratic dependence on N is maintained, so that the capacity is still very large when compared with the recall capacity. This demonstrates that although the energy is perhaps difficult to read out biologically, the familiarity signal is also encoded in easily accessible quantities. Nevertheless, because we are interested in the fundamental performance limitations of the network, in this article, we preferred to use the energy to calculate the capacity.