Universal Memory Mechanism for Familiarity Recognition and Identification

Volodya Yakovlev,¹ Daniel J. Amit,^{2,3†} Sandro Romani,⁴ and Shaul Hochstein¹

¹Life Sciences Institute & Neural Computation Center and ²Racah Institute of Physics, Hebrew University, Jerusalem 91904, Israel, and ³Department of Physics and ⁴Doctoral Program in Neurophysiology, Department of Human Physiology, University of Rome “La Sapienza,” 00185 Rome, Italy

Macaque monkeys were tested on a delayed-match-to-multiple-sample task, with either a limited set of well trained images (in randomized sequence) or with never-before-seen images. They performed much better with novel images. False positives were mostly limited to catch-trial image repetitions from the preceding trial. This result implies extremely effective one-shot learning, resembling Standing's finding that people detect familiarity for 10,000 once-seen pictures (with 80% accuracy) (Standing, 1973). Familiarity memory may differ essentially from identification, which embeds and generates contextual information. When encountering another person, we can say immediately whether his or her face is familiar. However, it may be difficult for us to identify the same person. To accompany the psychophysical findings, we present a generic neural network model reproducing these behaviors, based on the same conservative Hebbian synaptic plasticity that generates delay activity identification memory. Familiarity becomes the first step toward establishing identification. Adding an inter-trial reset mechanism limits false positives for previous-trial images. The model, unlike previous proposals, relates repetition–recognition with enhanced neural activity, as recently observed experimentally in 92% of differential cells in prefrontal cortex, an area directly involved in familiarity recognition. There may be an essential functional difference between enhanced responses to novel versus to familiar images: The maximal signal from temporal cortex is for novel stimuli, facilitating additional sensory processing of newly acquired stimuli. The maximal signal for familiar stimuli arising in prefrontal cortex facilitates the formation of selective delay activity, as well as additional consolidation of the memory of the image in an upstream cortical module.

Key words: familiarity recognition; memory; one-shot learning; neural networks; multiple DMS; psychophysics

Introduction

Familiarity recognition implies only that subjects differentiate between already-perceived and novel items, and do not necessarily include representation of detailed features of the item nor of the context in which it was previously perceived. We will use the term “familiarity” recognition to emphasize its psychological distinction from identification (Mandler, 1980). Identification memory involves associations among the image, its details, and contextual elements. In primates, long, repeated training is required for establishing associations that are reflected in selective “delay activity” with representations including interimage binding (Miyashita, 1988; Sakai and Miyashita, 1991; Nakamura and Kubota, 1995; Asaad et al., 1998; Erickson and Desimone, 1999). Delay activity serves as “working memory” for the brain to work

on, even long after stimulus disappearance (Amit, 1995, 1998). Enabled by long-term synaptic structuring, it is activated by a current stimulus (Amit and Brunel, 1997; Amit and Mongillo, 2003). It depends on context and can engender the representation of context (Miyashita, 1988; Sakai and Miyashita, 1991; Erickson and Desimone, 1999). Following the terminology of Mandler, we couple the terms “recognition” with “familiarity” and “context-dependent memory” with “identification,” and relate the latter with “delay activity.”

Monkey memory processes are studied by delayed-match-to-sample (DMS) and paired associate (PA) tasks. In the first, monkeys report whether a second image is identical with a first; in the PA task, whether the second is the (fixed) associate of the first. These tasks require that a memory trace be maintained during the delay period between the two stimuli. DMS can be performed by recognizing whether the current (test) image was recently presented (cue), even without identification of image details or context. In PA task, the context of the image, its associated image, is crucial.

In DMS studies, with cue followed by several test images, including perhaps two identical ones (ABBA paradigm, repeated nonmatch stimuli), recalling the recency of previous presentation is insufficient for correct performance. Moreover, it is necessary to recognize whether the test image was seen in the same or in a preceding trial. In inferotemporal cortex responses decrease or increase for match compared with nonmatch stimuli (Miller

Received Aug. 17, 2007; revised Oct. 23, 2007; accepted Nov. 20, 2007.

This work was supported by the Change in Your Mind “Center of Excellence” of the Israel Science Foundation of the Israel Academy of Science; the Statistical Mechanics and Complexity “Center of Excellence,” Istituto Nazionale per la Fisica della Materia, Roma; the National Institute for Psychobiology in Israel; and the United States–Israel Binational Science Foundation. We are indebted to Drs. Y. Amit, M. Ahissar, R. Shapley, A. Treisman, and U. Zohary for helpful comments during the course of this study, to A. Miller for useful discussions of familiarity recognition, to T. Orlov for help with the experiments, and to D. Darom for help with images. Our coauthor and friend, Prof. Daniel J. Amit, passed away suddenly on November 2, 2007. With deep sorrow and mourning, we dedicate this paper to his memory.

[†]Deceased, Nov. 2, 2007.

Correspondence should be addressed to Dr. Shaul Hochstein at the above address. E-mail: shaul@vms.huji.ac.il.
DOI:10.1523/JNEUROSCI.4799-07.2008

Copyright © 2008 Society for Neuroscience 0270-6474/08/280239-10\$15.00/0

and Desimone, 1994). Neurons with decreased responses are present even without ABBA training, but they also show weaker responses for repeated nonmatch stimuli in the ABBA task; hence these neurons cannot improve performance in this task. However, enhanced match responses are observed only for match stimuli, not for repeated nonmatch ones. Thus, the prefrontal cortex appears central to the performance of DMS tasks unsolvable by novelty/familiarity judgments (Passingham, 1975; Bauer and Fuster, 1976; Mishkin and Manning, 1978; Miller et al., 1996).

Xiang and Brown (1998, 2004) differentiated these possibilities confronting perirhinal with prefrontal neurons, by presenting well trained and new images in a match-to-sample task. The response of most anterior inferotemporal cortex cells was lower for recently seen images, and higher for most prefrontal cells under these conditions. [Note that our terminology, adapted from Mandler (1980), differs from that of Xiang and Brown (1998, 2004): except when reporting their results, we do not discuss what they call familiarity (well trained vs infrequently seen images). We use the term familiarity for what they call non-novelty (images that were seen)].

Materials and Methods

Behavior

In our extended DMS task (Yakovlev et al., 2005), samples are sequences of images. The trial protocol, its parameters, and schematic trial outcomes are shown in Figure 1. This is a modified, delayed-match-to-multiple-sample task (Yakovlev et al., 2005) (see also Hölscher and Rolls, 2002). In this paradigm, the sample was a sequence of images. The animal viewed a sequence of random length (with two to seven images) and reported whether any (test) image was a repetition of any one of the previous images in the sequence. In one variant, the images of the sequence of every trial were selected from a limited fixed set of 16 images, so that every image had been seen hundreds of times. In a second experiment, sequences were taken from an “infinite” set, so that each image was used only once, except when it appeared a second time in the same trial as a test-repeat, or when planted in “catch” trials [to study false positives (FPs)] (see below). Any image in the sequence of a trial (except the last) can be the one repeated at the end of the trial, and any image (but the first) can be this final match. When any image is repeated (e.g., A, B, C, C or A, B, A) the subject is supposed to respond. Good performance requires a memory trace of all of the images in the sample sequence. The experiment was conducted with sequences of trained images (multicolored drawings of 200×200 pixels) and also, separately, with novel images [multicolored clip art images of patterns and cartoon-like drawings of objects (Art Explosion), each of 200×200 pixels]. Trial sequences varied in length (n), cue position (q), and selection of images. Details beyond the description given in the present paper were presented previously (Yakovlev et al., 2005). Animal experimentation was performed according to National Institutes of Health and Hebrew University guidelines.

Statistical test in the experiment. Statistical tests for performance and FPs was based on the following tables (Tables 1–3) of incidence for each pair n, q . The p value reported is calculated for the total number of hits and misses for trained and novels, and performing the Fisher test for novels higher than trained, using the corresponding confidence intervals. The individual values of p , for each pair n, q separately, using the Fisher test, are given in Table 2.

Modeling

Neurons, synapses, stimuli, and one-shot learning. We consider a network (Fig. 2A) composed of N excitatory neurons. Neuron i expresses, at time t , spike rate $v_i(t)$ ($i = 1, \dots, N$). Neuron j is connected to neuron i by a synapse J_{ij} , which modulates the current afferent to neuron i caused by the spike rate of neuron j . Hence, the total excitatory recurrent current afferent to neuron i at any given time is the following:

Figure 1. Multi-item DMS. **A**, Example of an experimental trial: After appearance of a fixation point, the monkey presses the bar to initiate the trial; in this example, five sample stimuli are presented sequentially, followed by a match to sample 2 and monkey bar release. Definitions: n , Number of samples in trial; q , position of cue from beginning of trial; d , cue–match distance; ($1 \leq n, q, d \leq 6$ or 5 for earlier sessions; data averaged over both); $d = n + 1 - q$; n and q are pseudorandom within the range. **B**, Three possible outcomes that end the trial are as follows: hit, monkey recognizes repetition and releases bar; false positive, monkey incorrectly responds with a bar release to a sample image; miss, no response to a repeated image. No response to a sample stimulus is a correction rejection.

$$j_i = \sum_{j=1}^N J_{ij} v_j. \quad (1)$$

Inhibition is schematized by a single inhibitory neuron feeding hyperpolarizing current to every excitatory neuron, proportional to the total excitatory activity in the network (Wilson and Cowan, 1972; Hopfield, 1984; Salinas and Abbott, 1996), in other words, the following:

$$j_{\text{inh}} = -A_i \sum_{j=1}^N v_j. \quad (2)$$

Presentation of a stimulus (image) increases the spike rate of a subset of the excitatory neurons, by means of a selective external current. The list of cells receiving this additional selective current specifies the stimulus presented, and can be characterized by an N -bit word, with a given number of 1's, corresponding to the cells that respond to this stimulus, the rest of the bits being 0. The total current to cell i in the presence of stimulus number μ is then the following:

$$J_i^{\text{tot}} = \sum_{j=1}^N J_{ij} v_j - A_i \sum_{j=1}^N v_j + A_{\text{stim}} \xi_i^{\mu}, \quad (3)$$

where $\xi_i^{\mu} = 1$ if neuron i is responsive to stimulus μ , and $= 0$, otherwise ($i = 1, \dots, N$). A_{stim} is the amplitude of the current to neuron i attributable to the stimulus (or its contrast). This total afferent current is transduced into a rate via a neural, sigmoid gain function, $v_i^{\text{tot}} = \Phi(J_i^{\text{tot}})$.

Table 1. Performance

	<i>q</i> = 1	<i>q</i> = 2	<i>q</i> = 3	<i>q</i> = 4	<i>q</i> = 5	<i>q</i> = 6
Results with well trained images						
Hits						
<i>n</i> = 1	861					
<i>n</i> = 2	191	437				
<i>n</i> = 3	131	233	288			
<i>n</i> = 4	93	88	182	213		
<i>n</i> = 5	62	79	123	134	125	
<i>n</i> = 6	5	15	24	9	8	32
Misses						
<i>n</i> = 1	75					
<i>n</i> = 2	184	20				
<i>n</i> = 3	101	50	9			
<i>n</i> = 4	112	23	27	4		
<i>n</i> = 5	49	30	24	9	1	
<i>n</i> = 6	10	6	7	2	0	2
Results with novel images						
Hits						
<i>n</i> = 1	1270					
<i>n</i> = 2	506	656				
<i>n</i> = 3	359	402	405			
<i>n</i> = 4	253	288	287	288		
<i>n</i> = 5	177	205	197	227	226	
<i>n</i> = 6	13	23	18	20	22	20
Misses						
<i>n</i> = 1	76					
<i>n</i> = 2	116	23				
<i>n</i> = 3	74	26	8			
<i>n</i> = 4	65	23	15	4		
<i>n</i> = 5	47	19	8	4	2	
<i>n</i> = 6	4	2	2	0	0	0

For each sequence length (*n*) and cue position (*q*), the number of trials that ended in a hit or miss, for trained and for novel images.

Table 2. Value of *p* (Fisher's one-tail test) for each pair *n, q*

	<i>q</i> = 1	<i>q</i> = 2	<i>q</i> = 3	<i>q</i> = 4	<i>q</i> = 5	<i>q</i> = 6
<i>n</i> = 1	0.016					
<i>n</i> = 2	0.000	0.241				
<i>n</i> = 3	0.000	0.000	0.243			
<i>n</i> = 4	0.000	0.000	0.001	0.468		
<i>n</i> = 5	0.000	0.000	0.000	0.022	0.885	
<i>n</i> = 6	0.017	0.074	0.224	0.137	1.000	0.684

Novel performance is higher with *p* < 0.05, except for cases in which *q* = *n* (i.e., last image repeated) and for *n* = 6 because of insufficient data.

Table 3. False-positive and correction-rejection incidence for images previously seen, as a function of number of trials before current trial when seen, and for images never seen before (for novel images) or seen longer ago than five trials (for trained images)

	Seen (no. of trials ago)					Unseen/more than 5 ago
	1	2	3	4	5	
Novel						
FPs	102	24	22	7	5	344
CRs	592	535	458	295	203	18,119
Trained						
FPs	202	67	55	39	31	142
CRs	2065	1898	1622	1396	1221	5095

Data for Figure 4, averages and confidence intervals.

The evolution of the rates in the network is described for simplicity by the set of *N* equations (Wilson and Cowan, 1972; Salinas and Abbott, 1996) as follows:

$$\tau \frac{dv_i}{dt} = -v_i + \Phi(J_i^{\text{tot}}(v_1, \dots, v_N)). \quad (4)$$

Eventually, the network reaches a stationary state, at which $v_i = \Phi(J_i^{\text{tot}}(v_1, \dots, v_N))$

(see below). The gain function tested was as follows:

$$\Phi(j) = \frac{1}{2} \left(1 + \tanh \left(\frac{j - \theta}{w} \right) \right), \quad (5)$$

where θ and w are, respectively, the threshold and width.

As mentioned above, each stimulus is specified by the set of excitatory neurons responsive to it (the set ξ_i^{μ} , $i = 1, \dots, N$). In the simulation, the neurons responsive to a given stimulus are selected at random, with probability *f* (coding level), independently from neuron to neuron and between stimuli [i.e., *fN* neurons are responsive (on average) to a stimulus].

Synapses are assumed to be in one of two efficacy states: potentiated, J_p , or depressed, J_D (Amit and Fusi, 1994; Petersen et al., 1998; Fusi et al., 2000). Learning is driven by the stimuli, and transitions between the two states of each synapse are provoked by the activity of the two neurons connected by the synapse, as elicited by the stimulus. These transitions are stochastic (Tsodyks, 1990; Amit and Fusi, 1994; Brunel et al., 1998; Fusi et al., 2000). On presentation of a stimulus, if a pair of neurons is activated by it, and the synapse between them is in its low state (J_D), it is potentiated ($J_D \rightarrow J_p$) with a given probability q_+ ; if one neuron in the pair is activated by the stimulus and the other is not, and the synapse between them is in its potentiated state, it is depressed ($J_p \rightarrow J_D$) with probability q_- (Fig. 2*B*); in all the remaining cases, the synapse is unchanged. In summary, on presentation of a stimulus, one-shot learning increases the number of potentiated synapses between the cells responsive to this stimulus; reduces the number of potentiated synapses between active neurons and inactive ones, and leaves all other synapses unchanged.

Hence, a one-shot presentation of a stimulus drives the network toward stability with a configuration of activities similar to the distribution elicited by the stimulus, and eventually toward a delay activity correlated with the particular stimulus. Yet, after a single presentation of a stimulus, no stable delay activity is formed, because the potentiation has not gone far enough (Brunel et al., 1998), so that even the most recently learned stimulus does not lead to delay activity (Fig. 2*C,D*). Moreover, because different stimuli share cells in their representation, the presentation of a stimulus causes some damage to the engram of previous ones because of the depressions that it induces in the synapses between the shared neurons and other neurons of the previous stimulus. As the number of stimuli in a sequence increases, the number of synapses supporting stimuli of the earlier representations degrades progressively. The network functions as a palimpsest, with older memories replaced by new ones.

With increasing number of successive stimuli (*p*), the fraction of synapses maintaining a trace of the earliest shown stimulus decreases as follows (Amit and Fusi, 1994):

$$F = K(1 - f^2 q_+ - 2f(1 - f)q_-)^{(p-1)}, \quad (6)$$

where the constant *K* is independent of the number of stimuli presented and *f* is the probability of a neuron responding to a particular stimulus (see above). For low values of *f*, this leads to an estimate of network memory capacity, as follows:

$$p_c \sim 1/(f^2 q_+ + 2f(1 - f)q_-). \quad (7)$$

As the fraction of active neurons per stimulus decreases, the capacity of the synaptic system to preserve a significant trace for the oldest stimulus increases (Willshaw et al., 1969; Tsodyks and Feiglman, 1988; Amit and Fusi, 1994). Equation 7 implies also that the capacity is optimized if, in addition, $q_- = A_{\text{LTD}} f q_+$, where A_{LTD} is a constant (LTD refers to “long-term depression,” the synaptic change to the depressed state). The intuition behind this is that the fraction of modified synapses depends on the number of synapses that are candidates for modification. A fraction f^2 of the depressed synapses see two active cells per stimulus, and the condi-

tional probability per potentiation is $f^2 q_+$. However, a fraction $2f(1 - f)$ of the potentiated synapses see an active–inactive pair of cells per stimulus (a much larger fraction for small f), and the conditional probability per depression is $2f(1 - f)q_-$. [For the model by Willshaw et al. (1969), $q_- = 0$, and on presentation of a very large number of stimuli, all synapses become potentiated, and the system ceases to function.] If A_{LTD} is a constant ~ 1 , q_- becomes smaller than q_+ by a factor f , then the network can maintain a significant synaptic trace (Eq. 6) for as many as $1/f^2$ stimuli (Amit and Fusi, 1994), which for $f = 1\%$ would be 10,000!

The coding level, f , in turn, cannot become arbitrarily small, because the number of active neurons (fN) elicited by a stimulus, must be sufficiently large to allow for a signal that can be detected. A theoretical limit is $fN > \ln(N)$, which is mostly satisfied with $f = 1\%$, even for $N = 5000$.

Simulation method. First, we construct two sets of p random stimuli each, by a random process in which each cell is chosen to be responsive, with average probability f (see exception below). One set of stimuli models the familiar, one-shot learned ones, and the other set models a control group of unfamiliar, never-seen stimuli. Learning is unsupervised and is performed (as described above), one stimulus at a time, on a background synaptic distribution, with potentiated and depressed synapses distributed at random (uncorrelated with the stimuli). Given the resulting (after learning) synaptic matrix, the N equations (4) above are integrated (using the Euler scheme with time step dt), starting from an initial state with a random distribution of rates, presenting one stimulus at a time and waiting until the network reaches a stationary state. In the simulation, we “turn off” learning during the testing stage so that testing order is irrelevant. This corresponds, in Standing’s experiment, to the fact that he only tested a small fraction of the images viewed as samples, so there was little “forgetting” during testing.

Figure 2, C and D, shows sample plots of the evolution of the average rate for a familiar stimulus (red) and for a novel one (blue). The familiar stimulus is the most recently learned. In Figure 2C, rates are averaged over the selective population; in Figure 2D, over the entire network. In all cases, on reaching a stationary state, the stimulus is turned off, and the activity decays (i.e., there is no sustained delay activity). Note the much lower average rates in Figure 2D, because of the very low rates in 98% of the nonselective cells, provoked by the feedback inhibition during stimulation.

Parameters common to all simulations. Parameters common to all simulations were as follows: $A_{stim} = 0.1$; $q_+ = 0.4$; $q_- = A_{LTD} f q_+$; $\tau = 10$ ms; $dt = 0.01$ ms.

Stimulus coding for Standing’s 10,000 stimuli. To simulate the case with 10,000 stimuli (see Fig. 5C), the number of active cells per stimulus was chosen by a Gaussian process with a mean Nf and SD 3% relative to the mean. This SD is much lower than if the stimuli were generated by a Bernoulli process of mean probability f . The motivation is that, for a network of 100,000 cells (the size of a cortical column) with $f = 1\%$, the average number of neurons per image is $Nf = 1000$, and the relative SD is $\sqrt{1000/1000} \approx 0.03$. To avoid a simulation with 100,000 neurons (and 10^{10} synapses), we used 7000 neurons and imposed a Gaussian distribution with the relative SD expected in the larger population (i.e., 3%).

Receiver operating characteristics for learning before and after reset. One way of tracing the network performance is to plot the number of hits versus number of FPs [receiver operating characteristics (ROC) curve].

Figure 2. The network and neuronal dynamics. **A**, Schematic view of the network. Solid circles, Excitatory cells; internal lines, recurrent excitatory synapses; dashed circle, inhibitory neuron; arrows, external afferents feeding a selective subset of cells. **B**, Schematics of learning process. Pairs of cells (circles) and the synapse connecting them. Full/open circle, Responsive/nonresponsive cell; J_D/J_P , synaptic efficacy; to each combination of a pair of activities corresponds a probability for a transition; other activity combinations and initial synaptic states leave the synapse unchanged. **C**, **D**, Evolution of the average neural activity: averaged over population of cells selective to the current stimulus (**C**) and averaged over entire network (**D**). Red, For a familiar image most recently (one-shot) learned; blue, for a novel (never-seen) stimulus. The lower scale in **D** is attributable to low activity of 98% of cells, caused by inhibition (rates: relative to maximum rate). When the stimulus is removed, the rate decays (i.e., no delay activity). Parameters: $N = 2000$, $p = 200$, $f = 2\%$, $q_- = 0.004$ ($A_{LTD} = 0.5$), $A_i = 13$, $J_D = 6$, $J_P = 20$, $\theta = 0.11$, $w = 0.07$ (and Parameters in Materials and Methods). Note in **C** the selective signal even for unfamiliar stimulus during stimulation expressed by the higher average rate for selective cells for other cells.

Hits and FPs are generated using a variable discrimination threshold for discriminating familiar and unfamiliar stimuli, starting from a low threshold at which hits are very probable, but so are FPs. As the threshold increases, FPs are eliminated, but miss errors multiply and performance decreases. See, for example, Figure 7E, in which it is assumed that either 100% of the trials contain only unfamiliar stimuli or 100% contain images seen in the preceding trial. Hits and FPs in this case are computed as follows: (1) trials with unfamiliar stimuli: hits, P_h = (fraction of once-seen images with rate above threshold, from Fig. 7A, red curve); FPs from images never seen, P_f = (fraction of unseen images with rate above threshold, from Fig. 7A, blue curve); (2) catch trials without reset: hits, P_h = (fraction of once-seen images in absence of reset with rate above threshold, from Fig. 7D, red curve); FPs from catch images, P_f = (fraction of images with rate above threshold, from Fig. 7A, red curve); (3) catch trials with reset: hits, P_h = (fraction of once-seen images with rate above threshold, from Fig. 7D, red curve); FPs from catch images, P_f = (fraction of images after a reset with rate above threshold, from Fig. 7B, red curve).

The ROC curve (Fig. 7E) expresses the probability of hits (P_h) versus the probability of FPs (P_f). [For a description of the ROC curve, see Green and Swets (1966) and MacMillan and Creelman (1991).]

Results

Multi-item memory (behavior)

The first goal of the current study is to derive behavioral characteristics of familiarity detection, in terms of the accuracy level of hits versus misses when a familiar image is presented, as well as false positives vs correct rejections (CRs) when a novel image is presented.

Figure 3. Monkey performance level for well trained and for novel images: performance (hit rate) in the multitier DMS task versus cue position (q), for different trial lengths (n). **A**, Results for task performance with sets of 16 well trained images; performance decreases with n for fixed q and increases with q (i.e., decreasing d) for fixed n . **B**, Novel (12,000) images; performance is significantly higher than with trained images (**A**). **C**, Superposition of data for trained (black; **A**) and novel (red; **B**) images, separated by trial length. The bars here and in Figure 4 represent 95% confidence intervals for proportions (DeGroot, 1986) (these are larger for the fewer trials of $n = 6$).

Fixed image sets

We trained two monkeys with fixed sets of 16 images: The monkeys saw each image hundreds of times, in different sequences. Figure 3A shows their average task performance [i.e., hits/(hits + misses)] versus cue position (q), for each trial length (n). For constant n , performance improves as the cue is presented later (i.e., closer to the match image) (decreasing distance d between cue and match). We accounted for this result assuming that working memory depends on delay activity (Amit et al., 2003). This delay activity survives intervening sequence stimuli (Miller et al., 1996), which induce their own delay activity. Behavioral errors are attributed to spontaneous jumps of an image into or out of working memory, corresponding to spontaneous (noise driven) initiation or termination of delay activity (leading to FPs and misses, respectively); the probability for such jumps increases with time or number of presented stimuli.

Novel images

The same monkeys were tested on the same task using novel images: No image was seen twice, except when it was a match stimulus, or when included as a catch stimulus in a later trial (see below, False positives). Figure 3B shows the surprising result: Contrary to expectation, performance is significantly better for the never-before-seen (novel) images than for the well trained ones (Charles et al., 2004), nearly perfect (>90% correct; except for the first cue position, where presumably monkeys relate differently to the image because it cannot be itself a match). This finding calls for an explanatory framework: As shown in Figure 3C, performance is significantly better than with trained images ($p < 0.005$, paired t test between the entire set of performance values for novel and trained images). This high performance for novel images, in which no delay activity is expected, requires a different strategy.

False positives

We included a small number of catch trials to measure the effect of repeated stimuli in generating false-positive responses. This was done in the following way: In generating the sample sequences, we inserted catch trials as follows: after producing a long set of trial sequences in which each image appears only once, we replace in 30% of the trials one (single) image with an image already seen in one of the five preceding trial sequences. Figure 4 shows average FP rate as a function of backward distance (1–5) to the trial that contained the image provoking the FP, for the trained set (Fig. 4A) and for novel images (Fig. 4B), compared with the rate for further back (Fig. 4A) or never-seen images (Fig. 4B). FP rates for images of the one-back trial are somewhat higher for novel than for trained images. Given the large capacity of familiarity memory [implied by Standing (1973)], in absence of some intertrial reset mechanism, one would expect a FP rate for novel images seen in any previous trial as high as the hit rate in the same trial (i.e., ~100%). The much lower FP rate suggests that some intertrial reset mechanism must be implemented in this circumstance.

Network model (theory)

As mentioned above, for trained images the accepted physiological correlate and explanatory paradigm has been reverberating, selective neural delay activity [working memory (Miyashita, 1988; Miyashita and Chang, 1988; Sakai and Miyashita, 1991; Amit and Brunel, 1997; Camperi and Wang, 1998)]. However, this option is not available for novel images. Tens to hundreds of presentations per image are required to form selective delay activity (Miyashita, 1993; Erickson and Desimone, 1999).

We suggest that prefrontal enhancement [as found by Xiang and Brown (2004)] signals presence of the match, similar to the suggestion that inferotemporal enhancement underlies successful performance of the basic DMS task with novel images (Miller and Desimone, 1994), in the absence of delay activity (Li et al., 1993).

One issue naturally arises when finding similar effects in two cortical regions (even when they are in opposite directions): Are these independent mechanisms or does one drive the other? On the basis of learning dynamics, it has been suggested that prefrontal feedback signals drive inferotemporal plasticity (Xiang and Brown, 2004).

A number of models have been suggested to account for the decreased response for repeated stimuli [whether familiar, recent, or non-novel (Brown and Xiang, 1998; Bogacz et al., 2001; Bogacz and Brown, 2003; Norman and O'Reilly, 2003; Meeter et

al., 2005)] and some of them also account for the enormous memory capacity observed by Standing (1973) (Bogacz et al., 2001; Bogacz and Brown, 2003). Here, we propose a particularly simple neural system, endowed with conservative Hebbian synaptic plasticity, capable of accounting for the monkey psychophysics experiments we present and for the related findings of Standing. The computational output of the network is a directly observable familiarity signal and the model neurons exhibit characteristic properties typical of neural responses in prefrontal cortex. Moreover, the very same synaptic plasticity underlies both familiarity recognition and identification: two apparently quite different memory phenomena.

Identification memory expressed by delay activity implies a set of strengthened synapses among neurons representing an image. These strengthened synapses are a passive, long-term memory of past experience, which allows future recall. The synapses of a given cortical module contain the passive memory of all images that have been learned, and as such this memory is nonselective. If the synapses have been sufficiently potentiated by past presentations of an image, then, after removal of the image, reverberating delay activity (working memory), selective to the image presented, can be sustained. This is an active reminder that the image was perceived recently. Delay activity is also critical for representing the context of an image by allowing interaction between the active memory of one image and activity induced by presentation of an immediately following (associated) image, referred to as prospective activity (Sakai and Miyashita, 1991; Amit et al., 1994; Mongillo et al., 2003).

The proposed familiarity mechanism still depends on changes of synaptic efficacy. But, after one-shot presentation, the degree of synaptic potentiation cannot sustain selective delay activity (Miyashita, 1993; Erickson and Desimone, 1999), although a trace of passive memory is in place (Xiang and Brown, 2004). Hence on second presentation of an image, the population response will be higher (on average) than for an image never seen. This higher average response acts as a signal for “having been seen before” as opposed to “never seen.”

This signal is nonselective but would generate a proper behavioral response: The rate evoked by the test (repeat) image can be compared with a threshold (in a one-way decision, as in the behavioral results, above), or a comparison can be made between two rates in a two-alternative forced-choice test (Standing, 1973). There is a selective component underlying the response to a familiar image, but it is attributable to the stimulus presented, and would be there also on first presentation of an image. Familiarity recognition may be an effective first incremental step to identification (working memory), as expressed by delay activity. This first step suffices to reproduce one-shot encoding of a familiarity memory trace of an enormous number of images, without generating delay activity.

Familiarity recognition signal and readout

How can the traces preserved in the synapses distinguish first from second presentations of a stimulus [as found in humans by Standing (1973) and by ourselves, above, in monkeys], without delay activity? Figure 5A (red curve) presents stationary state spike rates during stimulation, averaged over the active neurons responsive to each stimulus, after learning 200 stimuli, as a function of stimulus number; no. 1 is the oldest, first-learned stimulus (somewhat depressed by subsequent learning), and no. 200, the most recently learned. The blue curve shows rates for unfamiliar stimuli. Fluctuations are attributable to variability of the actual number of active neurons/stimulus and stochasticity in the plas-

Figure 4. FP rates [FP/(FP + CR)] for images presented within one of the preceding five trials versus backward distance to that trial. **A**, Test with a limited set of well trained images (as in Fig. 4A) showing FP rates for images seen versus not seen in the last five trials. **B**, Test with novel, never-before-seen images (as in Fig. 3B), showing FP rates for catch images seen in the last five trials versus never-seen images. The FP rate for an image seen in the immediately preceding trial is higher for novel than for trained images ($p < 10^{-4}$), and both rates are significantly greater than for a FP with images seen in earlier trials ($p < 0.001$, Fisher's test). The FP rates of images from earlier trials are about the same for further back trials and are about the same as for unseen (blue bar). All rates are much lower than same-trial hit performance, the FP rate that would have been expected if memory were fully maintained across trials.

ticity dynamics. Every red (familiar image) rate is higher than any blue (novel image) rate. Were it possible for the brain to probe exclusively selective cells for each stimulus, this would be a clear familiarity signal.

However, readout must be based on sampling (nonselectively) the entire network of excitatory cells (i.e., independently of the current stimulus). Figure 5B shows this signal (i.e., rates averaged over all network cells) using the same simulation. Along with the 2% (~40) selective cells in Figure 5A, the data in Figure 5B include in their average the 98% (~1960) nonresponsive cells. Nevertheless, the red–blue curve difference continues to form a clear familiarity signal. The greater fluctuations in Figure 5B cause more performance errors (misses and/or FPs). The rate distributions over the stimuli (Fig. 5A,B, right) show greater overlap in Figure 5B corresponding to lower signal detection theory d' (Green and Swets, 1966). The near-quantitative match to

Figure 5. Network familiarity signal for up to 10,000 pictures. **A**, **B**, Average rate signal versus presentation order (**A** for selective neurons; **B** for entire network) after one-shot learning of 200 stimuli. The red/blue curves represent rates for familiar/unfamiliar stimuli. Separation of rates is much larger for selective cells. The horizontal full and dashed lines are averages and SDs over stimuli for each curve. The amount of overlap between the two is an indication of expected performance. (Parameters are as in Fig. 2.) **C**, Average overall rates for familiar (red) and unfamiliar (blue) stimuli ($N = 7000$; $p = 10,000$; $f = 1\%$; $q_- = 0.002$ ($A_{LTD} = 0.5$); other parameters are as in Fig. 2) (i.e., stimuli coded on average by 70 cells, with actual number having a Gaussian distribution of relative width 3%) (see Materials and Methods). Rates were sampled every 50 images. The format is as in **A** and **B**. Rate distributions across stimuli for the two curves are shown to the right of each panel, superposed by a Gaussian with mean and SD from the data. The distance between these curves corresponds to the detectability, d' .

behavioral performance (see above), including a small but significant rate of misses and false positives, supports the chosen model parameters including the use of a global readout signal.

Standing's 10,000

Figure 5C exhibits the dramatic result that a network of reasonable size, with quasirealistic neural dynamics, is capable of discriminating between 10,000 familiar (one-shot learned) and unfamiliar images, even by nonselective sampling. The decay in the amplitude of the signal going to older stimuli is attributable to the system's approaching its capacity limit.

Standing's two-alternative forced choice test, in which subjects chose between two concurrent images, one familiar and one unfamiliar, is modeled by comparing average rates invoked by a familiar and an unfamiliar stimulus. When the former is lower than the latter, the subject makes the wrong choice, producing both miss and FP errors. For the simulation of Figure 5C, performance would be 98% correct, sufficiently close to Standing's result. In our multiple-DMS paradigm, a single image serves as a test, so rate comparison is impossible. During training, a rate threshold must be established to separate familiar from unfamiliar images. This leads to the fourfold variety of responses: hit/miss, familiar match image rate above/below threshold; FP/CR, unfamiliar sample stimulus rate above/below threshold.

Palimpsest effect, learning 1000 + 1000

The model system is a palimpsest [i.e., it memorizes ("recognizes") the latest images and gradually "forgets" earlier ones (Nadal et al., 1986)]. We simulated learning 1000 stimuli in a network with an expected capacity of ~ 1000 (Fig. 6A). Average rates for the first 1000 familiar images (orange curve) dips into the unfamiliar stimulus (blue) curve, as one moves to the oldest stimulus (no. 1), so discrimination becomes difficult at an age of 1000 (rate-over-stimulus distributions, familiar and novel) (Fig. 6B). Another 1000 stimuli are then presented on top of the first 1000. The (red) signal for the first 1000 drowns completely in the sea of unfamiliar signals (rate distribution in Fig. 6C), but the familiarity discriminant of images 1000–2000 is as effectively expressed as were the first 1000 (Fig. 6D).

The palimpsest property is attributable to the same mechanism that limits the memory capacity, namely that every new image erases (by LTD induced by neurons common to the representations of different images) part of the engram embedded by the previously seen ones, and thus liberating synapses for new images to be memorized. To exhibit the phenomenon, we chose $f = 0.02$ so that capacity is not too high (~ 1000 images).

Intertrial memory reset

The high level of performance found for novel images in the multiple-DMS paradigm (Fig. 3B) suggests that a Standing-like phenomenon may be involved. However, familiarity memory would be too good! It raises an excess memory problem: An image presented in trial k would produce a FP

when shown in trial $k + 1$ or later, because it has been seen recently. Given the expected enormous size of familiarity memory, all catch images, regardless of distance back, should produce FPs. Instead, the rate of FPs was relatively low (Fig. 4A, B), $\sim 15\%$ for one back and 2–3% for images of earlier trials. The fact that the FP rate for repeated images is relatively low indicates that monkeys activate a memory reset mechanism between trials; they effectively confine their familiarity search to the limited sequence contained within each trial (Miller et al., 1993; Hölscher and Rolls, 2002).

The intertrial reset mechanism may be learned in reaction to error signals produced by FPs. It is modeled by activating, in the intertrial interval, a large fraction of the cells in the module, with particular plasticity probabilities. To exemplify the consequences, we use a network expressing familiarity for 50 stimuli (Figure 7A). We choose the optimal discrimination threshold from the intersection of Gaussian fits. The expected rate of FPs, for novel images seen in the immediately preceding trial, in absence of a reset, is equal to the hit rate, given by the fraction of red rates above the threshold (92%).

Figure 7B presents the effect of the reset mechanism. The percentage of FPs is the fraction of "red" rates above threshold (as in Fig. 7A). The reset lowers the overall rate in the network and

hence a large number of red rates go below threshold. The percentage of FPs induced in this case is 32%. In the experiment, there were 15% FPs for catch images in the preceding trial. After a second reset (Fig. 7C), FP rate decreases to 8% (experiment, ~2%). After the second reset, the distributions of rates for familiar and unfamiliar stimuli become indistinguishable for additional resets, in accordance with the experimental result (Fig. 4). Figure 7D shows that, after a double reset, the network is able to learn new stimuli as well (as required by the behavioral results). Figure 7E compares network performance ROC curves in the absence of a reset mechanism, when there are no catch trials (i.e., no images repeated from previous trials) or with 100% catch trials, and in the presence of a reset mechanism, again with 100% catch images (see Materials and Methods). Note the high performance level without catch trials, very poor discrimination between familiar and novel images for catch trials without a reset mechanism. The reset restores performance to quite near the level without catch trials (see Materials and Methods).

Discussion

Testing monkeys on a delay-match-to-sample task involving memory of multiple items, we find performance divides into two modes: Using well trained images, capable of inducing working memory delay activity, performance is good, with few false positives, occurring mainly for images presented in the preceding trial. The hit rate in this mode is a function of the length of the trial and the distance (in terms of number of intervening images) from cue to match. When using novel images that were never before seen by the monkey, performance was surprisingly even better. Here, delay activity could not have been initiated by image presentation, because it is well known that an experience of hundreds of presentations of an image is required to develop appropriate (attractor) delay activity. Performance hit rate in this mode is less dependent on cue–match distance, suggesting a long-term passive memory. At the same time, the false positive rate, even for “catch” images presented in the preceding trial, is lower than in the trained images mode. This result suggests that a between-trial reset mechanism is activated in this mode.

The suggested model accounts for all of these behavioral performance characteristics within a rather generic scheme of plasticity and of neural dynamics. Interestingly, the two modes are interpreted by two different mechanisms, interstimulus delay activity and enhanced stimulus response, but both mechanisms are captured by different stages of the same Hebbian synaptic learning rules. The near-quantitative accord between model and behavioral characteristics suggests that additional work in this direction, comparing additional experiments with quantitative model analysis, is worthwhile.

It has been widely believed that familiarity recognition relies on the hippocampus and perirhinal cortices (Brown and Xiang, 1998; Brown and Aggleton, 2001), where, as mentioned, cells responsive to visual stimuli respond at a lower rate on second and additional presentations of the same image (Brown et al., 1987; Fahy et al., 1993; Li et al., 1993; Miller et al., 1993; Sobotka and Ringo, 1993; Brown and Xiang, 1998; Xiang and Brown, 1998; Bogacz et al., 2001). It is im-

Figure 6. Palimpsest memory. **A**, Familiarity signal for 1000 familiar stimuli (orange relative to blue). Same stimuli are erased after another 1000 are learned (red overlaps blue). Clear discrimination of new 1000 stimuli, learned on top of the first 1000 (red to blue, 1000–2000). Network has capacity of 1000. ($N = 3000$; $f = 2\%$; $q_- = 0.04$ ($A_{LD} = 5$); $J_D = 10$; $J_P = 25$; other parameters are as in Fig. 2). **B–D**, The three panels are the rate distributions (introduced in Fig. 5) in the three cases: stimuli 0–1000 (orange–blue), 0–1000 (red–blue), and 1000–2000 (orange–blue). Note that, although the nonselective rate distribution is well described by a Gaussian, the familiar ones have a systematic tendency to lower rate, as a result of the decay of selective image rate with “age.” In a larger network, and/or lower coding level, the decay would have been less pronounced.

portant to note that, in every one of these cases, the response change is for a particular image, and not for the general category. That is, the response of each of these neurons only signals the relative familiarity, recency, or novelty of a particular image, and without knowing which image was presented, one cannot tell whether the image seen was or was not in the appropriate category.

Although it may seem unimportant whether changes are signaled by increment or decrement, it is significant that these reflect whether images are highlighted when they are relatively familiar or unfamiliar, relatively novel or non-novel, seen relatively recently or longer ago. Xiang and Brown (2004) note that the functional difference may be considerable. The direction of change is such that the maximal signal from temporal cortex is for novel stimuli, facilitating additional sensory processing of such stimuli. The maximal signal for familiar stimuli arising in prefrontal cortex facilitates readout, as well as additional potentiation of the synapses selective for the particular stimulus. This would lead, eventually, to the formation of selective delay activity. Familiarity is inherently a first step toward full identification: One of the first things we know about an image is per force that we have seen it before. It is hard to conceive of a mechanism that includes a decreased response as a step on the way to establishing a poststimulus “delay activity.”

Moreover, the higher rates for familiar stimuli have another important potential role. (We are grateful for this suggestion to Dr. Yali Amit, University of Chicago, IL.) A constant flow of novel stimuli is a source of noise in the system and gradually effaces previous memory traces. Hence, the higher rates associated with familiar stimuli may serve as a filter, to accumulate evidence for the significance of a stimulus. These higher rates may lead to sufficient selective activity in a second module, in which structuring can begin, exclusively for familiar stimuli, thus eliminating noise, reinforcing encoding, and suggesting a mechanism of consolidation (McGaugh, 2000). We might add that, on methodological grounds, it is more likely that

Figure 7. Intertrial memory reset. **A–C**, FPs for catch trials. **A**, Rates for familiar (red) and unfamiliar (blue) one-shot memorized stimuli ($N = 2000$; $p = 50$; $f = 2\%$; $q_- = 0.04$ ($A_{LTD} = 5$); $\theta = 0.13$; $w = 0.05$; $J_D = 11$; $J_P = 22$; $A_I = 12$). Forty-six of 50 stimuli produce FPs. Rate histograms and Gaussian fit are as in Figure 2. Dashed line, Optimal discrimination threshold (intersection of the familiar and unfamiliar Gaussian fits) unaffected by resets. **B**, First reset. Familiar and unfamiliar stimuli tested after an intertrial reset of **A** (50% of cells stimulated; $q_{+,reset} = 0.12$; $q_{-,reset} = 0.95$). FP rate, 32%. Rate histogram shown only for familiar stimuli. **C**, Second reset for same stimuli: FP rate, 8%. **D**, Learning 50 new stimuli, after the second reset in **C**. Familiar rates (red), produced by the newly learned stimuli, are as before reset (**A**). **E**, ROC curve (see Materials and Methods) obtained from **B** (light gray curve) shows the good performance after a first reset for the hypothetical case of 100% catch trials incidence. For comparison, the ROC curve in absence of catch trials (black) shows a better d' , whereas the one with 100% repetition incidence of image from previous trial, but without a reset mechanism, shows chance performance.

enhanced activity underlies a cognitively significant signal than decreased activity. There may be too many novel stimuli in the world, as well as too ubiquitous spontaneous activity, which would be best ignored.

The model described is quite successful in providing a unified framework for familiarity detection as found experimentally in very long sequences in humans (Standing, 1973) and in variable length short trials in monkeys (current study). It also provides a strong indication that the same Hebbian mechanism may underlie both the creation of familiarity detection as well as the more exacting process of generating selective delay activity [the neural correlate of short-term memory (Miyashita and Chang, 1988)], essential for generating context dependence in neural representations. A future choice will have to be made between the model presented here and other models (Bogacz and Brown, 2003). Recent neurophysiology (Xiang and Brown, 2004; Charles et al., 2004) tends to favor the present model. We argue that the model proposed here subtends a different functionality and is necessary on methodological grounds. The dialectic between the psychophysics experiments and modeling is also very rich in proposing new experiments and opening novel queries for the theory, as follows.

Experiment

Standing's forced-choice test could and should be introduced in the monkey experiment, both with novel and well trained images. In this way, one could probe the dividing line between the perceptive mechanism for novel, never-before-seen images, and that for well trained images. Monkeys should be trained first with novel images, without catch trials, which should be introduced later. The relevant questions would be the following: (1) Will reset be acquired when testing in presence of infrequent catch trials? (2) Will

images in catch-trials initially produce false positives? (3) What if one mixes well trained images among the novel ones? Will these act as catch trials *ab initio*? (4) Would turning over from novel images to a fixed small set produce the result of training *ab initio* with a fixed limited set? (Preliminary data indicates a negative response.) (5) Would there be initial good performance (the images are novel), followed by a period of confusion, before passing to a (delay activity) strategy for known images?

Physiology (single-cell recording) is a tool for testing the proposed reset mechanisms as well as an eventual decision mechanism (not modeled). The model requires a high fraction of highly active cells in the relevant cortical area during the intertrial interval. There are some preliminary indications that this is the case; more work is required to quantify the effect and to infer the synaptic transition parameters accompanying it.

Modeling

We have not attempted here a detailed quantitative comparison of the network dynamics with the psychophysics, although the level of agreement at this preliminary stage is quite promising for future developments. The main reasons are that the model is not at the state of the art regarding the physiology and

anatomy of the elements and the network. Moreover, to go further requires closer attention to the details of the experimental procedure, in particular the distribution of trial lengths, and the seeding of the catch trials.

Even before rendering the network model more biologically plausible (see below), the present model should be investigated to explore the constraints imposed by the requirement that repeated presentations of the stimuli (with the accompanying plasticity) lead to stable delay activity steady states. Although this is to be expected, it is far from trivial. The appropriate structure of the synaptic matrix is a necessary condition for sustaining selective delay activity, but it is not sufficient. Other parameters must be adjusted, such as the neural parameters, the inhibition mechanism, the amplitude of the synaptic efficacies, etc. Once this is achieved, one could estimate the number of presentations required to reach stable delay activity at the present level of modeling. Moreover, it will allow comparison of the capacity for familiarity detection with that of identification.

The model allows for testing the effect of catch trials on performance and on FPs, by allowing in the simulations a certain fraction of the images to be seen twice (30%) in the experiment reported. This may also account for the discrepancy in the numerical effect of the reset mechanism between the experiment and the simulation.

The present model should be tested for realistic, random connectivity levels between cells, to test the relationship between familiarity memory capacity and the network size. One would expect (Sompolsky, 1986; Derrida et al., 1987) the constraint $f > \ln N/N$ to be replaced by $\ln(cN)/(cN)$, where c is the fraction of neurons synapsing on a given neuron (anatomically $\sim 10\%$). Hence, to allow for $f \sim 1\%$, one would need $cN \sim 9000$, or 90,000 cells, a rather realistic number.

To explore more effectively the space of constitutive parameters, simulations of networks with a high number of neurons should be replaced by a new mean-field approach. Eventually, the properties of the model should be put to test implementing spiking neurons and spike-driven synaptic plasticity (Mongillo et al., 2005).

References

- Amit DJ (1995) The Hebbian paradigm reintegrated: local reverberations as internal representations. *Behav Brain Sci* 18:617–657.
- Amit DJ (1998) Simulation in neurobiology—theory or experiment? *Trends Neurosci* 21:231–237.
- Amit DJ, Brunel N (1997) Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cereb Cortex* 7:237–252.
- Amit DJ, Fusi S (1994) Learning in neural networks with material synapses. *Neural Comput* 6:957–982.
- Amit DJ, Mongillo G (2003) Spike-driven synaptic dynamics generating working memory states. *Neural Comput* 15:565–596.
- Amit DJ, Brunel N, Tsodyks MV (1994) Correlations of cortical Hebbian reverberations: theory versus experiment. *J Neurosci* 14:6435–6445.
- Amit DJ, Bernacchia A, Yakovlev V (2003) Multiple-object working memory—a model for behavioral performance. *Cereb Cortex* 13:435–443.
- Asaad WF, Rainer G, Miller EK (1998) Neural activity in the primate prefrontal cortex during associative learning. *Neuron* 21:1399–1407.
- Bauer RH, Fuster JM (1976) Delayed-matching and delayed-response deficit from cooling dorsolateral prefrontal cortex in monkeys. *J Comp Physiol Psychol* 90:293–302.
- Bogacz R, Brown MW (2003) Comparison of computational models of familiarity discrimination in the perirhinal cortex. *Hippocampus* 13:494–524.
- Bogacz R, Brown MW, Giraudo-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* 2:51–62.
- Brown MW, Xiang JZ (1998) Recognition memory: neuronal substrates of the judgment of prior occurrence. *Prog Neurobiol* 55:149–189.
- Brown MW, Wilson FA, Riches IP (1987) Neuronal evidence that inferomedial temporal cortex is more important than hippocampus in certain processes underlying recognition memory. *Brain Res* 409:158–162.
- Brunel N, Carusi F, Fusi S (1998) Slow stochastic Hebbian learning of classes of stimuli in a recurrent neural network. *Network* 9:123–152.
- Camperi M, Wang X-J (1998) A model of visuospatial working memory in prefrontal cortex: recurrent network and cellular bistability. *J Comput Neurosci* 5:383–405.
- Charles DP, Gaffan D, Buckley MJ (2004) Impaired recency judgments and intact novelty judgments after fornix transection in monkeys. *J Neurosci* 24:2037–2044.
- DeGroot MH (1986) Probability and statistics, Ed 2. Reading, MA: Addison-Wesley.
- Derrida B, Gardner E, Zippelius A (1987) An exactly soluble asymmetric neural network model. *Eurphys Lett* 4:167–173.
- Erickson CA, Desimone R (1999) Responses of macaque perirhinal neurons during and after visual stimulus association learning. *J Neurosci* 19:10404–10416.
- Fahy FL, Riches IP, Brown MW (1993) Neuronal activity related to visual recognition memory: long term memory and the encoding of recency and familiarity information in the primate anterior and medial inferior temporal and rhinal cortex. *Exp Brain Res* 96:457–472.
- Fusi S, Annunziato M, Badoni D, Salamon A, Amit DJ (2000) Spike-driven synaptic plasticity: theory, simulation, VLSI implementation. *Neural Comput* 12:2227–2258.
- Green DM, Swets JA (1966) Signal detection theory and psychophysics. New York: Wiley.
- Hölscher C, Rolls ET (2002) Perirhinal cortex neuronal activity is actively related to working memory in the macaque. *Neural Plast* 9:41–51.
- Hopfield JJ (1984) Neurons with graded response have collective computational properties like those of two-state neurons. *Proc Natl Acad Sci USA* 81:3088–3092.
- Li L, Miller EK, Desimone R (1993) The representation of stimulus familiarity in anterior inferior temporal cortex. *J Neurophysiol* 69:1918–1929.
- MacMillan NA, Creelman CD (1991) Detection theory: a user's guide. Cambridge, UK: Cambridge UP.
- Mandler G (1980) Recognizing: the judgment of previous occurrence. *Psychol Rev* 87:252–271.
- McGaugh JL (2000) Memory—a century of consolidation. *Science* 287:248–251.
- Meeter M, Myers CE, Gluck MA (2005) Integrating incremental learning and episodic memory models of the hippocampal region. *Psychol Rev* 112:560–585.
- Miller EK, Desimone R (1994) Parallel neuronal mechanism for short-term memory. *Science* 263:520–522.
- Miller EK, Li L, Desimone R (1993) Activity of neurons in anterior inferior temporal cortex during a short-term memory task. *J Neurosci* 13:1460–1476.
- Miller EK, Erickson CA, Desimone R (1996) Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *J Neurosci* 16:5154–5167.
- Mishkin M, Manning FJ (1978) Non-spatial memory after selective prefrontal lesions in monkeys. *Brain Res* 143:313–323.
- Miyashita Y (1988) Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature* 335:817–820.
- Miyashita Y (1993) Inferior temporal cortex: where visual perception meets memory. *Annu Rev Neurosci* 16:245–263.
- Miyashita Y, Chang HS (1988) Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature* 331:68–70.
- Mongillo G, Amit DJ, Brunel N (2003) Retrospective and prospective persistent activity induced by Hebbian learning in a recurrent cortical network. *Eur J Neurosci* 18:2011–2024.
- Mongillo G, Curti E, Romani S, Amit DJ (2005) Learning in realistic networks of spiking neurons and spike-driven plastic synapses. *Eur J Neurosci* 21:3143–3160.
- Nadal JP, Toulouse G, Changeux JP, Dehaene S (1986) Networks of formal neurons and memory palimpsests. *Eurphys Lett* 1:535–539.
- Nakamura K, Kubota K (1995) Mnemonic firing of neurons in the monkey temporal pole during a visual recognition memory task. *J Neurophysiol* 74:162–178.
- Norman KA, O'Reilly RC (2003) Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. *Psychol Rev* 110:611–646.
- Passingham R (1975) Delayed matching after selective prefrontal lesions in monkeys (*Macaca mulatta*). *Brain Res* 92:89–102.
- Petersen CC, Malenka RC, Nicoll RA, Hopfield JJ (1998) All-or-none potentiation at CA3–CA1 synapses. *Proc Natl Acad Sci USA* 95:4732–4737.
- Sakai K, Miyashita Y (1991) Neural organization for the long-term memory of paired associates. *Nature* 354:152–155.
- Salinas E, Abbott LF (1996) A model of multiplicative neural responses in parietal cortex. *Proc Natl Acad Sci USA* 93:11956–11961.
- Sobotka S, Ringo JL (1993) Investigation of long-term recognition and association memory in unit responses from inferotemporal cortex. *Exp Brain Res* 96:28–38.
- Sompolinsky H (1986) Neural networks with non-linear synapses and static noise. *Phys Rev A* 34:2571–2574.
- Standing L (1973) Learning 10,000 pictures. *Q J Exp Psychol* 25:207–222.
- Tsodyks M (1990) Associative memory in neural networks with binary synapses. *Mod Phys Lett B* 4:713–716.
- Tsodyks M, Feigelson MV (1988) Enhanced storage capacity in neural networks with low level of activity. *Eurphys Lett* 46:101–103.
- Willshaw DJ, Buneman OP, Longuet-Higgins HC (1969) Non-holographic associative memory. *Nature* 222:960–962.
- Wilson HR, Cowan JD (1972) Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys J* 12:1–24.
- Xiang JZ, Brown MW (1998) Differential neuronal encoding of novelty, familiarity and recency in regions of the anterior temporal lobe. *Neuropharmacology* 37:657–676.
- Xiang JZ, Brown MW (2004) Neuronal responses related to long-term recognition memory processes in prefrontal cortex. *Neuron* 42:817–829.
- Yakovlev V, Bernacchia A, Orlov T, Hochstein S, Amit DJ (2005) Multi-item working memory—a behavioral study. *Cereb Cortex* 15:602–615.