



## Model of Familiarity Discrimination in the Perirhinal Cortex

RAFAL BOGACZ

*Department of Computer Science, University of Bristol, Bristol BS8 1UB, UK*

r.bogacz@bristol.ac.uk

MALCOLM W. BROWN

*MRC Centre for Synaptic Plasticity, Department of Anatomy, University of Bristol, Bristol BS8 1TD, UK*

m.w.brown@bristol.ac.uk

CHRISTOPHE GIRAUD-CARRIER

*Department of Computer Science, University of Bristol, Bristol BS8 1UB, UK*

cgc@cs.bns.ac.uk

*Received December 29, 1999; Revised June 12, 2000; Accepted June 19, 2000*

Action Editor: Barry Richmond

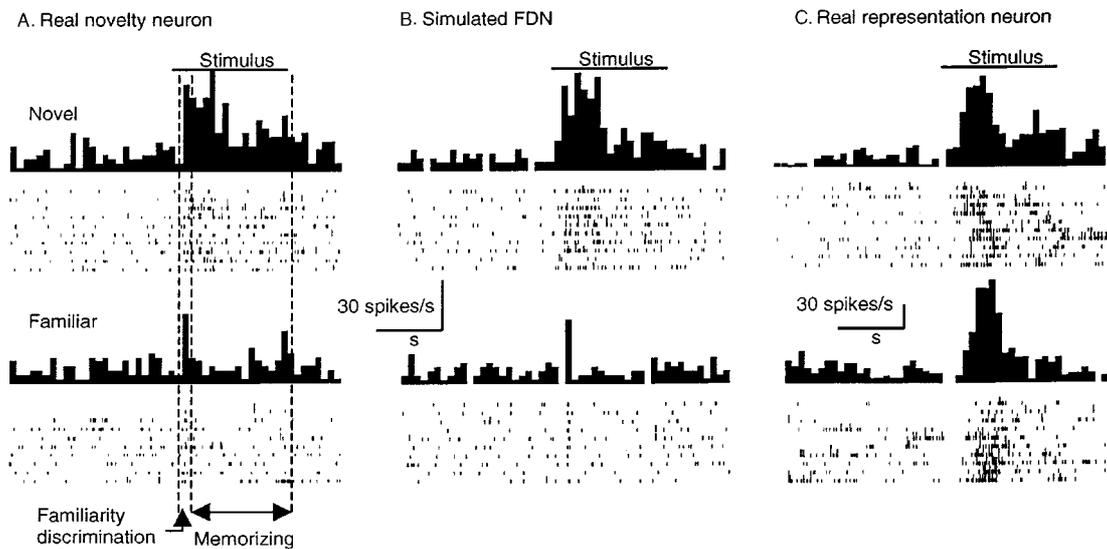
**Abstract.** Much evidence indicates that recognition memory involves two separable processes, recollection and familiarity discrimination, with familiarity discrimination being dependent on the perirhinal cortex of the temporal lobe. Here, we describe a new neural network model designed to mimic the response patterns of perirhinal neurons that signal information concerning the novelty or familiarity of stimuli. The model achieves very fast and accurate familiarity discrimination while employing biologically plausible parameters and Hebbian learning rules. The fact that the activity patterns of the model's simulated neurons are closely similar to those of neurons recorded from the primate perirhinal cortex indicates that this brain region could discriminate familiarity using principles akin to those of the model. If so, the capacity of the model establishes that the perirhinal cortex alone may discriminate the familiarity of many more stimuli than current neural network models indicate could be recalled (recollected) by all the remaining areas of the cerebral cortex. This efficiency and speed of detecting novelty provides an evolutionary advantage, thereby providing a reason for the existence of a familiarity discrimination network in addition to networks used for recollection.

**Keywords:** recognition memory, novelty detection, hippocampal region, computational model, spike-response model

### 1. Introduction

A large body of psychological evidence suggests that recognition memory involves two separable processes, recollection and familiarity discrimination (see, e.g., Mandler, 1980; Hintzman et al., 1998). Indeed, per-

sonal experience indicates that it is not uncommon to be able to recognize that a person is familiar to us even though we cannot immediately recollect anything more about the person or our previous encounters with them. Work in amnesic patients and in monkeys has established that discrimination of the relative familiarity or



*Figure 1.* Comparison of activity of actual and theoretical perirhinal neurons. In each display, a cumulated peristimulus-time histogram is depicted together with a line-by-line raster display of the occurrence times of the neuron's action potentials during the presentation of each of 10 different pictures. **A:** Novelty neuron recorded from monkey perirhinal cortex (Xiang and Brown, 1998). Shown is the strong response of the neuron to 10 novel and short response to 10 familiar pictures. Approximate boundaries of two processing phases are indicated: familiarity discrimination period and memorizing period. **B:** Computer simulation of an FDN using Spike-Response Model obtained for a network consisting of 40 representation neurons, 40 FDNs, and 1 decision neuron. The illustrated FDN has similar response properties to the novelty neuron in panel A. **C:** Visually responsive neuron, which does not differ in its response to novel and familiar patterns, recorded from monkey perirhinal cortex.

novelty of visual stimuli is dependent on the perirhinal cortex (Aggleton and Brown, 1999; Aggleton and Shaw, 1996; Brown and Xiang, 1998; Murray, 1996; Murray and Bussey, 1999). The perirhinal cortex (Brodman's areas 35 and 36) is located along the border of the parahippocampal gyrus, inferomedially in the temporal lobe. The cortical inputs to the perirhinal cortex carry multimodal information and come from association cortical, including visual areas. The perirhinal cortex also has close connections with the amygdala, the hippocampus (mainly via the entorhinal cortex), and the striatum (Murray and Bussey, 1999). Damage to the perirhinal cortex results in impairments in recognition memory tasks that rely on discrimination of the relative familiarity of objects (Murray, 1996). Within the monkey's perirhinal cortex,  $\sim 25\%$  of neurons respond strongly to the sight of novel objects but respond only weakly or briefly when these objects are seen again (Brown and Xiang, 1998; Xiang and Brown, 1998) (Fig. 1a). The population of such responses attests to very fast discrimination of the novelty or familiarity of stimuli: response differences occur within 100 ms of stimulus onset (Xiang and Brown, 1998). This finding accords with the ability of human subjects to make such discriminations rapidly (Seeck et al.,

1997; Hintzman et al., 1998). Additionally, the population of these neuronal responses manifests a very large storage capacity as the responses of individual neurons continue to signal the novelty or familiarity of objects even when many hundreds of objects have been seen (Xiang and Brown, 1998). This capacity is in accordance with the ability of human subjects who, after seeing thousands of different pictures once, can still recognize the individual pictures as familiar (Standing, 1973).

The above evidence suggests that the familiarity discrimination aspect of recognition memory is dependent on the perirhinal cortex. This raises the question of how this relatively small area of the brain can perform familiarity discrimination for such large numbers of stimuli with such efficiency. A further, related question is why the brain should possess separate familiarity discrimination and recollection processes. The purpose of this study is to provide answers to these questions through the description of a neural network model that performs familiarity discrimination and whose behavior is consistent with experimental observations. By varying synaptic weights, the model efficiently stores information about the occurrence of stimuli. The network determines whether a stimulus has been encountered

before but not its associations; accordingly, the model cannot be used to achieve associative recall. With this restriction, the network achieves much higher storage capacity for familiarity discrimination than other neural networks achieve for recall. A brief theoretical analysis of such a method of familiarity discrimination was presented by Bogacz et al. (1999). Here we show how a neural network may be constructed using realistic parametric assumptions and plausible learning rules and compare the activity of its neurons with that of real perirhinal neurones.

To our knowledge, this is the first published model of familiarity discrimination designed to mimic processing in the perirhinal cortex. It differs from previous neural network models of how familiarity discrimination may be performed in the brain, as these use networks possessing the ability to recall information and hence do not achieve such high storage capacity (e.g., Kazer and Sharkey, 1999; Borisjuk et al., 1999). The model also differs from the artificial neural networks used for familiarity discrimination in industrial applications (e.g., Roberts and Tarassenko, 1995; Granger et al., 1998). In these approaches familiarity discrimination is considered as detecting typical patterns of device behavior, since untypical (that is, novel) patterns may be a sign of malfunctions. Hence such models assume that familiar patterns create clusters in representation space (and the synaptic weights of their neurons often encode prototypes of familiar patterns)(e.g., Granger et al., 1998). By contrast, the model outlined here does not require any assumptions concerning the distribution of patterns, and it discriminates whether a particular pattern was presented previously rather than whether the pattern is typical. The proposed network differs also from the novelty filter (Kohonen, 1989), which determines which bits of the delivered pattern differ from the closest familiar pattern. The proposed model has just a single output but may discriminate familiarity for many more patterns than a novelty filter. Although the information processing in the proposed network is similar to that in a novelty detector (Kohonen et al., 1974; Kohonen, 1989), the novelty detector is an abstract model of a single neuron, with correspondingly limited storage capacity. The proposed model is a network of neurons with a very large storage capacity. Furthermore, the performance of the novelty detector was analyzed for an abstract case in which each familiar pattern was presented infinitely many times (Kohonen et al., 1974; Kohonen, 1989), while for the network described here, the storage capacity is calculated for the

case in which each familiar pattern is presented only once before testing.

The network is designed to model familiarity discrimination and not other perirhinal functions (Suzuki, 1996; Murray and Bussey, 1999). Among perirhinal neurons, ~25% respond differently depending on the previous occurrence of visual stimuli (Xiang and Brown, 1998). Another ~30% are visually responsive but do not change their response depending on prior occurrence (see Fig. 1c). (For a computational model of such neurons, see Saksida and Bussey, 1998.) Furthermore, there are three types of neurons with responses carrying information of use to recognition memory—novelty, recency, and familiarity neurons—each with a different pattern of responsiveness (Xiang and Brown, 1998). This article focuses on modeling computations performed by novelty neurons (~10% of perirhinal neurons) that respond strongly to the first presentations of novel stimuli but only briefly to presentations of familiar stimuli. Neuronal simulations were made using the spike-response model (Gerstner, 1998b). This model takes into account the timing of action potentials, conduction delays, neuronal refractory periods, and shapes and linear summation of postsynaptic potentials. This spike-response model is used later in the article. For simplicity of explanation and mathematical analysis, a model based on binary neurons (McCulloch and Pitts, 1943) is used to introduce the network.

The article starts with a description of the proposed network in Section 2. Then the storage capacity and other properties of the proposed network are presented in Section 3. Section 4 discusses how the model relates to experimental observations. The storage capacity of the network is calculated in Appendices A, B, and C. Appendices D and E give all the implementation details to allow replication of the presented simulations.

## 2. Description of the Model

The computations performed by the network are similar to the discrimination of familiarity by checking the energy function of the Hopfield network. As shown in Appendix A, this provides an efficient method of familiarity discrimination. Checking the Hopfield energy is an abstract algorithm, which cannot be performed directly by a biologically plausible network. Simulations have established that there are several neural networks performing analogous computation and achieving a similarly high storage capacity. However, few of those networks exhibit behavior consistent with

experimental observations. Networks that have behavior consistent with experimental observations have been found to share many features. Accordingly, the most biologically plausible of such networks is presented here, with possible variants being discussed in Section 4.

The model will be described in detail in the following sections. In essence, it operates in two periods, a familiarity discrimination period followed by a memorizing period. The critical elements are familiarity discrimination neurons (FDNs) that correspond to the novelty neurons of the perirhinal cortex and make individual decisions about the novelty or familiarity of a stimulus during the familiarity discrimination period. The synapses of FDNs are modified according to Hebbian rules during the memorizing period. The other features of the model are to optimize the operation of the network.

### 2.1. Phases of Processing

The model operates in two phases: a brief initial period when the discrimination of familiarity is achieved (familiarity discrimination period) and a subsequent, longer period during which information storage is effected (memorizing period). Work in information theory and in analysis of experimental observations relating to information theory indicates that the first impulses in a spike train carry the most information (Rieke et al., 1997; Rolls and Tovee, 1994). Consonant with this, the model assigns a special computational role in judging familiarity to the first spike(s) after stimulus presentation—that is, the spikes occurring in a brief initial interval, in which novelty neurons are active for both novel and familiar patterns (see Fig. 1a). The subsequent memorizing period has high-frequency spike activity for a novel but not for a familiar stimulus. This high-frequency activity produces differential modifications of synaptic weights, thereby storing the occurrence of the novel stimulus. At this time, there is no high-frequency activity for familiar stimuli (Fig. 1a) because their occurrence has already been stored in the network and hence further synaptic weight modification is not necessary.

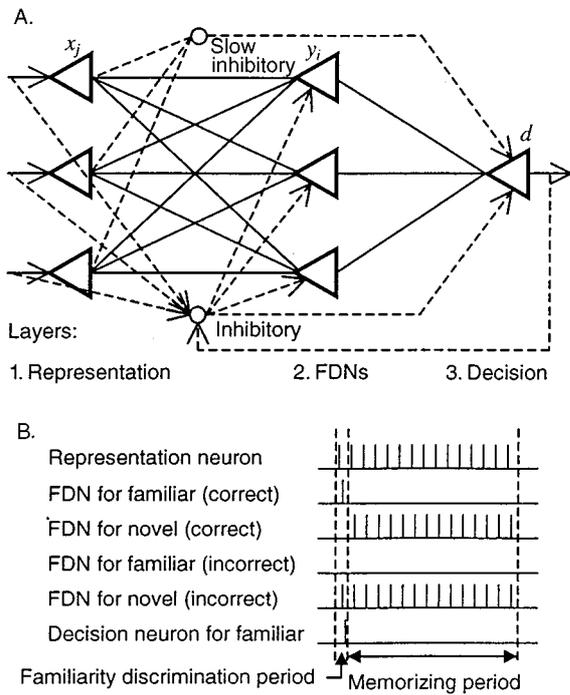
After presentation of a novel stimulus, the high-frequency activity during the memorizing period (Fig. 1a) effects the redistribution of synaptic weights according to Hebbian learning rules: increases in weights simulate, for example, long-term potentiation (LTP), and decreases simulate, for example, long-

term depression (LTD) (Bliss and Collingridge, 1993; Ito, 1989). The redistribution of weights results in an increase in the magnitude of postsynaptic potentials produced by the first spikes evoked by a future occurrence of that stimulus. The consequent increase in neuronal firing during the familiarity discrimination period when a stimulus reappears results in the stimulus being judged familiar. It might be thought that this increase during the familiarity discrimination period would be observable in the histogram of Fig. 1a. However, simulations demonstrate that the increase need not be large, need not affect many neurons, and hence cannot be expected to be readily observable on such histograms (see also Fig. 1b). Moreover, the memorizing period follows the familiarity discrimination period immediately, so that any increase in the histograms will tend to be masked by the onset of the faster activity of the memorizing period.

### 2.2. Information Processing in the Familiarity Discrimination Period

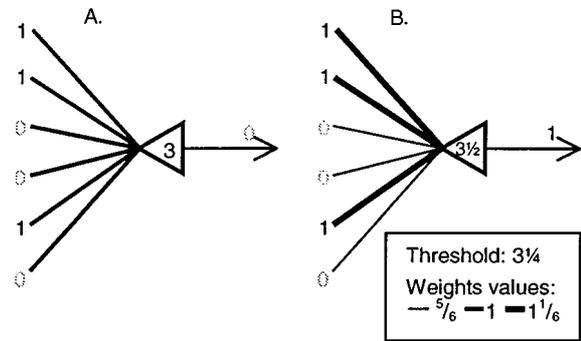
The model has three layers (Fig. 2a). The first layer of  $N$  neurons provides inputs to the familiarity discrimination network. These neurons correspond to visually responsive neurons, which do not alter their responses according to the novelty or familiarity of a pattern (Fig. 1c) and are here referred to as *representation neurons* because they collectively provide a representation of the stimulus. Almost all visually responsive neurons in the perirhinal cortex are stimulus selective (Brown and Xiang, 1998) so there is a particular pattern of activity of representation neurons associated with each presented stimulus. Here, for simplicity, the pattern for each stimulus is encoded as a sequence of  $N$  bits, where 1 denotes that the corresponding representation neuron belongs to the assembly of neurons that fires for that stimulus, while 0 indicates that it does not. In this model the representation neurons are considered just as an input to the network. One possible model of how such representation neurons may be produced is given in Saksida and Bussey (1998).

The second layer consists of *familiarity detection neurons (FDNs)*; these correspond to experimentally observed novelty neurons (Xiang and Brown, 1998). FDNs make individual, independent decisions about the familiarity of the stimulus. The modifiable connections of the model result in more FDNs being active for familiar patterns than for novel ones during the initial period of processing (Fig. 3). The synaptic efficacies of



**Figure 2.** Architecture of the neural network model. **A:** 3-layer network (of representation, familiarity discrimination, and decision neurons) performing 3 steps of processing. Differential activity after the first period of processing may be achieved using inhibition. After presentation of a familiar stimulus, decision neurons (here for simplicity only one is shown) are activated and trigger the excitation of inhibitory neurons, denoted in the figure by “Inhibitory,” which prevent activity in the FDNs during the memorization period. For novel patterns, the decision neuron is not active; hence the inhibition is not increased, and the FDNs are active during the memorization period. Since FDNs are active in the memorizing period for novel patterns, they could potentially activate the decision neuron for a novel pattern. To prevent the decision neurons from firing during pattern memorization, inhibitory neurons denoted by “Slow Inh.” (here again for simplicity only one is shown) are introduced. The activity of these inhibitory neurons is delayed compared to activity from the FDNs, so allowing the decision neurons to fire only in the initial familiarity discrimination period (until the “Slow inh.” neurons turn the decision neurons off). **B:** Schematic activity of the perirhinal neurons from the model. In the brief initial period, the majority of FDNs are active for familiar patterns and inactive for novel (correct). Although a fraction of FDNs may behave in the opposite way (incorrect), the network results in decision neurons being active for familiar and inactive for novel stimuli with high accuracy. Hence in the memorizing period appropriate FDNs are active for novel and inactive for familiar stimuli.

inputs to FDNs active during the memorizing period for novel patterns are redistributed according to Hebbian rules. In the model, the weights from active inputs to an FDN are increased by  $1/N$  (simulating long-term potentiation) (LTP). To keep average excitability lev-



**Figure 3.** Behavior and synaptic plasticity of FDNs. **A:** The membrane potential of an FDN (the number in the triangle) is determined by the sum of the weights of its active inputs (the sample FDN has  $N = 6$  inputs). Hence the higher the correlation between the weights and the input pattern for any individual FDN, the higher its membrane potential in the initial period. The FDN will be active only if its membrane potential exceeds a threshold. For simplicity, initially all synaptic weights are equal (the initial value of all weights in the example is 1). The threshold of each FDN is set such that most FDNs are inactive for novel patterns—that is, it is set above the value of the membrane potential averaged across the whole population. It is defined in the example as  $1/2(N \times \text{initial value of weights}) + 1/4$ ; it equals  $3\frac{1}{4}$ . In the example, after delivery of a sample pattern (110010), the membrane potential is equal to 3; this is below the threshold and the FDN is inactive. **B:** Modification of synaptic weights according to Hebbian rules during the memorizing period. When the same pattern is seen again, the value of the membrane potential is higher in the initial period ( $3\frac{1}{2}$ ) and the FDN is active.

els within the network constant, other synapses are weakened (simulating hetero-synaptic long-term depression) (LTD) (Fig. 3). Thus, during learning, synaptic weights become correlated to the FDN inputs active for that pattern. Hence, when the same pattern is seen again, the value of the membrane potential is higher in the initial period, and the FDN is more likely to fire. To ensure the decisions made by FDNs are independent of the number of representation neurons that are active (that is, to compensate for differences in the level of activity for different patterns), inhibitory neurons (denoted “Inhibitory” in Fig. 2a) increase the threshold of FDNs in proportion to the number of such neurons that are active. A similar function for inhibitory neurons was proposed in the neocortex model of Marr (1970). In order for the adjustment to be effective, the inhibition must reach FDNs approximately at the same time as the excitation from the representation neurons. Hence, the inhibitory neurons receive input from axons afferent to the representation neurons rather than from axons of the representation neurons themselves. Such a mechanism is satisfactory under the plausible assumption

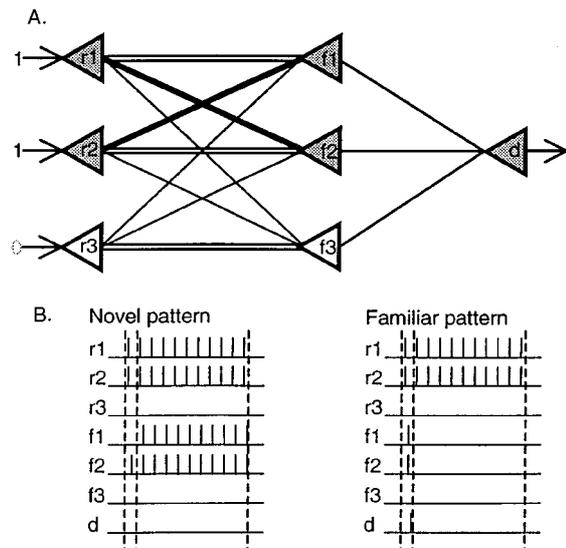
that this afferent activity predicts the activity of the representation neurons.

Since any individual FDN may make a mistake—that is, be active for a novel stimulus during this initial period—a mechanism is required to sample the activity of the population of FDNs and make a judgment based on their conjoint activity. In the current model, this mechanism is supplied by a third layer of *decision neurons*. Decision neurons receive inputs from the FDNs and are activated only when a majority of their inputs are active. These decision neurons govern the subsequent activity of the network, particularly the activity that will produce changes in synaptic efficacy according to Hebbian learning (LTP or LTD) (Bliss and Collingridge, 1993; Ito, 1989) for novel but not familiar patterns. Essentially, these neurons act as enforcers of the network decision. These neurons could be regarded as the output of the network in that they carry its decision; however, the FDNs provide a signal that is more biologically useful for distribution to other networks (brain regions) as it potentially carries more information, and, moreover, this information is accurate following the enforcement of the FDNs' activity provided by the decision neurons. As all the decision neurons perform the same task, theoretically there need be only one decision neuron in the whole network. However, limitations on the number of synapses of a single neuron mean that one neuron could not sample the whole population of perirhinal FDNs. Assuming a perirhinal decision neuron has  $10^4$  synapses, there need be only one decision neuron to every  $10^4$  FDNs. However, to make the network resistant to damage, the number of decision neurons may be increased to, for example, one to every 1000 FDNs without significantly increasing total network size.

Various neuronal circuits can implement the conditions necessary for controlling the modification of synaptic weights. Here, inhibitory connections are used (Fig. 2a); these result in the behavior of the neurons in the network mimicking those of experimentally observed perirhinal neurons (Fig. 1b) (Xiang and Brown, 1998) and are consistent with findings on homosynaptic LTD (described in Section 4). Whatever circuitry or underlying synaptic plastic mechanism is used, it needs to be sufficiently fast to produce the observed reduction in perirhinal neuronal responses that occurs even when a stimulus is repeated within less than a second (Miller and Desimone, 1993).

To maximize the information storage capacity of the network, it is necessary to ensure that individual FDNs

remain independent assessors of familiarity. For example, if all FDNs were active during the memorizing period, then the synaptic weights of all FDNs would be modified in the same way and all the FDNs would come to have highly correlated weights. Hence, eventually, they would all be active or inactive together, and the whole network would have the same capacity as a single FDN. To avoid this problem, the number of active FDNs must be limited during memorizing. It is therefore necessary that only a subset of FDNs respond to any given stimulus. One simple way of ensuring this selectivity of response is to provide specific connections with high synaptic weights that are not changed during learning (double lines in Fig. 4) from the representation neurons to the FDNs. For simplicity of analyzing network behavior, in this article we consider such connections as one-to-one connections, but they may be realized as few-to-few connections (For example, within local groups of neurons in the perirhinal cortex) without necessarily compromising storage



**Figure 4.** Synaptic weight modification in the network after delivery of a sample novel pattern. **A:** Weight modification. Double lines denote powerful, nonmodifiable connections; thick lines, weights that are increased during learning; thin lines, weights that are decreased. The  $f_3$  neuron does not receive activation through a strong nonmodifiable connection; hence it is not activated even after a decrease of inhibition. The weights of the FDNs are modified in a Hebbian manner—for example, weights between active units  $r_1$ - $f_2$ ,  $r_2$ - $f_1$  are increased (as if due to LTP); weights from inactive to active units  $r_3$ - $f_1$ ,  $r_3$ - $f_2$  are decreased (as if due to heterosynaptic LTD); and weights from active to inactive units  $r_1$ - $f_3$ ,  $r_2$ - $f_3$  are decreased (as if due to homosynaptic LTD). **B:** Corresponding network activity, as for Fig. 2b.

capacity (as indicated by simulations). These connections ensure that for an FDN to be active, the corresponding neuron in the first layer must also be active. The FDNs are thus made input sensitive and respond to only a subset of patterns—as do real perirhinal neurons (Xiang and Brown, 1998). Limitation of the number of active FDNs may additionally or alternatively be achieved by inhibition. Further to uncorrelate the activity of FDNs, the weights of connections between active representation neurons and inactive FDNs are reduced—as if, for example, by homosynaptic LTD (see Fig. 4; a more formal explanation of this weight reduction may be found in Appendices B and D). This reduction also ensures that the average synaptic weight and hence the mean excitability of the network remains constant. Mathematical analysis of this network is shown in Appendix B. Appendix D gives details of the simple binary model, and Appendix E of the spike-response model. Modeling of precise spike timing described in Appendix E does not bring any new computational properties, but it allows observation of the behavior of neurons in the network (Fig. 1b) and hence comparison with the activity of real perirhinal neurons.

### 3. Storage Capacity and Other Properties of the Network

The computations performed by the network in the initial familiarity discrimination period are similar to those of the abstract algorithm of familiarity discrimination by checking the energy function of the Hopfield network (Bogacz et al., 1999). This method allows a neural network having  $N$  neurons to discriminate familiarity with 99% reliability for  $0.023N^2$  patterns (see Appendix A). This capacity is much greater than the standard capacity of the Hopfield network for retrieval—namely,  $0.145N$  (Amit, 1989). The storage capacity of the network in the model is two times smaller—that is, assuming the probability of error 1%, the network including  $N$  FDNs can discriminate familiarity for  $0.012N^2$  patterns (which is proven in Appendix B). If a higher reliability is required, the familiarity discrimination capacity slightly decreases, but it still is of order  $N^2$  (For example, for an error probability of  $10^{-6}$  the capacity is about  $0.003N^2$ ). Figure 5 compares the capacities of the perirhinal network model and the Hopfield model found in simulations, with the corresponding theoretical predictions.

The above formulae are true in the case of a fully connected network. However, the capacity of the pro-

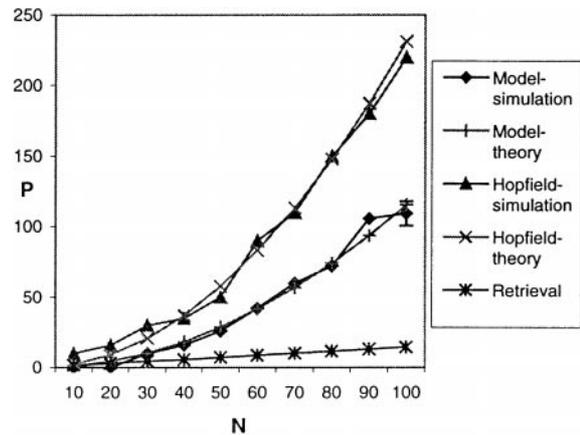


Figure 5. Comparison of the simulated familiarity discrimination capacities of the perirhinal network and the Hopfield model, with the corresponding theoretical predictions (standard capacity for retrieval is also shown). For each number of inputs  $N$ , and for each number of previously stored patterns  $P$ , the behavior of the network was tested repeatedly with sets of random patterns until it had been tested with 500 previously stored patterns and 500 random patterns for which the absolute value of the correlation with each stored pattern was less than 0.5. For each number of inputs  $N$ ,  $P_{max}$  is taken as the maximum number of stored patterns  $P$ , for which the error rate is  $\leq 1\%$ . To illustrate the precision of the simulation process, for one data point ( $N = 100$ ), the capacity was estimated 10 times using the above method. The standard deviation of the maximum number of stored patterns was  $\pm 8.64$  (shown by error bar).

posed familiarity discrimination network remains proportional to the number of synapses rather than the number of neurons, even for networks that are not fully connected (see Appendix C and Fig. 6). Making a familiarity judgment may be considered to be a single bit classification, and storage capacity has been also found to be proportional to the number of synapses for classification networks based on a single neuron classifier (Cover, 1965) and for a multilayer perceptron with one output (Kowalczyk, 1997).

In the Hopfield network the inverse patterns to stored patterns (where all the zeros are replaced by ones and ones by zeros, such as 1001011 and 0110100) have the same energy value as corresponding original patterns. Therefore, in theory such inverted patterns would be classified by the proposed network as familiar, although they are very different from those stored. However, experimental and theoretical work indicates that patterns are likely to be encoded in a sparse way—that is, for a given pattern the fraction of neurons that are simultaneously active will be much less than a half (there will be fewer ones than zeros in the binary representations of patterns) (Foldiak and Young, 1995). With this

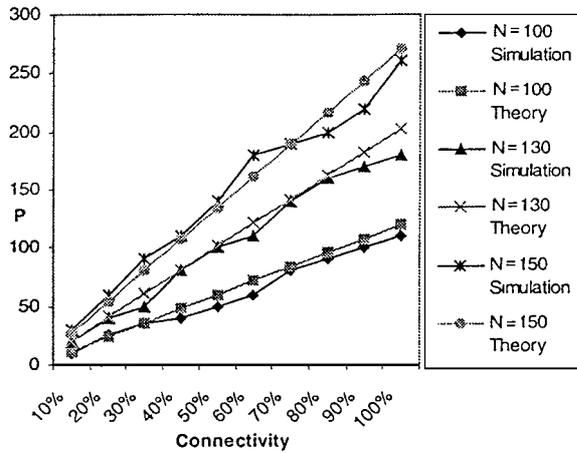


Figure 6. Comparison of the simulated familiarity discrimination capacities of the perirhinal network for different degrees of sparseness of connectivity, with the corresponding theoretical predictions. Method of simulation as in Fig. 5.

constraint, the inverted patterns cannot exist (because they would consist of more ones than zeros) so that mistakes involving inverted patterns would not occur. In this article, however, sparse coding is not used. For simplicity it is assumed that the probabilities of neuron being active and inactive in a particular pattern are equal. This simplification was made to make analysis of network properties in the Appendices simpler and clearer. However, simulations establish that the operation of the network is essentially unchanged should encoding be sparse.

The proposed network may still under certain circumstances classify a novel pattern as familiar (that is, mimic false alarms during recognition memory tests). The patterns for which this error may be made correspond to spurious attractors in the Hopfield network—states having lower values of the energy function than the stored patterns (Amit, 1989). By decreasing the number of stored patterns  $P$ , the probability of such errors may be decreased, but spurious attractors exist even for very small  $P$  (Amit, 1989). For example, mixtures of stored patterns (that is, patterns showing substantial overlap with many of the stored patterns) have a high probability of being spurious attractors (Amit, 1989). Although for a given set of stored patterns it is not difficult to construct novel patterns that will be classified as familiar (as illustrated in Section 4), the probability that a random novel pattern is classified as familiar is very small until the network is near saturation (Appendix B).

The network demonstrates generalization and is resistant to disruption by noise. A pattern will still be classified as familiar even if it differs in a substantial proportion of its bits from its previous representation (as might result from changes in, for example, the orientation of the stimulus or part of the stimulus being obscured). More precisely, a pattern will be classified as familiar if the Hamming distance (number of different bits) between the pattern and one of the stored patterns is small. It is plausible that presentations of very similar stimuli would indeed result in similar patterns of activity of the representation neurons. It is predicted by the majority of algorithms modeling feature extraction in the cortex, such as independent component analysis (Olshausen and Field, 1996), and is consistent with the neurophysiological observations in area TE adjacent to the perirhinal cortex (Kobatake et al., 1998). The network is also very resistant to damage. The model will work after the removal of synapses of FDNs, or even whole FDNs, with the capacity per synapse remaining the same (see Appendix C). As few decision neurons are needed, it is not costly to build redundancy into the network by increasing their numbers. Moreover, as all the decision neurons perform the same task, the network's capacity will not be affected by loss of their synapses or even of whole decision neurons.

The proposed network has two processing layers. In artificial neural networks, multilayer networks are used when the problem is nonlinearly separable (Hertz et al., 1991). Familiarity discrimination may be perceived as a simple linearly separable problem and may be solved by a single neuron (Bogacz et al., 1999). However, the single neuron has a very limited storage capacity. The two layers of the proposed network are required to achieve high storage capacity.

If the human perirhinal cortical network operates on similar principles to the proposed model, its theoretical capacity may be estimated on the assumption that it contains  $\sim 4 \times 10^7$  pyramidal neurons (Insausti et al., 1998), of which  $\sim 4 \times 10^6$  are novelty neurons, each with  $\sim 10^4$  potentially modifiable synapses. With a probability of error of  $10^{-6}$ , the perirhinal network of novelty neurons could discriminate familiarity for  $\sim 10^8$  patterns. For comparison, the human hippocampal subfield CA3, which has been modeled as an autoassociative memory (Marr, 1971; Rolls, 1996), contains  $\sim 2.3 \times 10^6$  pyramidal neurons (Cassell, 1980)—a number in the same range as that of perirhinal novelty neurons—with  $\sim 40,000$  potentially modifiable synapses per neuron (Cragg, 1975). If the storage

capacity of this network were calculated according to the model proposed by Rolls (1996) with the model's parameters (sparseness = 0.02 and structure of connectivity factor = 0.2), then the human CA3 could perform recall for only  $\sim 10^5$  patterns.

Furthermore, familiarity discrimination in the proposed network is very fast. To discriminate familiarity after the information about a stimulus reaches the representation neurons, the information needs to be processed by only two layers of neurons and so may take less than  $\sim 10$  ms. In contrast, associative memories that perform recall require a longer relaxation process (Amit, 1989).

#### 4. Model Behavior and Experimental Observations

Computer implementation of this network using parameters concerning excitatory and inhibitory postsynaptic potential durations, refractory periods, and conduction delays for cortical neurons (Fohlmeister et al., 1995) (see Appendix E for details) obtained the behavior for the FDNs exemplified in Fig. 1b. The activity patterns of neurons in the network are closely similar to those observed experimentally for neurons in the perirhinal cortex (Xiang and Brown, 1998). In particular, the responses of FDNs are much shorter and/or weaker for subsequent than for first presentations, as for real novelty perirhinal neurons (Xiang and Brown, 1998). Studies of cross-correlograms of neurons recorded from the perirhinal cortex suggest that there are likely to exist direct connections from representation neurons to FDNs (Brown and Xiang, 1998), consistent with the structure of connections in the proposed network.

According to the model, homosynaptic LTD should occur at synapses between active representation neurons and inactive FDNs for novel stimuli but not for familiar stimuli. In this context it may be significant that in slices of the perirhinal cortex maintained *in vitro*, homosynaptic LTD is more readily induced when the level of inhibition in the network is decreased (Ziakopoulos et al., 1999; Cho et al., 2000). This finding is consistent with the model, where the level of inhibition is low for novel and high for familiar patterns.

In the model, the computation of stimulus familiarity is performed by the first spike(s) of the train. Markram and Tsodyks (1996) have found that LTP at synapses of cortical neurons can increase the probability of the first spike in a train generating a *large* postsynaptic po-

tential. Thus under these conditions, synaptic efficacy changes may be utilized by computations involving the first spike(s) of the train.

The model is also consistent with the observation that gross stimulation or epileptic attacks involving the temporal lobe in human epileptic patients may result in *déjà vu*, the patients reporting inappropriate subjective feelings of familiarity (Bancaud et al., 1994). In the model, epileptic seizures would be likely to cause simultaneous excitation of large numbers of FDNs, so leading to the network classifying (falsely) many stimuli as familiar.

If a novel pattern shows substantial overlap with a number of previously stored patterns, it may mistakenly be judged as familiar by the proposed network (see Section 3). In this respect it demonstrates false memory, as do human subjects under similar conditions. Thus when subjects are read a list of semantically related words (such as sugar, taste, sour, chocolate, etc.) and then asked about words that are semantically related to the presented words but that were not on the list (such as sweet), they commonly mistakenly claim such words were presented (they produce false alarms) (Roediger and McDermott, 1995). A similar, simulated recognition memory test was presented to the proposed network. During simulations, first a random pattern  $A$  of  $N = 50$  bits was generated. Then 15 patterns sharing 20% bits (that is, 10 bits) with pattern  $A$  (different bits were shared by different patterns) were presented to the network but not pattern  $A$  itself. Such overlap in the binary representations of the patterns was chosen to simulate semantic overlap between words in psychological experiments. When the network judged the familiarity of pattern  $A$ , it was mistakenly classified as familiar in 62.4% of 1000 performed experiments, a figure within the range of false alarms reported for human subjects in tests of false memory (Roediger and McDermott, 1995). Note that the generation of such false memories requires there to be numerous similar repetitions. In the simulations, overlap between patterns was only 20%; hence such an error would not occur if the number of similar patterns were small (such as 3 patterns). Analogously, in the Roediger-McDermott paradigm false memories do not occur if the list of words has just a few items. The suggestion that the perirhinal network may be involved in creation of false memories under such conditions is consistent with studies of amnesic patients with damage in the medial temporal lobes (Schacter et al., 1998). Amnesic patients during such tests make more nonrecognition errors and

more false alarm errors for words nonsemantically related to the presented words than do controls, but they make fewer false alarms for words semantically related to the presented words. PET studies also demonstrate that the parahippocampal region is involved in recognition memory and that during recognition memory tests according to the Roediger-McDermott paradigm, the neuronal activity during recognition of false memories and real memories are similar (Schacter et al., 1996). This is also consistent with the behavior of the model, where neuronal activity is similar for stored patterns and “false memory patterns.” This model does not allow explanation of all types of false memories (Schacter, 1999), but its behavior is consistent with false memories during recognition memory tests.

The model furthermore makes the following testable predictions:

1. There will be specific connections between, for example, representation neurons (visually responsive neurons) and FDNs (differentially responsive neurons), and particular connections will, as indicated in Fig. 4, variously undergo LTP-like or LTD-like modifications or be not plastic.
2. Drugs that block changes in synaptic efficacy without affecting normal transmission (for example, blockers of N-methyl-d-aspartate or metabotropic glutamate receptors) will leave familiarity discrimination unimpaired unless a novel stimulus presented during drug action is re-presented (when it will be judged to be still novel). Behaviorally, it would be as if familiarity discrimination were normal except for the rapid forgetting of the occurrence of novel items.
3. Interfering stimulation at the time of the first few spikes evoked by a stimulus will disrupt familiarity discrimination and storage, whereas such stimulation slightly later in the train will disrupt only storage. Consistent with this suggestion, brief electrical stimulation of a monkey’s temporal cortex during acquisition disrupts subsequent familiarity discrimination (Ringo, 1995).
4. Increasing inhibition (for example, by  $\gamma$ -aminobutyric acid agonists) will impair storage (due to inhibition of FDNs during the memorizing period) and may lead to increases in nonrecognition errors during recognition memory tests (due to inhibition of FDNs during familiarity discrimination).
5. Overexcitation of the network by decreasing inhibition or prolonged abnormal stimulation or epilepsy

will potentially result in increases in false alarm errors in recognition memory tests (due to excitation of a large number of FDNs).

In creating the model of the network, different architectures were tested before finally developing the current model whose behavior parallels experimental findings. For example, initially recurrent networks were used (inspired by the model of Hopfield, 1982). However, the spike trains generated by these networks were very regular—that is, the intervals between spikes in trains were very similar (as is typical for recurrent networks: Gerstner, 1998a). The irregularity of the real spike trains of perirhinal neurons (see Fig. 1a) led to the use of a network with primarily feed-forward rather than recurrent connections.

The computational part of the network (that is, FDNs) is consistent with many experimental observations. Here we consider how the control part of the network (decision neurons, slow inhibitory neurons, triggering connections) might be recognized experimentally. The number of inhibitory neurons that are involved in controlling the activity of the FDNs and decision neurons need not to be large. Moreover, not all perirhinal inhibitory neurons will be involved in these processes; others will control the activity of representation neurons, for example. If as is possible, these neurons can produce the necessary levels of inhibition without large changes in firing rate; then identifying such neurons will be experimentally difficult. According to the model, decision neurons also do not need to be numerous (see Section 2.2), and therefore they are likely to be difficult to find in recording experiments *in vivo*. The current architecture results in decision neurons having a brief response to familiar stimuli during the familiarity discrimination period. However, the model indicates that this response need only be weak, and therefore, even if the activity of such neurons has been sampled, their responses are unlikely to have been specifically identified. Other architectures, which would also allow familiarity discrimination and be consistent with observations concerning FDNs, would make decision neurons even more unlikely to be identified. For example, if “Slow inh.” neurons have projections to “triggering” synapses connecting the decision to the inhibitory neurons (instead of directly to the decision neurons), decision neurons would be briefly active for familiar patterns in the familiarity discrimination period and more active for novel patterns during memorizing period. Thus in this model,

the decision neurons would have responses similar to FDNs (novelty neurons). Alternatively, if the FDNs projected directly to the inhibitory neurons with the “triggering” synapses (and “Slow inh.” neurons projected to these synapses), the network would contain no decision neurons. Thus, because such architectures are quite plausible, searching experimentally for decision neurons is unlikely to yield a critical test of the model. Only further, precise experimental observations will be able to discriminate between such variants of the model.

Simulations have also showed that behavior consistent with experimental observations is exhibited by a network in which the weights of FDNs are modified according to anti-Hebbian learning—that is, in the opposite way to the current model (using LTD instead of LTP). In this model FDNs detect novelty (instead of familiarity), and the decision neurons are active for novel patterns.

The model is potentially capable of being extended to take account of the interactions and the temporal dependencies between different types of neurons (novelty, recency, and familiarity) in the perirhinal cortex (Xiang and Brown, 1997, 1998; the extension is described in Bogacz et al., in press). Additionally, the model uses an idealized representation of information—that is, the representation neurons belonging to an assembly representing a particular stimulus start to fire approximately synchronously after stimulus presentation (see Appendix E). Such a high reliability and synchronization of representation neurons has not been observed in the perirhinal cortex (Brown and Xiang, 1998; Xiang and Brown, 1997). However, such variability need not cause a major problem as long as the familiarity discrimination period is not too brief, and the activity afferent to the FDNs is required to exceed a threshold before the FDNs become active. Thus in the brain the exact timing of the familiarity discrimination period may vary from one stimulus to another. It will be possible to update the model once further research has established more details of how the information is represented in the perirhinal cortex.

The consistency of the model with observations suggests that the fundamental computational principles presented here are correct. However, it is probable that the structure of circuits in the real perirhinal cortex differs from those proposed here, and further experiments (such as those suggested above) are necessary to disclose to what extent the model’s proposed architecture does indeed parallel that in perirhinal cortex.

## 5. Conclusion

If, as experimental evidence makes plausible, the brain possesses a familiarity discrimination neural network operating on similar principles and with a speed, capacity, and accuracy even approaching those of the model, neural networks involved in categorization, association, and recall would not need to perform familiarity discrimination. Storage of the occurrence of a novel stimulus could be achieved within the familiarity discrimination network, rendering it unnecessary to change synaptic connections within perceptual and categorization networks to perform this function (Xiang and Brown, 1998). In contrast, storing new and recollecting old associations of a stimulus, including its context of occurrence, necessitates the existence of systems additional to such a familiarity discrimination network. The model thus suggests why recognition memory involves two separable processes, one for familiarity discrimination and one for associative recollection.

## Appendix A: Familiarity Discrimination in the Hopfield Network

The Hopfield network provides a simple model of associative memory (Hopfield, 1982). It is a fully connected recurrent neural net consisting of  $N$  neurons, whose activations are denoted by  $x_i$ . The active state of a neuron is represented by 1, and the inactive state by  $-1$ . The patterns stored by the network are denoted by  $\xi^\mu$ , and the number of these patterns by  $P$ . The weight of the connection between neurons  $j$  and  $i$  is denoted by  $w_{ij}$  and computed according to the Hebb rule (Hopfield, 1982):

$$w_{ij} = \begin{cases} \frac{1}{N} \sum_{\mu=1}^P \xi_i^\mu \xi_j^\mu & \text{for } i \neq j \\ 0 & \text{for } i = j \end{cases}. \quad (1)$$

The energy of the Hopfield network is defined by (Hopfield, 1982)

$$E(\bar{x}) = -\frac{1}{2} \sum_{i=1}^N x_i \sum_{j=1}^N x_j w_{ij}. \quad (2)$$

The value of the energy function is usually lower for stored patterns and higher for other patterns (Amit, 1989). Therefore, the value of the energy may be used for familiarity discrimination, which in this context corresponds to checking whether a pattern is stored in the

Hopfield network (Bogacz et al., 1999). Normally, the Hopfield network is used for retrieval of information by updating one-by-one the activities of the neurons (a process called *relaxation*), and the network can perform this task if the number of stored patterns does not exceed  $0.145N$  (Amit, 1989). In the approach presented in this Appendix, the neurons do not perform any computations (there is no relaxation), but the familiarity discrimination is done by checking the value of the energy function after delivering a pattern. In other words, the discrimination is done not by the Hopfield network itself but by an external agency that sets up the activations of the neurons according to a discriminated pattern and calculates the network's energy for this pattern. The remainder of this Appendix shows that this agency is able to discriminate familiarity even if the number of stored patterns is so high that the network cannot retrieve any information (that is, it is in the spin-glass state).

The (double) value of the energy function after delivering to the network one of the stored patterns, (such as  $\xi^1$ ) is given by (Bogacz et al., 1999):

$$\begin{aligned}
2E(\bar{\xi}^1) &= -\sum_{i=1}^N \xi_i^1 \sum_{j=1}^N \xi_j^1 w_{ij} \\
&= -\sum_{i=1}^N \xi_i^1 \sum_{j=1}^N \xi_j^1 \frac{1}{N} \sum_{\mu=1}^P \xi_i^\mu \xi_j^\mu \\
&= -\frac{1}{N} \sum_{i=1}^N \xi_i^1 \sum_{\mu=1}^P \sum_{j=1}^N \xi_j^1 \xi_i^\mu \xi_j^\mu \\
&= -\frac{1}{N} \sum_{i=1}^N \xi_i^1 \left( \sum_{j=1}^N \xi_j^1 \xi_i^1 \xi_j^1 + \sum_{\mu=2}^P \sum_{j=1}^N \xi_j^1 \xi_i^\mu \xi_j^\mu \right) \\
&= -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N (\xi_i^1 \xi_j^1)^2 \\
&\quad - \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N \sum_{\mu=2}^P \xi_i^1 \xi_j^1 \xi_i^\mu \xi_j^\mu. \tag{3}
\end{aligned}$$

According to signal-to-noise analysis (Amit, 1989), the first term in Eq. (3) is called “signal” and the second “noise.” Since  $\xi_j^\mu$  may be equal to  $-1$  or  $+1$ , elements of the summations in the signal term are equal to 1, so the signal term reduces to  $-N$ . The elements of the summations in the noise term are equal either to  $-1$  or  $+1$ , so assuming random patterns they may be treated as discrete random variables, which can be equal to  $-1$  or  $+1$  with equal probabilities. So these variables have mean 0 and variance 1. The noise term is a sum of

$N^2(P-1)$  such variables (so it may be approximated by a sum of  $N^2P$  variables) divided by  $N$ . The noise has binomial distribution and may be approximated with a normal distribution  $\theta(\mu, \sigma)$ , with mean  $\mu = 0$ . From the above it could be expected that the standard deviation of the noise  $\sigma = P^{1/2}$ . However, due to the symmetry of weights in the Hopfield model, the elements of the summations in the noise term are symmetrical after exchanging indices  $i$  and  $j$ . Since the summing is done over both indices (and  $w_{ij} = 0$ ), there are pairs of the same elements in the sum. Hence, one can add them once and multiply the result by 2, as follows (Bogacz et al., 1999):

$$\begin{aligned}
2E(\bar{\xi}^1) &= -N - \frac{2}{N} \sum_{i=1}^N \sum_{j=i+1}^N \sum_{\mu=2}^P \xi_i^1 \xi_i^\mu \xi_j^1 \xi_j^\mu \\
&\approx -N + \theta(0, \sqrt{2P}). \tag{4}
\end{aligned}$$

Hence due to the symmetry of weights in the network the variance of the noise is two times higher. Similarly, the (double) value of the energy after delivering a random pattern that is not correlated with any stored pattern satisfies

$$2E(\bar{R}) \approx \theta(0, \sqrt{2P}). \tag{5}$$

In both cases (Eqs. (4) and (5)) the noise has mean zero, and the average value of  $2E$  for stored patterns is  $-N$ , while for novel patterns it is 0. Therefore, by taking as threshold the middle value  $-N/2$ , we can define a familiarity discrimination criterion—namely, if  $2E < -N/2$ , or equivalently  $E < -N/4$ , then the pattern is considered familiar; otherwise, it is novel.

This familiarity discrimination algorithm works well when the noise  $\theta$  is small. We consider the algorithm as working well if the probability of error is less than 1%. An error occurs if the noise is higher than the threshold  $-N/2$ . To calculate the maximum acceptable number of stored patterns  $P_{\max}$ , we must solve the following equation:

$$\Pr\left(\theta(0, \sqrt{2P_{\max}}) < \frac{N}{2}\right) = 0.99, \tag{6}$$

where  $\Pr$  denotes probability. Equation (6) is equivalent to

$$\Pr\left(\theta(0, 1) < \frac{N}{\sqrt{8P_{\max}}}\right) = 0.99. \tag{7}$$

Since the noise may be estimated by a normal distribution, Eq. (7) may be solved by checking the value of

the inverted standard normal cumulative distribution for 0.99:

$$\frac{N}{\sqrt{8P_{\max}}} \approx 2.33. \quad (8)$$

Solving Eq. (8) with respect to  $P_{\max}$ , we get (Bogacz et al., 1999)

$$P_{\max} \approx \frac{0.185}{8} N^2 \approx 0.023 N^2. \quad (9)$$

Figure 5 compares the capacities found in simulations with this theoretical prediction. In the case of familiarity discrimination, the storage capacity is dependent on the assumed acceptable probability of error. In contrast, in the case of the retrieval from the Hopfield network, there exists an independent capacity boundary—namely,  $0.145N$  (Amit, 1989). This difference comes from the fact that during the retrieval process the associative memories undergo a relaxation process, during which different attractors compete between themselves and so to calculate capacity mean-field methods are used (Amit, 1989). In contrast, the familiarity discrimination is performed simply by calculating the value of the energy function, and the capacity is obtained from the signal to noise analysis.

## Appendix B: Familiarity Discrimination by a Feed-Forward Network

The energy of the Hopfield network is an artificial function whose value is calculated by a double summation (see Eq. (2) in Appendix A). The model described in Section 2 effectively calculates a similar function also by a double summation: one summation is performed by FDNs, and the other by decision neurons. This Appendix shows how the energy function may be transformed to a form computed by a feed-forward network and what is the storage capacity for familiarity discrimination of the resulting network.

We wish to define a decision function that is positive for familiar patterns and negative for novel ones. From Appendix A, the simplest such function is given by

$$d_1(\bar{x}) = \sum_{i=1}^N x_i \sum_{j=1}^N x_j w_{ij} - \frac{N}{2}. \quad (10)$$

We now transform  $d_1$  to a form that may be computed by a neural net. By bringing the threshold term into the

sum and splitting the sum into two terms, we get

$$d_1(\bar{x}) = \sum_{\substack{i=1 \\ x_i=1}}^N \left( \sum_{j=1}^N x_j w_{ij} - \frac{1}{2} \right) + \sum_{\substack{i=1 \\ x_i=-1}}^N - \left( \sum_{j=1}^N x_j w_{ij} - \frac{1}{2} \right). \quad (11)$$

We create a decision function using only the first term of Eq. (11), as follows:

$$d_2(\bar{x}) = \sum_{\substack{i=1 \\ x_i=1}}^N \left( \sum_{j=1}^N x_j w_{ij} - \frac{1}{2} \right). \quad (12)$$

Assuming that the inner summation is computed by FDNs and the outer by decision neurons, the decision neurons sum only over the subset of FDN<sub>*i*</sub> such that  $x_i = 1$ . In the model this is implemented by strong nonmodifiable (“driving”) connections between representation neurons and corresponding FDNs (see Section 2). These connections ensure that to activate a FDN, the corresponding representation neuron must also be active, so Eq. (12) is already close to the required form.

Intuitively, dropping one term from Eq. (11) (as we did to get Eq. (12)) does not change the critical properties of this function very much (it is still positive for familiar patterns and negative for novel ones). However, the storage capacity of the network becomes two times smaller because the network “uses information from half of the neurons.” Here, we show using signal-to-noise analysis the value of the storage capacity for familiarity discrimination when the decision function  $d_2$  (Eq. (12)) is applied. The value of  $d_2$  after delivering to the network one of the stored patterns, (such as  $\xi^1$ ) is given by

$$\begin{aligned} d_2(\bar{\xi}^1) &= \sum_{\substack{i=1 \\ \xi_i^1=1}}^N \left( \sum_{j=1}^N \xi_j^1 \frac{1}{N} \sum_{\mu=1}^P \xi_i^\mu \xi_j^\mu - \frac{1}{2} \right) \\ &= \sum_{\substack{i=1 \\ \xi_i^1=1}}^N \left( \frac{1}{N} \sum_{j=1}^N \xi_j^1 \xi_i^1 \xi_j^1 \right. \\ &\quad \left. + \frac{1}{N} \sum_{j=1}^N \xi_j^1 \sum_{\mu=2}^P \xi_i^\mu \xi_j^\mu - \frac{1}{2} \right) \end{aligned}$$

$$\begin{aligned}
&= \sum_{\substack{i=1 \\ \xi_i^1=1}}^N \left( \xi_i^1 - \frac{1}{2} \right) + \frac{1}{N} \sum_{\substack{i=1 \\ \xi_i^1=1}}^N \sum_{j=1}^N \sum_{\mu=2}^P \xi_j^1 \xi_i^\mu \xi_j^\mu \\
&= \sum_{\substack{i=1 \\ \xi_i^1=1}}^N \frac{1}{2} + \frac{1}{N} \sum_{\substack{i=1 \\ \xi_i^1=1}}^N \sum_{j=1}^N \sum_{\mu=2}^P \xi_j^1 \xi_i^\mu \xi_j^\mu. \quad (13)
\end{aligned}$$

Again, the first term corresponds to the signal, and the second to noise. The signal has a binomial distribution, so it may be approximated with a normal distribution with mean  $N/4$  and variance  $N/16$ . As in Appendix A, assuming random patterns, the elements of the summation in the noise term may be treated as discrete random variables, having value  $-1$  or  $+1$  with equal probabilities. The noise term is the sum of approximately  $N^2 P/2$  such variables (on average) divided by  $N$ , so it has binomial distribution and may be approximated with normal distribution with mean 0. As with the Hopfield network (Appendix A), it could be expected that the noise has variance equal to  $P/2$ ; however, the variance of the noise is actually twice as large due to symmetry of the weights:

$$\begin{aligned}
d_2(\bar{\xi}^1) &\approx \theta\left(\frac{N}{4}, \sqrt{\frac{N}{16}}\right) + \theta(0, \sqrt{P}) \\
&= \frac{N}{4} + \theta\left(0, \sqrt{\frac{N+16P}{16}}\right). \quad (14)
\end{aligned}$$

We consider the algorithm as working well if the probability of error is less than 1%. An error occurs when  $d_2$  is negative for a familiar pattern. To calculate the maximum acceptable number of stored patterns  $P_{\max}$ , we must solve the following equation:

$$\Pr\left(\theta\left(0, \sqrt{\frac{N+16P_{\max}}{16}}\right) \leq \frac{N}{4}\right) = 0.99. \quad (15)$$

This is equivalent to

$$\Pr\left(\theta(0, 1) < \frac{N}{\sqrt{N+16P_{\max}}}\right) = 0.99. \quad (16)$$

Solving Eq. (16) with respect to  $P_{\max}$  (as in Appendix A) we obtain

$$P_{\max} \approx \frac{0.185N^2 - N}{16} \approx 0.012N^2. \quad (17)$$

Equation (17) shows the maximum number of patterns for which probability of nonrecognition error is 1%. Similarly, it can be shown that the maximum number of patterns for which the probability of false alarm error is equal to 1% is exactly the same. Therefore, in either case, Eq. (17) is a general capacity for a network implementing function  $d_2$  for familiarity discrimination with the probability of error 1%.

For large  $N$ , Eq. (12) may be approximated, almost without loss in capacity (as shown by simulations), by

$$d(\bar{x}) = \operatorname{sgn}\left(\sum_{\substack{i=1 \\ x_i=1}}^N \operatorname{sgn}\left(\sum_{j=1}^N x_j w_{ij} - \frac{1}{2}\right)\right). \quad (18)$$

Figure 5 shows that the storage capacity for familiarity discrimination of function  $d$  obtained in simulations matches closely the theoretical predictions of Eq. 17. The inner part of Eq. (18), as it calculates the sum of the weights of its active synapses relative to  $1/2$ , may be performed by a neuron; hence, let us denote it by  $y_i$ . Function  $d$  may thus be implemented by a single neuron, as follows:

$$\begin{aligned}
d(\bar{x}) &= \operatorname{sgn}\left(\sum_{\substack{i=1 \\ x_i=1}}^N y_i\right) \quad \text{where:} \\
y_i &= \operatorname{sgn}\left(\sum_{j=1}^N x_j w_{ij} - 1/2\right). \quad (19)
\end{aligned}$$

Hence, familiarity discrimination may be performed by a neural network with two processing layers, as shown in Fig. 2a ( $x$  corresponds to representation neurons,  $y$  to FDNs, and  $d$  to decision neurons). In Eq. (19), the summation is done only over the activation of those FDN $_i$  for which  $x_i = 1$ . In the network, this is implemented by driving connections.

### Appendix C: Capacity of Familiarity Discrimination Network with Diluted Connectivity

Appendix B concerns the fully connected familiarity discrimination network. This Appendix shows that the network also works with diluted connections with the capacity per synapse remaining the same.

Assume that the probability that a representation neuron is connected to an FDN is equal to  $c$  and that  $C_i$  denotes the set of representation neurons connected to FDN $_i$ . For simplicity we have assumed that driving connections are not lost, though, in reality, these

connections are easily protected by increasing their redundancy. The Hebbian definition of the weights of the FDNs from Eq. (1) becomes

$$w_{ij} = \begin{cases} \frac{1}{Nc} \sum_{\mu=1}^P \xi_i^\mu \xi_j^\mu & \text{for } i \neq j \wedge j \in C_i \\ 0 & \text{otherwise.} \end{cases} \quad (20)$$

The value of  $d_2$  (defined in Eq. (12)) after delivering to the network one of the stored patterns (such as  $\xi^1$ ) becomes

$$\begin{aligned} d_2(\bar{\xi}^1) &= \sum_{\substack{i=1 \\ \xi_i^1=1}}^N \left( \sum_{\substack{j=1 \\ j \in C_i}}^N \xi_j^1 \frac{1}{Nc} \sum_{\mu=1}^P \xi_i^\mu \xi_j^\mu - \frac{1}{2} \right) \\ &= \sum_{\substack{i=1 \\ \xi_i^1=1}}^N \left( \frac{1}{Nc} \sum_{\substack{j=1 \\ j \in C_i}}^N \xi_j^1 \xi_i^1 \xi_j^1 \right. \\ &\quad \left. + \frac{1}{Nc} \sum_{\substack{j=1 \\ j \in C_i}}^N \xi_j^1 \sum_{\mu=2}^P \xi_i^\mu \xi_j^\mu - \frac{1}{2} \right) \\ &\approx \sum_{\substack{i=1 \\ \xi_i^1=1}}^N \theta \left( \frac{1}{2}, \sqrt{\frac{1-c}{Nc}} \right) + \theta \left( 0, \sqrt{\frac{P}{c}} \right) \\ &\approx \theta \left( \frac{N}{4}, \sqrt{\frac{1-c}{2c} + \frac{N}{16}} \right) + \theta \left( 0, \sqrt{\frac{P}{c}} \right) \\ &= \frac{N}{4} + \theta \left( \frac{8(1-c) + Nc + 16P}{16c} \right). \end{aligned} \quad (21)$$

Performing calculations analogous to those of Appendix B (Eqs. (15) and (16)), we obtain the expression for the number of stored patterns for which the probability of error is 1%:

$$\begin{aligned} P_{\max} &\approx \frac{0.185N^2c - Nc - 8(1-c)}{16} \\ &\approx 0.012N^2c. \end{aligned} \quad (22)$$

Equation (22) shows that the storage capacity of the familiarity discrimination network with diluted connectivity is proportional to the number of connections between representation neurons and FDNs. Hence, the capacity per synapse is unaltered by the dilution of connections.

## Appendix D: Binary Neuron Model of the Perirhinal Network

In the model described in Section 2 and detailed here, the active state of a neuron is denoted by 1, and the inactive state by 0 (in contrast to models described in Appendices A and B, where the inactive state is denoted by  $-1$ ). Additionally, both the decision neurons and the FDNs have positive synaptic weights (denoting excitatory connections). The positive weights are compensated for by inhibition. In this Appendix, first we introduce all the changes to the model description necessary to satisfy the above assumptions, and then we show that computations performed by the network described here are equivalent to those of the network from Appendix B (Eq. (19)).

In this Appendix the variables corresponding to those from Appendix B are denoted by the same symbols but with a caret. Hence with the transformation from the  $\{-1, 1\}$  notation to the  $\{0, 1\}$  notation  $\hat{\xi}_i^\mu = \frac{1}{2}(\xi_i^\mu + 1)$  and  $\hat{x}_i = \frac{1}{2}(x_i + 1)$ . In the case of the FDNs the introduction of positive weights is achieved by initializing the weights with a positive constant  $K$  (in simulations,  $K = 5/N$ ). The constant value of the weights of the driving connections is denoted by  $R$  (in simulations,  $R = 4$ ). The Hebbian definition of the weights of the FDNs from Eq. (1) becomes (for simplicity the fully connected network is considered again)

$$\hat{w}_{ij} = \begin{cases} \frac{8}{N} \sum_{\mu=1}^P \left( \hat{\xi}_i^\mu - \frac{1}{2} \right) \left( \hat{\xi}_j^\mu - \frac{1}{2} \right) + K & \text{for } i \neq j \\ R & \text{for } i = j. \end{cases} \quad (23)$$

As a result of the driving connections, the FDNs have the same activation as that of the corresponding representation neurons ( $\hat{y}_i = \hat{x}_i = \hat{\xi}_i^\mu$ ) during the memorizing period for novel patterns. Hence, the term  $(\hat{\xi}_i^\mu - 1/2)(\hat{\xi}_j^\mu - 1/2)$  from Eq. (23) is equivalent to  $(\hat{y}_i - 1/2)(\hat{x}_i - 1/2)$  and Eq. (23) expresses a Hebbian learning rule, where synaptic weights are changed according to the activity of presynaptic and postsynaptic neurons (LTP for  $\hat{x}_i = 1, \hat{y}_i = 1$ ; heterosynaptic LTD for  $\hat{x}_i = 0, \hat{y}_i = 1$ ; homosynaptic LTD for  $\hat{x}_i = 1, \hat{y}_i = 0$ ). In the current form Eq. (23) implies that the weights should also be increased when both presynaptic and postsynaptic neurons are inactive ( $\hat{x}_i = 0, \hat{y}_i = 0$ ). However, the general form of the above mentioned term from Eq. (23) is  $(\hat{y}_i - a)(\hat{x}_i - a)$ , where  $a$  denotes sparseness of coding (that is, the

probability that a neuron is active in a given pattern). As discussed in Section 3, it is assumed that  $a = 1/2$ , but in reality  $a$  is likely to be smaller. In the general case the increase of weight when both neurons are inactive is  $8a^2/N$ , which for small  $a$  is very close to 0 and may be discarded. Thus the rule in practice approximates closely to Hebbian learning.

To make the membrane potential of each FDN independent of the number of active bits in the input pattern, each FDN receives inhibition proportional to this number:

$$\text{Inh} = \sum_{i=1}^N K \hat{x}_i. \quad (24)$$

The threshold of  $\text{FDN}_i$ , which is equal to  $1/2$  in Eq. (19), must be increased by  $R$  to ensure that only neurons receiving activation through driving connections are active. An additional term must also be added to the definition of the threshold, to effect the transformation from the  $\{-1, 1\}$  notation to the  $\{0, 1\}$  notation (as will be seen from Eq. (28)):

$$T_i = \frac{1}{2} + R + \sum_{\substack{i=1 \\ j \neq i}}^N \frac{4}{N} \sum_{\mu=1}^P \left( \hat{\xi}_i^\mu - \frac{1}{2} \right) \left( \hat{\xi}_j^\mu - \frac{1}{2} \right). \quad (25)$$

The additional term in Eq. (25) depends on the stored patterns, so it could suggest that the threshold should be modified during learning. However, this term comes from the fact that the weight modification according to Eq. (23) creates a possibility that for a particular neuron the total change of weights (summed over all synapses) is not zero. This term allows the level of excitability of an FDN to be increased (or decreased) in compensatory way. Real neurons also have a mechanism that balances their level of excitability. For example, if the neuron is very active, its excitability decreases (Kirov et al., 1999). Alternatively, the additional term in Eq. (25) may be removed if the learning rule of Eq. (23) is modified in a way ensuring that the total change of weight of a neuron is equal to 0 (that is,  $\forall i : \sum_{j=1}^N \Delta w_{ij} = 0$ ), which is also biologically plausible (Sorra and Harris, 1998).

The output of  $\text{FDN}_i$  is defined by

$$\hat{y}_i = \sigma \left( \sum_{j=1}^N \hat{w}_{ij} \hat{x}_j - \text{Inh} - T_i \right), \quad (26)$$

where  $\sigma(z)$  is equal to 1 for positive  $z$  and 0 for negative ones. The decision neuron is active if the majority of FDNs receiving excitation through driving connections

are active. The information about the total number of FDNs receiving excitation through driving connections is carried by the same inhibition, as those received by FDNs:

$$\hat{d} = \sigma \left( \sum_{i=1}^N 2K \hat{y}_i - \text{Inh} \right). \quad (27)$$

We now prove that the computation of the above network is equivalent to those of Appendix B (Eq. (19))—that is,  $\hat{d} = \frac{1}{2}(d + 1)$ . First, we consider FDNs. Due to driving connections, an FDN is inactive when the corresponding representation neuron is inactive—that is,  $\hat{x}_i = 0 \Rightarrow \hat{y}_i = 0$ . Let us prove that if a representation neuron is active, the activity of the corresponding FDN defined by Eqs. (26) and (19) are equivalent—that is,  $\hat{x}_i = 1 \Rightarrow \hat{y}_i = \frac{1}{2}(y_i + 1)$ :

$$\begin{aligned} \hat{y}_i &= \sigma \left( \sum_{j=1}^N \hat{w}_{ij} \hat{x}_j - \ln h - T_i \right) \\ &= \sigma \left( \sum_{j=1}^N \left( \frac{8}{N} \sum_{\mu=1}^P \left( \hat{\xi}_i^\mu - \frac{1}{2} \right) \left( \hat{\xi}_j^\mu - \frac{1}{2} \right) + K \right) \hat{x}_j \right. \\ &\quad \left. + R \hat{x}_i - \sum_{i=1}^N K \hat{x}_i - \frac{1}{2} - R \right. \\ &\quad \left. - \sum_{\substack{j=1 \\ j \neq i}}^N \frac{4}{N} \sum_{\mu=1}^P \left( \hat{\xi}_i^\mu - \frac{1}{2} \right) \left( \hat{\xi}_j^\mu - \frac{1}{2} \right) \right) \\ &\approx \sigma \left( \sum_{\substack{j=1 \\ j \neq i}}^N \frac{8}{N} \sum_{\mu=1}^P \frac{1}{2} \hat{\xi}_i^\mu \frac{1}{2} \hat{\xi}_j^\mu \frac{1}{2} (x_j + 1) \right. \\ &\quad \left. - \sum_{\substack{j=1 \\ j \neq i}}^N \frac{4}{N} \sum_{\mu=1}^P \frac{1}{2} \hat{\xi}_i^\mu \frac{1}{2} \hat{\xi}_j^\mu - \frac{1}{2} \right) \\ &= \sigma \left( \sum_{j=1}^N x_j \frac{1}{N} \sum_{\mu=1}^P \hat{\xi}_i^\mu \hat{\xi}_j^\mu - \frac{1}{2} \right) \\ &= \frac{1}{2} \left( \text{sgn} \left( \sum_{j=1}^N x_j w_{ij} - \frac{1}{2} \right) + 1 \right) \\ &= \frac{1}{2} (y_i + 1). \end{aligned} \quad (28)$$

The decision neuron defined by Eq. (27) is active when the number of active FDNs is higher than half of the number of active representation neurons. Since FDNs

not receiving input through driving connections are inactive, Eq. (27) is equivalent to Eq. (19), where the decision neuron is active when the majority of FDNs receiving input through driving connections are active.

### Appendix E: Spike Response Model of the Perirhinal Network

The network was implemented in a realistic spiking neuronal framework by adapting Eqs. (23) to (27) to the spike-response model (Gerstner, 1998b) with parameters similar to those used in the model of cortical neurons by Fohlmeister et al. (1995). In this Appendix, the spike-response model is briefly introduced, and its application to the simulation of the perirhinal network is described. In the spike-response model, the set of time moments in which the neuron  $i$  has fired is denoted by  $F_i$ :

$$F_i = \{t_i : y_i(t_i) = 1\}. \quad (29)$$

In the implementation, only the most recent spike times are remembered—those that are important for the computation of current membrane potential (in the implementation, the 5 most recent spikes are remembered). The membrane potential of neuron  $i$  is influenced by three sources: excitation, inhibition, and refractoriness:

$$h_i(t) = h_i^{exc}(t) + h_i^{inh}(t) + h_i^{ref}(t). \quad (30)$$

The excitation of neuron  $i$  is defined as a weighted sum of excitatory postsynaptic potentials (EPSPs) caused by all the excitatory neurons connected to neuron  $i$ :

$$h_i^{exc}(t) = \sum_{j=1}^N w_{ij} \sum_{t_j \in F_j} \varepsilon(t - t_j - \Delta). \quad (31)$$

In Eq. (31),  $\Delta$  denotes the axonal delay of excitatory neurons (in the model  $\Delta$  is taken randomly as 2 or 3), and  $\varepsilon(r)$  is the kernel function for EPSPs, which is equal to 0 for negative  $r$ , and for positive  $r$ :

$$\varepsilon(r) = R_\varepsilon \frac{r}{r_\varepsilon^2} \exp\left(-\frac{r}{r_\varepsilon}\right). \quad (32)$$

In Eq. (32),  $r_\varepsilon$  denotes the period after which the EPSP reaches its maximum (in the model  $r_\varepsilon = 2$  ms), and  $R_\varepsilon$  denotes the magnitude of the EPSP (it was chosen so that the maximum value of  $\varepsilon(r)$  is equal to 1—that is,  $R_\varepsilon = 5.5$ ). Inhibition is defined similarly to excitation as a weighted sum of inhibitory postsynaptic potentials (IPSP) caused by inhibitory neurons connected to

neuron  $i$ :

$$h_i^{inh}(t) = \sum_{j \in Inh} w_{inh} \sum_{t_j \in F_j} \varepsilon^{inh}(t - t_j - \Delta^{inh}). \quad (33)$$

As above,  $\varepsilon^{inh}$  is the inhibitory kernel, which is defined similarly to the excitatory kernel (Eq. (32)), but with different parameter values (in the model  $r_{inh} = 6$  ms,  $R_{inh} = 16.3$ ). The refractoriness of neuron  $i$  is defined as

$$h_i^{ref}(t) = \sum_{t_i \in F_i} \eta(t - t_i). \quad (34)$$

In Eq. (34)  $\eta(r)$  is a kernel function of refraction, which is defined as

$$\eta(r) = \begin{cases} -\infty & \text{for } r \leq \delta^{abs} \\ -R_{ref} \exp\left(-\frac{r - \delta^{abs}}{r_{ref}}\right) & \text{for } r > \delta^{abs}. \end{cases} \quad (35)$$

The refraction in a period just after a spike has very strong negative value, which blocks the neuron from firing (this period is called the *absolute refractory period*). Subsequently, the refraction converges to 0. In the model the following parameters were taken:  $r_{ref} = 4$  ms,  $R_{ref} = 16$ ,  $\delta^{abs} = 1$  ms.

The behavior of the neuron is stochastic, and the probability of firing depends on the value of the membrane potential according to the following:

$$\Pr(y = 1) = \frac{1}{2}(1 + \tanh \beta(h_i(t) - T_i)). \quad (36)$$

In Eq. (36), the parameter  $\beta$  determines the randomness of the neuron ( $\beta = 25$ ).

The representation neurons are activated in a way that simulates observed behavior in the perirhinal cortex (Fig. 1c). In particular, when the stimulus is not present, they fire with a very low frequency; when a stimulus is present, some of them (belonging to the assembly representing the stimulus) start to fire within  $\sim 2$  ms of each other with higher frequency (about 35 Hz). The intervals between spikes have an approximate Poisson distribution (Schadlen and Newsome, 1998).

FDNs behave according to the presented spike-response model. The threshold in Eq. (36) is defined by Eq. (25), and the weights in Eq. (31) are defined by Eq. (23). Since all the inhibition neurons denoted “Inhibitory” in Fig. 2a have the same function, for simplicity they are not implemented. Instead, each FDN

receives inhibition proportional to its estimated population activity (as in Fohlmeister et al., 1995). Hence, the inhibition received by FDNs is defined by Eq. (33), where the set  $Inh$  describing inhibitory neurons is equal to the set of all representation neurons (analogously to Fohlmeister et al., 1995) and  $w_{inh}$  is equal to  $K$  during the familiarity discrimination period. This ensures that the FDNs receive inhibition analogous to that defined in Eq. (24). In the memorizing period,  $w_{inh}$  is modified according to the activity of the decision neurons. If the decision neuron is not active in the familiarity discrimination period,  $w_{inh}$  decreases by 50% during the memorizing period. In the brain such an effect can be produced by GABA<sub>B</sub> presynaptic autoinhibition (Mott et al., 1993). GABA<sub>B</sub> autoinhibition has been observed in the perirhinal cortex (Ziakopoulos et al., 2000). This reduction in inhibition results in high-frequency activity of all FDNs receiving input through driving connections. If the decision neurons are active in the familiarity discrimination period,  $w_{inh}$  increases by 50% during the memorizing period (due to a triggering input from the decision neurons) (see Fig. 2a) and blocks the high-frequency activity of FDNs.

The behavior of the decision neurons is also given by the spike response model where the weight in Eq. (31) is equal to  $2K$  and the decision neurons have two sources of inhibition. The first one is the same as the inhibition received by the FDNs, and it ensures that the decision neurons behave as defined in Eq. (27). The second source comes from inhibitory neurons denoted by “Slow inh.” in Fig. 2a, which block the activity of the decision neurons during the memorizing period. Again, these neurons have not been implemented individually, but the decision neurons receive inhibition according to their estimated population activity. This inhibition is defined by Eq. (33), where the set  $Inh$  describing inhibitory neurons is equal to the set of all representation neurons and  $w_{inh}$  is equal to  $K$ . Additionally, a longer axonal delay of  $\Delta = 6$  ms is introduced to model the information processing by these inhibitory neurons (as in Fohlmeister et al., 1995).

All software is available from <http://www.cs.bris.ac.uk/~bogacz/bbc>.

## Acknowledgments

This work is supported in part by an ORS grant held by the first author and Wellcome Trust and MRC grants held by the second.

## References

- Aggleton JP, Brown MW (1999) Episodic memory, amnesia and the hippocampal-anterior thalamic axis. *Behavioral Brain Sci.* 22:425–498.
- Aggleton JP, Shaw C (1996) Amnesia and recognition memory: A re-analysis of psychometric data. *Neuropsychol.* 34:51–62.
- Amit DJ (1989) *Modeling Brain Function*. Cambridge University Press, Cambridge.
- Bancaud J, Brunet-Bourgin F, Chauvel P, Halgren E (1994) Anatomical origin of déjà vu and vivid “memories” in human temporal lobe epilepsy. *Brain* 117:71–90.
- Bliss TVP, Collingridge GL (1993) A synaptic model of memory: Long-term potentiation in hippocampus. *Nature* 361:31–39.
- Bogacz R, Brown MW, Giraud-Carrier C (1999) High capacity neural networks for familiarity discrimination. In: *Proceedings of International Conference on Artificial Neural Networks '99, Edinburgh*. Institution of Electrical Engineers, London, pp. 773–776.
- Bogacz R, Brown MW, Giraud-Carrier C (in press) Model of co-operation between recency, familiarity and novelty neurons in the perirhinal cortex. In: Bower J, ed. *Computational Neuroscience: Trends in Research* Elsevier, New York.
- Borisjuk RM, Denham MJ, Hoppensteadt, FC (1999) An oscillatory model of novelty detection in hippocampus (Abstract). In: *Proceedings of the Third International Conference on Cognitive and Neural System*, Boston University Press, Boston.
- Brown MW, Xiang JZ (1998) Recognition memory: Neuronal substrates of the judgement of prior occurrence. *Prog. Neurobiol.* 55:149–189.
- Cassell M (1980) The number of cells in the striatum pyramidal of the rat and human hippocampal formation. Ph.D. thesis, University of Bristol.
- Cho K, Kemp N, Noel J, Aggleton JP, Brown MW, Bashir ZI (2000) A new form of long-term depression in the perirhinal cortex. *Nature Neurosci.* 3:150–156.
- Cover TM (1965) Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Trans. Elec. Computers* 14:326–334.
- Cragg BG (1975) The density of synapses and neurons in normal, mentally defective and ageing human brains. *Brain* 98:81–90.
- Fohlmeister C, Gerstner W, Ritz R, van Hemmen JL (1995) Spontaneous excitation in the visual cortex: Stripes, spirals, rings and collective bursts. *Neural Comput.* 7:905–914.
- Foldiak P, Young M (1995) Sparse coding in the primate cortex. In: Arbib MA, ed. *Handbook of Brain Theory and Neural Networks*. MIT Press, Cambridge, MA.
- Gerstner W (1998a) Populations of spiking neurons. In: Maas W, Bishop C, eds. *Pulsed Neural Networks*. MIT Press, Cambridge, MA.
- Gerstner W (1998b) Spiking neurons. In: Maas W, Bishop C, eds. *Pulsed Neural Networks*. MIT Press, Cambridge, MA.
- Granger E, Grossberg S, Rubin MA, Streilein WW (1998) Familiarity discrimination of radar pulses. *Adv. Neural Info. Proc. Sys.* 11:875–881.
- Hertz J, Krogh A, Palmer RG (1991) *Introduction to the Theory of Neural Computations*. Addison-Wesley, Redwood City, CA.
- Hintzman DL, Caulton DA, Levitin DJ (1998) Retrieval dynamics in recognition and list discrimination: Further evidence of

- separate processes of familiarity and recall. *Memory and Cognition* 26:449–462.
- Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci.* 79:2554–2558.
- Insausti R, Juottonen K, Soininen H, Insausti AM, Partanen K, Vainio P, Laakso MP, Pitkanen A (1998) MR volumetric analysis of the human entorhinal, perirhinal and temporopolar cortices. *Am. J. Neuroradiol.* 19:659–671.
- Ito M (1989) Long-term depression. *Ann. Rev. Neurosci.* 12:85–102.
- Kazer JF, Sharkey AJC (1999) The septo-hippocampal system and anxiety: A robot simulation. In: *Proceedings of the International Conference on Artificial Neural Networks '99, Edinburgh*. Institution of Electrical Engineers, London, pp. 389–394.
- Kirov SA, Sorra KE, Harris KM (1999) Slices have more synapses than perfusion-fixed hippocampus from both young and mature rats. *J. Neurosci.* 19:2876–2886.
- Kobatake E, Wang G, Tanaka K (1998) Effects of shape discrimination training on the selectivity of inferotemporal cells in adult monkeys. *J. Neurophysiol.* 80:324–330.
- Kohonen T (1989) *Self-organisation and Associative Memory*. (3rd ed.). Springer-Verlag, Heidelberg.
- Kohonen T, Oja E, Ruohonen M (1974) Adaptation of a linear system to a finite set of patterns occurring in an arbitrarily varying order. *Acta Polytech. Scand. Elect. Eng.* 25.
- Kowalczyk A (1997) Estimates of storage capacity of multilayer perceptron with threshold logic hidden units. *Neural Networks* 10:1417–1433.
- Li L, Miller EK, Desimone R (1993) The representation of stimulus familiarity in anterior inferior temporal cortex. *J. Neurophysiol.* 69:1918–1929.
- Mandler G (1980) Recognising: The judgement of previous occurrence. *Psychol. Rev.* 87:252–271.
- Markram H, Tsodyks M (1996) Redistribution of synaptic efficacy between neocortical pyramidal neurons. *Nature* 382:807–810.
- Marr D (1970) A theory for cerebral neocortex. *Proc. Roy. Soc. Lond. B* 176:161–234.
- Marr D (1971) Simple memory: A theory of archicortex. *Phil. Trans. Roy. Soc. B* 262:23–81.
- McCulloch WS, Pitts W (1943) A logical calculus of ideas immanent in nervous activity. *Bull. Math. Biophys.* 5:115–133.
- Miller EK, Desimone R (1993) Scopolamine effects short-term memory but not inferior temporal neurons. *NeuroReport* 4:81–84.
- Mott DD, Xie C, Wilson WA, Swartzwelder HS, Lewis DV (1993) GABA<sub>B</sub> autoreceptors mediate activity-dependent disinhibition and enhance signal transmission in the dentate gyrus. *J. Neurophysiol.* 69:674.
- Murray EA (1996) What have ablation studies told us about the neural substrates of stimulus memory? *Seminars in Neurosci.* 8:13–22.
- Murray EA, Bussey TJ (1999) Perceptual-mnemonic functions of the perirhinal cortex. *Trends in Cog. Sci.* 3:142–151.
- Olshausen BA, Field DJ (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381:607–609.
- Rieke F, Warland D, de Ruyter van Steveninck R, Bialek W (1997) *Spikes: Exploring the Neural Code*. MIT Press, Cambridge, MA.
- Ringo JL (1995) Brevity of processing in a mnemonic task. *J. Neurophysiol.* 73:1712–1715.
- Roberts S, Tarassenko L (1995) A probabilistic resource allocating networks for novelty detection. *Neural Comput.* 6:270–284.
- Roediger HL III, McDermott KB (1995) Creating false memories: Remembering words not present in list. *J. Exper. Psychol: Learning, Memory and Cognition* 21:803–814.
- Rolls ET (1996) A theory of hippocampal function in memory. *Hippocampus* 6:601–620.
- Rolls ET, Tovee MJ (1994) Processing speed in the cerebral cortex and neurophysiology of visual masking. *Proc. Roy. Soc. Lond. B* 257:9–15.
- Saksida LM, Bussey TS (1998) Towards a neural network model of visual object identification in primate infero-temporal cortex. *Soc. Neurosci. Abst.* 24:1906.
- Schacter DL (1999) The seven sins of memory: Insight from psychology and cognitive neuroscience. *Am. Psychologist* 54:182–203.
- Schacter DL, Reiman E, Curran T, Yun L, Bandy D, McDermott KB, Roediger HL III (1996) Neuroanatomical correlates of veridical and illusory memories: Evidence from positron emission tomography. *Neuron* 17:267–274.
- Schacter DL, Verfaellie M, Anes MD, Racine CA (1998) When the true recognition suppresses false recognition: Evidence from amnesic patients. *J. Cog. Neurosci.* 10:668–679.
- Schadlen MN, Newsome WT (1998) The variable discharge of cortical neurons: Implications for connectivity, computations and information coding. *J. Neurosci.* 18:3870–3896.
- Seeck M, Michael CM, Mainwaring N, Cosgrove R, Blume H, Ives J, Landis T, Schomer DL (1997) Evidence for rapid face recognition from human scalp and intracranial electrodes. *NeuroReport* 8:2749–2754.
- Sorra KE, Harris KM (1998) Stability in synapse number and size at 2 hr after long-term potentiation in hippocampal area CA1. *J. Neurosci.* 18:658–671.
- Standing L (1973) Learning 10,000 pictures. *Q. J. Exper. Psychol.* 25:207–222.
- Suzuki WA (1996) The anatomy, physiology and functions of the perirhinal cortex. *Curr. Opinion in Neurobiol.* 6:172–186.
- Xiang JZ, Brown MW (1997) Processing visual familiarity and recency information: Neuronal interactions in area TE and rhinal cortex (Abstract). *Brain Res. Abs.* 14:69.
- Xiang JZ, Brown MW (1998) Differential neuronal encoding of novelty, familiarity and recency in regions of the anterior temporal lobe. *Neuropharmacol.* 37:657–676.
- Ziakopoulos Z, Tillett CW, Brown MW, Bashir ZI (1999) Input- and layer-dependent synaptic plasticity in the rat perirhinal cortex in vitro. *Neurosci.* 92:459–472.
- Ziakopoulos Z, Brown MW, Bashir ZI (2000) GABAB receptors mediate frequency dependent depression of excitatory potentials in rat perirhinal cortex in Vitro. *Eur. J. Neurosci.* 12:803–809.