INSTITUTE OF PHYSICS PUBLISHING NETWORK: COMPUTATION IN NEURAL SYSTEMS

Network: Comput. Neural Syst. **13** (2002) 457–485 PII: S0954-898X(02)52832-3

The restricted influence of sparseness of coding on the capacity of familiarity discrimination networks

Rafal Bogacz[1,3] **and Malcolm W Brown**[2,4]

[1] Department of Computer Science, University of Bristol, Bristol BS8 1UB, UK
[2] MRC Centre for Synaptic Plasticity, Department of Anatomy, University of Bristol, Bristol BS8 1TD, UK

E-mail: M.W.Brown@bristol.ac.uk

Received 5 January 2002, in final form 2 August 2002
Published 19 September 2002
Online at stacks.iop.org/Network/13/457

Abstract

Much evidence indicates that the perirhinal cortex is involved in the familiarity discrimination aspect of recognition memory. It has been previously shown under selective conditions that neural networks performing familiarity discrimination can achieve very high storage capacity, being able to deal with many times more stimuli than associative memory networks can in associative recall. The capacity of associative memories for recall has been shown to be highly dependent on the sparseness of coding. However, previous work on the networks of Bogacz *et al*, Norman and O'Reilly and Sohal and Hasselmo that model familiarity discrimination in the perirhinal cortex has not investigated the effects of the sparseness of encoding on capacity. This paper explores how sparseness of coding influences the capacity of each of these published models and establishes that sparse coding influences the capacity of the different models in different ways. The capacity of the Bogacz *et al* model can be made independent of the sparseness of coding. Capacity increases as coding becomes sparser for a simplified version of the neocortical part of the Norman and O'Reilly model, whereas capacity decreases as coding becomes sparser for a simplified version of the Sohal and Hasselmo model. Thus in general, and in contrast to associative memory networks, sparse encoding results in little or no advantage for the capacity of familiarity discrimination networks. Hence it may be less important for coding to be sparse in the perirhinal cortex than it is in the hippocampus. Additionally, it is established that the capacities of the networks are strongly dependent on the precise form of the learning rules (synaptic plasticity) used in the network. This finding indicates that the precise characteristics of synaptic plastic changes in the real brain are likely to have major influences on storage capacity.

[3] Present address: Department of Applied and Computational Mathematics, Princeton University, Princeton, NJ 08544, USA
[4] Author to whom any correspondence should be addressed.

0954-898X/02/040457+29$30.00

1. Introduction

Work in monkeys has established that discrimination of the relative familiarity or novelty of visual stimuli (i.e. determining whether a stimulus has been previously encountered or not) is dependent on the perirhinal cortex, and this finding is consistent with studies of amnesic patients (Eichenbaum *et al* 1994, Aggleton and Shaw 1996, Murray 1996, Brown and Xiang 1998, Buffalo *et al* 1998, Murray and Bussey 1999, Aggleton and Brown 1999, Brown and Aggleton 2001). Thus damage to the perirhinal cortex results in impairments in recognition memory tasks that rely on discrimination of the relative familiarity of objects (Murray 1996, Brown and Aggleton 2001). Moreover, 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 *et al* 1987, Riches *et al* 1991, Fahy *et al* 1993, Li *et al* 1993, Miller *et al* 1993, Sobotka and Ringo 1993, Brown and Xiang 1998, Xiang and Brown 1998).

Storage capacity for familiarity discrimination is defined as the number of presented stimuli for which a network can discriminate familiarity with an accuracy of 99%. Bogacz *et al* (1999, 2001) showed that neural networks performing familiarity discrimination can achieve very high storage capacity, much larger than similarly sized associative memory networks achieve for associative recall. The reason for this difference in capacities is that the familiarity discrimination networks need only to provide a binary answer (novel or familiar—one bit of information), while the associative memories have to recall the whole pattern of neuronal activity representing the associated stimulus (many bits of information). If the perirhinal cortex worked akin to the model of Bogacz *et al* (2001), it alone could 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.

It has been previously shown that the sparseness of coding is a major factor determining the capacity of associative memory networks. The sparseness of coding is the proportion of neurons active after presentation of a single visual stimulus. Sparse representation greatly increases the capacity of associative memories (Buhmann *et al* 1989, Amit 1989, Palm and Sommer 1992), and this has been suggested to be a reason for expecting sparseness of representation in the hippocampus (Marr 1971, Barnes *et al* 1990, Treves and Rolls 1994). Therefore, it might be expected that sparseness of coding would also influence the capacity of familiarity discrimination networks.

However, previous calculations of the capacity of familiarity discrimination networks have not systematically investigated the influence of sparseness of coding on capacity. Capacity calculations for the Bogacz *et al* (1999, 2001) model were performed under the simplifying assumption that there was an equal number of active and inactive neurons after presentation of each visual stimulus. Two other models of familiarity discrimination in the perirhinal cortex (Norman and O'Reilly 2001, Sohal and Hasselmo 2000) have also been proposed, but although these models have been simulated, their capacity has not been established for either sparse or non-sparse coding.

This paper establishes how sparseness of coding influences the capacity of each of the published models of familiarity discrimination in the perirhinal cortex under the simplifying assumption that the responses of the neurons providing input to the network are uncorrelated.

It is shown that the sparseness of coding influences the capacity of the different familiarity discrimination networks in different ways. It is shown that the influence of the sparseness of coding upon capacity depends also on the precise form of the learning rules and decision

functions. Therefore, the types of learning rule for synaptic weight modification that allow the different models to achieve high storage capacity are also analysed.

In the next section 2 all the previously proposed models of familiarity discrimination in the perirhinal cortex, and one new model based on anti-Hebbian learning, are introduced and reviewed. Section 3 analyses the influence of the sparseness of coding on capacity of each of the models. Section 4 discusses the consequences of these results.

The derivations of capacity are contained in the appendices. Accordingly, the main text of the paper contains only the mathematical description of the models and the results of capacity calculations. The paper focuses on modelling computations performed by 'novelty' neurons. These neurons comprise $\sim$10% of perirhinal neurons that respond strongly to the first presentations of novel stimuli but only briefly or weakly to presentations of familiar stimuli (Xiang and Brown 1998). This paper is not concerned with comparing the biological plausibility of the models, nor the efficiency of the models in the case when responses of neurons providing input to the networks are correlated: an analysis of these issues may be found elsewhere (Bogacz 2001, Bogacz and Brown 2002).

2. Description of the models

This section provides a review of all the published networks for familiarity discrimination in the perirhinal cortex and describes one new network. For ease of explanation and mathematical analysis, the networks are introduced using a simple model of neurons (similar to that of McCulloch and Pitts (1943)). This model does not consider changes of neurons' membrane potentials in time. We assume that each visual stimulus is represented by a specific pattern of activity of the neurons providing input to the familiarity discrimination network. We further assume that the neurons providing input to the network may be in one of two states: active or inactive. For example, after presentation of a visual stimulus the active state of an input neuron corresponds to an increase in its activity, i.e. a response, the inactive state to no increase, i.e. to no response. However, as demonstrated previously (Bogacz *et al* 2001), it is possible to extend a model of a perirhinal network based on binary neurons to a model based on more realistic spiking neurons (Gerstner 1998) with the operational principles, capacity and efficiency remaining essentially unchanged.

All the models reviewed here have a similar architecture: the essential element of each model is a layer of novelty neurons which receive projections from input neurons. Synaptic weight modifications result in the novelty neurons having on average lower activity for familiar than for novel stimuli. Hence the familiarity of a stimulus is represented by and can be measured from the distribution of the responses of the novelty neurons. The models differ in their assumed types of synaptic plasticity and in details of their operation. Each will be now described.

2.1. Hebbian model

The model of Bogacz *et al* (1999, 2001) assumes that a proportion of perirhinal neurons form a network specialized just for familiarity discrimination, and will be here called the Hebbian model. It is first introduced intuitively and then its formal description is given.

2.1.1. Introduction of the Hebbian model. A proportion of neurons in the perirhinal cortex have weaker responses after presentation of familiar stimuli than novel stimuli (Brown *et al* 1987, Xiang and Brown 1998, Li *et al* 1993, Sobotka and Ringo 1993). A number of synaptic and network mechanisms may underlie this decrease of response; figure 1(a) shows the mechanism employed in the 'Hebbian model' (Bogacz *et al* 1999, 2001) based on Hebbian

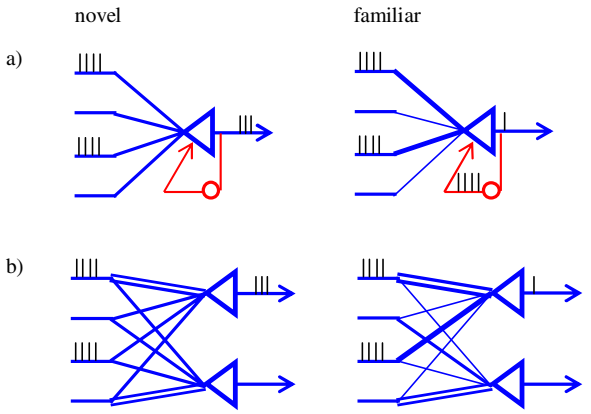

Figure 1. The Hebbian model. In each panel, the triangle represents an excitatory novelty neuron (Xiang and Brown 1998) and the circle represents an inhibitory interneuron. Lines on the left side of each panel denote inputs to the network, which are axons of neurons whose activity encodes visual stimuli. 'Spikes' over the lines indicate that the corresponding neuron is active, a lack of spikes that it is inactive. The thickness of the lines indicates the strength of the synaptic connections. The left column of panels illustrates synaptic weights and neuronal responses for a novel stimulus, and the right column of panels when this stimulus is presented again (i.e. for a familiar stimulus). (a) Synaptic plasticity in a single neuron. (b) Synaptic weight modification in the network. For simplicity the inhibitory neurons (mentioned in the text and shown in panel (a) are not shown in (b)). After presentation of a novel stimulus, the number of active novelty neurons is limited (only the upper one is active), for example by connections with high synaptic weights (denoted by double lines). The synaptic weights of the active novelty neurons are modified as in panel (a), while the weights of the inactive neurons are modified in the opposite way, e.g. the synaptic weight from the active input to the inactive novelty neuron is decreased as if by homo-synaptic LTD.

synaptic plasticity. After presentation of a novel stimulus, synaptic weights from active inputs are increased as if by long-term potentiation (LTP) (Bliss and Collingridge 1993), while weights from inactive units are decreased as if by hetero-synaptic long-term depression (LTD) (Ito 1989, Kemp and Bashir 2001). These changes produce an initially higher response of novelty neurons for familiar stimuli than for novel. However, in the network, the novelty neurons project to inhibitory neurons; this results in a higher level of inhibition for familiar than for novel stimuli, and the increased inhibition results in a smaller neuronal response for familiar stimuli than for novel (see figure 1(a)).

In the Hebbian model, the response of novelty neurons is lower for familiar stimuli due to inhibition. However, this model requires that the activity of novelty neurons should be higher for familiar stimuli in the brief initial interval before the response is suppressed by inhibition. Nevertheless, simulations (Bogacz *et al* 2001, Bogacz 2001) show that this interval may be very brief (e.g. 10 ms) and, due to temporal jitter, the increase in firing rate for familiar stimuli is not readily visible in peristimulus time histograms of simulated neuronal responses. For the same reason no increase could be expected to be (and is not) observable in the peristimulus-time histograms of the responses of real perirhinal neurones.

If each novelty neuron makes its own decision about stimulus familiarity, the overall response ('answer') of the network is encoded in the population activity of the novelty neurons. It is necessary to ensure that individual novelty neurons remain independent assessors of familiarity if the information storage capacity of the network is to be maximized (Bogacz *et al* 2001). Otherwise, should all the novelty neurons be active after the presentation of each of a series of novel stimuli, then the synaptic weights of each of the novelty neurons

would be modified in the same way, and hence all the novelty neurons would come to have highly correlated weights. Thus, eventually, they would all be active or inactive together and the whole network would have the same capacity as a single novelty neuron. To avoid this problem, the number of novelty neurons active for any one stimulus must be limited, i.e. only a subset of novelty neurons must respond to any given stimulus.

There are at least two means of limiting the number of active novelty neurons. The first means is inhibitory competition: only the fraction of neurons with the highest membrane potentials are selected to be active, the activity of the remainder being suppressed by inhibition, and only these most active neurons have their weights modified. This method of limiting the number of active novelty neurons is used in the Norman and O'Reilly (2001) model and will be described in section 2.2.

The second method of ensuring this selectivity of response of the novelty neurons is to provide specific connections with high synaptic weights from the network inputs to subsets of novelty neurons. Although this method requires the additional assumption of the existence of specialized connections (and hence may seem less plausible), it makes mathematical analysis of network behaviour simpler. Therefore, the Hebbian model as analysed in this paper assumes that the number of active neurons is limited by specific connections with high synaptic weights. As the more plausible models which limit the number of active novelty neurons by competition are much more difficult to analyse, for such models only approximate expressions for capacity may be found mathematically. However, many of the properties which may be proved mathematically for the Hebbian model with strong connections are also valid for other familiarity discrimination networks (as will be shown in simulations in section 3).

When a network rather than a single neuron is considered in the Hebbian model described by Bogacz *et al* (2001), another synaptic change is introduced: the weights of connections between active inputs and inactive novelty neurons are reduced as if, for example, by homosynaptic LTD; see figure 1(b). This decrease is required for the decision of the network to be given by its total activity rather than a more complex function (this will be discussed in detail in section 3.2).

2.1.2. Description of the Hebbian model. This section gives the mathematical description of the Hebbian model whose principles were introduced above. The description of the Hebbian model differs in detail from the original description given by Bogacz *et al* (2001), where sparse coding was not assumed. The original equations have been adapted to take into consideration sparse coding. Thus a constant representing the sparseness of coding has been introduced into the equations in a way similar to that in which the equations for associative memories are adapted when sparse coding is used (Amit 1989). The notation in this paper is similar to that used in previous work on auto-associative memories (Amit 1989). Let the active state of an input neuron be denoted by 1 and the inactive state by 0. Assume that the network consists of N novelty neurons, receiving information from N input neurons whose activity pattern represents a visual stimulus. For simplicity assume that each novelty neuron is connected to all the input neurons and denote the strength of the synaptic connection between input neuron j and novelty neuron i by w_{ij}. Denote the activity of input neuron j by x_j, and define the membrane potential of novelty neuron i as

$$h_i = \sum_{j=1}^{N} w_{ij} x_j. \tag{2.1}$$

The number of active novelty neurons in the Hebbian model must be limited (see section 2.1.1), so assume for simplicity of calculation that novelty neuron i may be active only if input neuron i is also active. This would correspond to the existence of strong non-modifiable connections

between a novelty neuron and a corresponding input, such that input through this connection is necessary for the novelty neuron to be active. This assumption is made for simplicity of notation and to allow mathematical analysis of the properties of the model. Limiting the number of active novelty neurons may be alternatively achieved by strong connections between groups of neurons rather than between pairs of neurons, or by competition (Bogacz *et al* 2001).

The strong connections ensure that the initial network response in the Hebbian model is equal to the response (proportional to the membrane potential) of neurons receiving input through these strong connections (i.e. $x_i = 1$) (Bogacz *et al* 2001):

$$d(x) = \sum_{i=1}^{N} x_i h'_i = \sum_{\substack{i,j=1 \\ i \neq j}}^{N} x_i w_{ij} x_j. \tag{2.2}$$

Thus $d(x)$ is a dot product of the input pattern and a vector of membrane potentials. The detailed explanation of how such a function may be calculated by a biologically plausible neural network can be found in Bogacz (2001). In equation (2.2), h'_i denotes the membrane potential of novelty neuron i as a result of activity in all connections except the strong connection w_{ii} (the strong connections are assumed to be non-modifiable, hence their weights do not encode occurrences of the stimuli):

$$h'_i = \sum_{\substack{j=1 \\ j \neq i}}^{N} w_{ij} x_j. \tag{2.3}$$

Due to the Hebbian weight modifications produced by previous occurrences, d is higher for familiar patterns than for novel. In the Hebbian model, d regulates the level of inhibition and hence population activity in the network after a brief period of greater excitatory spiking, and the familiarity of stimuli may be discriminated reliably by evaluating d. Function d is called a decision function in the remainder of the paper.

According to equation (2.2) the decision about stimulus familiarity is based on the activity of the most active neurons. This type of decision function is called 'act_win' by Norman and O'Reilly (2001) and this name will be also used in the remainder of this paper for similar decision functions.

Denote the number of presented stimuli (previously stored patterns) by P and the activity of input neuron j after presentation of stimulus μ by x_j^μ, so stimulus μ is represented by the pattern of activity of the input neurons given by vector x^μ. Denote the sparseness of coding by a, and assume for simplicity of calculations that it is equal for each pattern, i.e.

$$\sum_{j=1}^{N} x_j^\mu = aN. \tag{2.4}$$

Later in section 3.1, it is shown how this assumption may be relaxed. In our analysis we will also assume that $a \leqslant 0.5$.

Let us denote the pattern of activity of the novelty neurons after presentation of stimulus μ by y^μ. Let us assume that y_i^μ is equal to 1 when the activity of novelty neuron i is positive and equal to 0 when neuron i is inactive. Hence, y_i^μ is equal to 1 only if novelty neuron i receives activation through its strong connection, thus $y^\mu = x^\mu$.

The Hebbian weight modifications (figure 1) after presentation of stimulus μ may be expressed by modifying every synaptic weight by the following term (for $i \neq j$; as for associative memories; Amit, 1989):

$$\Delta w_{ij} = (y_i^\mu - a)(x_j^\mu - a) = (x_i^\mu - a)(x_j^\mu - a). \tag{2.5}$$

In the rule of equation (2.5), the weights of all the neurons are modified, hence let us call this type of weight modification rule 'mod_all'. In the rule of equation (2.5), synaptic weights are changed according to the activity of the pre-synaptic input neuron x_j^μ and the post-synaptic novelty neuron y_i^μ: LTP for $y_i^\mu = 1$, $x_j^\mu = 1$; hetero-synaptic LTD for $y_i^\mu = 1$, $x_j^\mu = 0$; homo-synaptic LTD for $y_i^\mu = 0$, $x_j^\mu = 1$; practically no change (for small a) for $y_i^\mu = 0$, $x_j^\mu = 0$.

The learning rule of the Hebbian model includes a small increase in the weights between inactive neurons ($y_i^\mu = 0, x_j^\mu = 0$). Although there is no obvious known biological mechanism that could produce such a change of weights between inactive neurons, this change is introduced to the model in order to make the decision function simple. Hence the change of weights between inactive neurons has been introduced to simplify the model, in the same way as for other memory models (Hopfield 1982, Amit 1989). The change of weights between inactive neurons is very small for sparse representations. For example, for $a = 0.1$, this change is 81 times smaller than the change between active neurons. Moreover, it will be shown in section 3.2 that this change is not a critical element of the Hebbian model. It is no longer needed if a more complex decision function is used.

Note that since we assumed that the sparseness of coding a is always constant for all patterns, the parameter a in the weight updating rule is fixed and hence the rule is locally determinable (i.e. all the terms in the weight updating rule are determinable at the individual synapses). Since the Hebbian model assumes that the neurons providing input to the novelty neurons are excitatory, the weights between input and novelty neurons should be positive, but let us assume for simplicity of analysis that all these synaptic weights of novelty neurons are initialized to 0 (a more realistic modification of the Hebbian model with positive w_{ij}, i.e. initialized to a positive constant, is described by Bogacz *et al* (2001)). The weights after presentation of P stimuli are equal to (for $i \neq j$)

$$w_{ij} = \frac{1}{Na^2(1-a)^2} \sum_{\mu=1}^{P} ((y_i^\mu - a)(x_j^\mu - a) - \varepsilon)$$

$$= \frac{1}{Na^2(1-a)^2} \sum_{\mu=1}^{P} ((x_i^\mu - a)(x_j^\mu - a) - \varepsilon). \tag{2.6}$$

Although the term preceding the summation in equation (2.6) contains a, the term is just a multiplicative constant introduced to simplify the derivation of capacity (similar to the constant used for associative memories; Amit (1989)).

The term ε inside the summation is a constant added to the weights during each modification to keep the average strength of synaptic weights constant. Since it was assumed that in every pattern representing a visual stimulus, there are exactly aN bits not equal to 0, there is a very small negative covariance, equal to $\varepsilon = -a(1-a)/(N-1)$, between the activities of the inputs. If the assumption of equation (2.4) were not made, then x_i^μ and x_j^μ would be independent and the average value of Δw_{ij} would be equal to 0. Thus subtracting ε is necessary only when the assumption of a constant number of active inputs is made. Hence it is not a critical element of the Hebbian model. This constant is very small (and may be shown to be too small to have any effect on capacity) and will be discarded in all derivations (the average value of Δw_{ij} is derived in Bogacz (2001)). In the other models described in later sections, the activities of input (x) and novelty (y) neurons are statistically independent so that the constant ε does not need to be introduced.

It is shown in appendix A.1 that in the Hebbian model the average value of the decision function d (equation (2.2); d corresponds to the initial network response) is equal to N for familiar stimuli and to 0 for novel stimuli. Therefore, by taking as the threshold the middle

value $N/2$, we can define a familiarity discrimination criterion; namely, if $d > N/2$ then the stimulus is considered familiar, otherwise it is novel.

2.2. *Combined competitive model*

Li *et al* (1993) suggested that the reduction of the number of perirhinal neurons active after presentation of familiar compared to novel stimuli was caused by the learning of a sparse representation of the stimuli. After presentation of a novel stimulus, synaptic weights are modified such that neurons that do not represent features of the stimulus very well will not be active during subsequent presentations of the stimulus. Thus during the process a more precise and sparse representation of a familiar stimulus is formed (Li *et al* 1993).

The above idea is implemented in models of Norman and O'Reilly (2001) and Sohal and Hasselmo (2000)—the essential mechanisms underlying familiarity discrimination in these two models are described in this and the next section.

Norman and O'Reilly (2001) proposed a detailed model of hippocampal and neocortical contributions to human recognition memory, explaining many psychological observations. Here we analyse a simplified network derived from the neocortical part of the Norman and O'Reilly (2001) model; we call it the combined competitive model. The combined competitive model is similar to the Hebbian model (figure 1), except for three features.

First, the limitation of the number of active novelty neurons is achieved not by special strong connections, but by inhibition and competition: the active novelty neurons are those which have the highest membrane potentials. In the original Norman and O'Reilly (2001) model, the competition and inhibition were simulated explicitly. However, for simplicity during simulations described here, the membrane potentials of the novelty neurons are evaluated after delivery of a pattern (according to equation (2.1)) and the exactly aN neurons with the highest membrane potentials are selected to be active by the simulator program. The pattern of activity of the novelty neurons after presentation of a stimulus μ is denoted by y^μ (i.e. $y_i^\mu = 1$ if neuron i belongs to the group of the neurons with the highest membrane potential; otherwise $y_i^\mu = 0$).

Second, only the weights of active novelty neurons are updated after presentation of a novel stimulus, i.e. there is no homo-synaptic LTD as illustrated in figure 1(b) for the Hebbian model. The weights of the novelty neurons are updated according to a rule developed from Norman and O'Reilly (2001) and given by

$$\Delta w_{ij} = \frac{\eta}{Na(1-a)} y_i^\mu (x_j^\mu - a). \tag{2.7}$$

In equation (2.7), η denotes the learning rate—a parameter determining the magnitude of weight modification. Its optimal value depends on N and a (so for a given network η may be fixed, e.g. encoded genetically, and hence the learning rule can be locally determinable). The values of η that resulted in the highest capacity in simulations in figure 7 comparing the efficiency of the models ranged from 0.15 to 0.6 (detailed values are given in Bogacz (2001)). The term η was not introduced explicitly when describing the Hebbian model, but comparing equations (2.6) and (2.7) one can see that η is equal to $1/a(1-a)$ for the Hebbian model. Such a value of η simplifies derivations of the capacity of the Hebbian model, and has been used in all simulations of the Hebbian model except those of figure 7.

The expression $1/Na(1-a)$ in equation (2.7) is a simplifying proportionality constant—as used in associative memories (Buhmann *et al* 1989, Amit 1989).

In the rule of equation (2.7), only the weights of active neurons are modified, this type of learning rule will be called 'mod_win' in this paper.

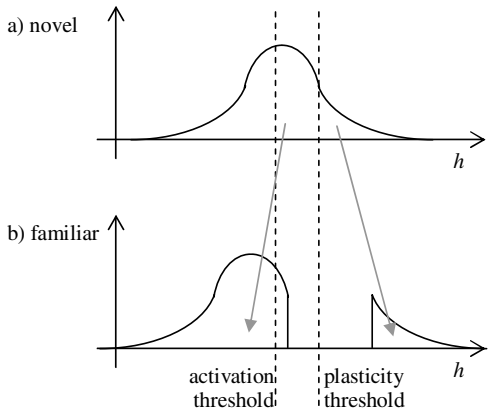

Figure 2. Intuitive explanation of the double-threshold model. The two panels show the distribution of membrane potentials for (a) a novel and (b) a familiar stimulus. The horizontal axis denotes membrane potential (h) and the vertical axis the number of neurons with a given membrane potential. Dashed lines show two thresholds: the activation threshold, above which neurons are active, and the plasticity threshold separating neurons whose synaptic weights are modified in different ways.

Third, although the decision about stimulus familiarity is based on the initial response of the novelty neurons, a learning rule of type mod_win is used. Hence a more complex decision function must be used if the capacity is to be maximized. It will be introduced in section 3.2.

2.3. Double-threshold model

Sohal and Hasselmo (2000) proposed a model to explain the responses of perirhinal neurons during recognition memory tasks. Two separate mechanisms were proposed for long-term and short-term recognition memory. Since this paper is concerned with long-term recognition memory, only a model simplified from the part of the Sohal and Hasselmo (2000) model concerned with long-term familiarity discrimination is analysed here. It is termed the double-threshold model.

The double-threshold model also employs Hebbian rules of learning, but the decrease in the number of neurons active for familiar stimuli is not caused by inhibition. The way in which this network discriminates familiarity is illustrated in figure 2. After presentation of a novel stimulus, the membrane potentials of the novelty neurons may be assumed to follow a normal distribution (figure 2(a)). The proportion of neurons with membrane potentials which are higher than a certain value, denoted as the plasticity threshold in figure 2, have their synaptic weights modified as for active novelty neurons in the Hebbian model (see figure 1(a)): the weights from active inputs are increased and the weights from inactive inputs are decreased. The weights of neurons with membrane potentials below the plasticity threshold are modified as for inactive novelty neurons in the Hebbian model (see the inactive neuron in figure 1(b)): the weights from active inputs are decreased. In the double-threshold model, the weights of all neurons are modified, i.e. a rule of type mod_all is used.

The double-threshold model represents a simplified version of the Sohal and Hasselmo (2000) model. The original equation for the weight updating rule in their model is very complex; the simplifications made below do not change the operational principles of the model and hence its capacity, while making possible its mathematical analysis. In the simplified

version of the model analysed here, the weights are updated according to the following rule:

$$\Delta w_{ij} = \frac{\eta}{Na(1-a)}(y_i^\mu - a)(x_j^\mu - a). \tag{2.8}$$

y_i^μ is equal to 1 if the membrane potential of neuron i (calculated from equation (2.1)) is above a plasticity threshold, and is 0 otherwise. Again, η denotes the learning rate. In the original Sohal and Hasselmo (2000) model the plasticity threshold is fixed. But here, for simplicity of analysis, it is assumed that the exactly aN novelty neurons with the highest membrane potentials are above the plasticity threshold. This would mean that the plasticity threshold is chosen for each stimulus such that the above criterion is satisfied. This criterion might be achieved in the brain via competitive inhibition. It acts to increase rather than decrease the capacity of the network (in a way analogous to the assumption about the constant number of active input neurons analysed in section 3.1).

These weight modifications mean that when the initially novel stimulus is subsequently re-presented, the membrane potentials are even higher for the neurons which were above the plasticity threshold on the first presentation (they follow the change indicated by the right arrow connecting figures 2(a) with (b)), while the membrane potentials for other neurons are even lower (they follow the left arrow in figure 2). In the double-threshold model there is an activation threshold—neurons with membrane potentials above this threshold are active—which is smaller than the plasticity threshold (see figure 2). If the activation threshold is set appropriately (for example by inhibition within the network—its optimal value is derived in appendix C), more neurons are active for novel than for familiar stimuli (compare the areas under the distribution density curves to the right of the activation threshold in figures 2(a) and (b). The number of active novelty neurons can thus be used as the familiarity criterion.

2.4. Anti-Hebbian model

The anti-Hebbian model is based on the ideas of Brown and Xiang (1998), and Kohonen *et al* (1974), but its full mathematical description is introduced for the first time in this paper. The anti-Hebbian model (like the Hebbian model) assumes that a proportion of perirhinal neurons forms a network specialized for familiarity discrimination. In the anti-Hebbian model, the neurons are selected to be active as in the combined competitive models: i.e. the aN neurons with the highest membrane potential are selected, and their pattern of activity is denoted by y (see section 2.2).

However, by contrast to all previously described models in which the synaptic weights are modified according to Hebbian rules, in the anti-Hebbian model they are modified in the opposite way. Figure 3 shows the modification of the weights of the most active novelty neurons for a sample pattern. After presentation of a novel stimulus the synaptic weights of connections from active input neurons are decreased as if by homo-synaptic LTD. This synaptic modification decreases the sum of the synaptic weights of the novelty neuron. Hence to maintain the overall excitability of the neuron, the synaptic weights of connections from inactive input neurons must be increased (see figure 3). When the same stimulus is presented again, the membrane potential of the novelty neuron will be lower (because the weights of synapses of inputs that were active for this stimulus have been reduced) and the novelty neuron will be inactive (or, more generally, less active). Thus the neuron responds more strongly to novel than familiar stimuli.

Hence, in the anti-Hebbian model, after presentation of each stimulus, the weights are updated according to the following rule:

$$\Delta w_{ij} = -\frac{\eta}{Na(1-a)} y_i^\mu (x_j^\mu - a). \tag{2.9}$$

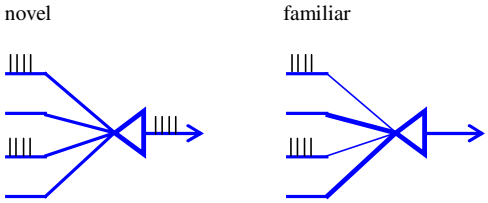

Figure 3. Modification of synaptic weights of active novelty neurons in the anti-Hebbian model. Notation as in figure 1.

In equation (2.9), η again denotes the learning rate. According to equation (2.9), the weights between active inputs and active novelty neurons (y_i^μ=1, x_j^μ=1) are decreased as if by homosynaptic LTD and, to balance the neurons' excitability, the weights between inactive inputs and active novelty neurons are increased ($y_i^\mu = 1, x_j^\mu = 0$). This latter change could be achieved simply by increasing the strength of the other synapses of a given neuron following homosynaptic LTD at some of its synapses, in such a way as to maintain the neuron's excitability.

In the rule of equation (2.9), only the weights of active neurons are modified, hence this is a rule of type mod_win. If the weights of inactive neurons had also to be modified (i.e. a rule of type mod_all were used), then for the anti-Hebbian model, modification of the weights of inactive novelty neurons would have to be according to a most unlikely rule: the weights between active inputs and inactive novelty neurons would have to be increased. There is no known biological synaptic modification mechanism to achieve such a change. Furthermore, there is no obvious compensatory mechanism by means of which the necessary change could be effected, as it must apply to only certain synapses on the inactive neurons (as opposed to a general increase in excitability). Therefore, the anti-Hebbian model is biologically plausible only if the weight modification rule is of type mod_win; a rule of type mod_all is not plausible for the anti-Hebbian model.

Synaptic plasticity, as illustrated in figure 3, decreases the responses of novelty neurons for familiar stimuli; hence the decision about stimulus familiarity may be based on the population activity of the novelty neurons. The precise form of the decision function will be introduced in section 3.2.

3. Influence of the sparseness of coding on capacity

This section analyses the influence of the sparseness of coding on capacity of the different models with the simplifying assumption that activities of all the novelty neurons and the network inputs are uncorrelated. This is an oversimplification of the situation in the real perirhinal cortex (see Erickson *et al* 2000). However, analysing the efficiency of familiarity discrimination networks when inputs are not independent is very difficult. Hence, it is useful first to analyse the influence of the sparseness of coding on the capacity of the models for the simple case of uncorrelated neuronal responses (as has been done for other memory networks, e.g. (Buhmann *et al* 1989, Amit 1989, Palm and Sommer 1992)).

3.1. Hebbian model

Appendix A.1 shows that a fully connected Hebbian network of N neurons may discriminate the familiarity with 99% accuracy for the following number of patterns:

$$P_{\max} = 0.023N^2. \tag{3.1}$$

Appendix A.1 shows that the capacity of the Hebbian model is the same (i.e. expressed by equation (3.1)) for any sparseness of coding. Hence the capacity of the Hebbian model does not depend on the sparseness of coding. This fact is unexpected as sparse representation greatly increases the capacity of associative memories (Amit 1989). Figure 4 shows that the capacity of the Hebbian model obtained in simulations matches the theoretical prediction and confirms the lack of dependence of the model's capacity on the sparseness of coding.

In section 2.1.3, equation (2.4), it was assumed for simplicity of calculation that in each pattern there were exactly aN bits not equal to 0. It would be more realistic to assume that the patterns have different numbers of active bits, and a is the probability of a given bit being active. In this case, the term ε in the weight modification rule of equation (2.6) is not necessary because the bits in the patterns are truly independent and hence there is no correlation to be counterbalanced by this term. Appendix A.2 shows that when the numbers of active inputs differ between patterns, the capacity is given by

$$P = (1-a)^2\left(0.023N^2 - \frac{N(1-a)}{a}\right). \tag{3.2}$$

Figure 5 compares the prediction of equation (3.2) with the results of simulations of the Hebbian network with different numbers of active inputs. For larger N, the second term in the bracket in equation (3.2) becomes insignificantly small in comparison to the first, and may be discarded. Hence in practice, for large N, when the number of active bits in patterns may differ, the capacity converges to $0.023N^2(1-a)^2$, and sparse coding slightly increases the capacity. This effect becomes visible in figure 5 in larger networks, e.g. note that in figure 5(c) (for $N = 300$ neurons) in series '$d = hx$', the capacity for $a = 0.2$ is larger than for $a = 0.5$.

Relaxing the assumption about there being a constant number of bits in the patterns does not decrease the capacity significantly for sparse representations, because $(1-a)^2 \approx 1$. But it decreases the capacity for less sparse representations; e.g. for $a = 0.5$, the capacity decreases fourfold. However this decrease may be avoided in at least two ways.

First, if the weights are modified according to equation (3.3) (instead of equation (2.6), then it can be shown, analogously to the derivation of appendix A.1, that the capacity remains approximately equal to $0.023N^2$. This is consistent with the results of simulations shown in figure 5 in series '$d = hx, w(a)$'. In equation (3.3), a^μ is the sparseness of pattern μ, defined as the number of bits equal to 1 in pattern μ:

$$w_{ij} = \sum_{\mu=1}^{P} \frac{1}{N(a^\mu)^2(1-a^\mu)^2}\left((x_i^\mu - a^\mu)(x_j^\mu - a^\mu) + \frac{a^\mu(1-a^\mu)}{N-1}\right). \tag{3.3}$$

Note that although the rule of equation (3.3) is more complex than that of equation (2.6), it is still locally determinable, i.e. the values of all the variables in equation (3.3) are determinable by novelty neuron i. According to equation (3.3), the novelty neuron must determine a^μ, the sparseness of the activities of the input neurons. This information could be provided to the neuron for example by feed-forward inhibition (i.e. inhibitory neurons receiving projections from the input neurons, that then inhibit the novelty neurons in proportion to the number of active inputs, see Bogacz *et al* (2001)). In particular, note that it is possible to implement the rule without the need to provide information from other novelty neurons.

A second method of avoiding the decrease in capacity was proposed by Bogacz *et al* (2001), and involves modification of the thresholds of the novelty neurons (for details see Bogacz *et al* (2001)). However, this other method is less effective: it avoids the effect due to different numbers of active inputs, but not the effect due to different numbers of potentially active novelty neurons. Therefore, when this other method is used and $a = 0.5$, the capacity is halved to $0.012N^2$ (Bogacz *et al* 2001). This difference explains the difference in capacity

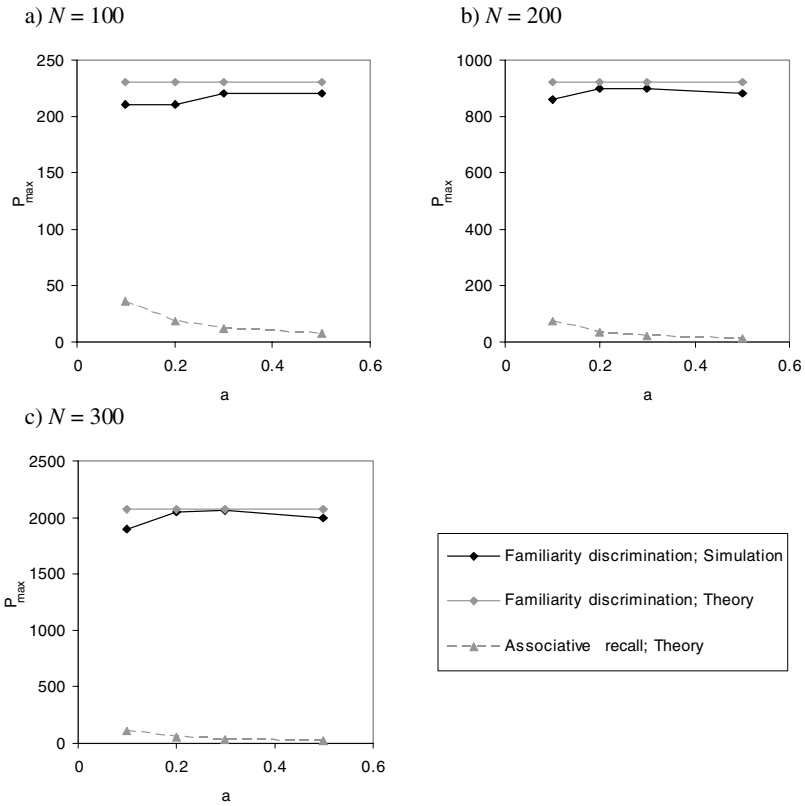

Figure 4. Comparison of capacity for familiarity discrimination and associative recall. The following convention is used in figures 4–8. Capacities are shown for fully connected networks of 100, 200 and 300 neurons (panels (a)–(c)). The sparseness of representation is shown on the x-axes; the capacity is shown on the y-axes. Black lines denote the results of simulations and grey lines show the theoretical predictions. The capacities ($P_{\max}$) of the Hebbian model and its variants are denoted by solid lines, and the capacities of the other models by dashed lines. Methods of simulation as in Bogacz *et al* (1999, 2001). For each number of neurons N, and for each number of previously stored patterns P, the behaviour of the network was tested repeatedly with sets of random patterns until it had been tested with 5000 previously presented patterns and 5000 random (novel) patterns. Namely, during each repetition, P patterns were presented to the network, and then accuracy was tested on all the presented patterns in the list and equal number of novel (i.e. random) patterns. Tests were repeated a number of times, such that the network was tested altogether with 5000 previously presented patterns and 5000 novel patterns, e.g. for $P = 100$, the simulations were repeated 50 times. The average accuracy is entered. For each number of neurons N, the network error was tested for different numbers of stored patterns P starting from $P = 1$ until the network error exceeded 1%. P was increased with different *steps*: for $P < 10$ every integer value of P was tested ($step = 1$); for $P \in (10, 50)$ every even value of P was tested ($step = 2$); for $P \in (50, 200)$, $step = 5$; for $P \in (200, 1000)$, $step = 10$; for $P > 1000$, $step = 20$. $P_{\max}$ is taken as the maximum number of stored patterns P for which the error rate is $\leqslant 1\%$. To illustrate the precision of the simulation process, for one data point (100 neurons, sparseness 0.5), the capacity was estimated ten times using the above method. The standard deviation of the estimated capacities was +/−5.4 (i.e. about 2.5% of the mean). The capacity for associative recall is taken as $0.145 \times N/(4 \times \text{sparseness})$ (Amit 1989).

given here and in Bogacz *et al* (2001). It highlights the importance to capacity of the precise learning rule.

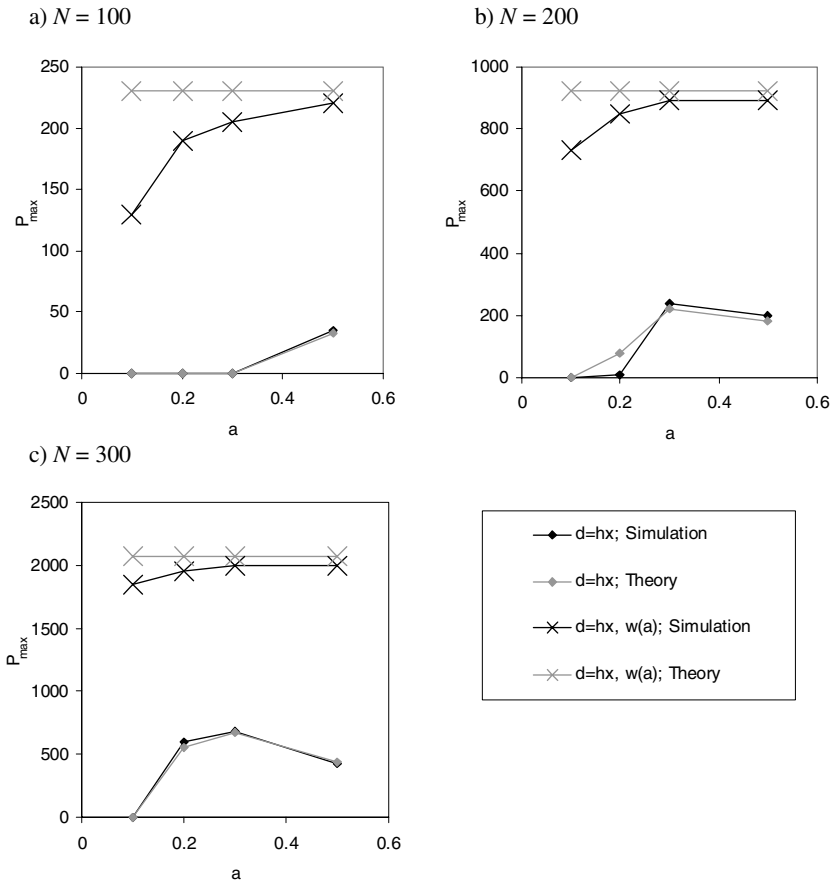

Figure 5. Capacity of the various versions of the Hebbian model when input patterns differ in the numbers of active inputs (varying sparseness across patterns). Capacities are shown for fully connected networks of 100, 200 and 300 neurons (panels (a)–(c)). The average sparseness of representation (a) is shown on the x-axes; the capacity ($P_{\max}$) is shown on the y-axes. The series marked '$d = hx$' shows the capacity for the standard decision function of equation (2.2) (the theoretical values are calculated according to equation (3.2)). The series marked '$d = hx$, $w(a)$' shows the capacity for the standard decision function, when the weights are updated according to equation (3.3) (the theoretical values according to equation (3.1)). The method of simulation is as for figure 4. Note the important influence of the combination of learning rule and decision function on capacity. In particular, in series '$d = hx$', for larger networks capacity increases for sparser representations; for example for $N = 300$, the capacity for $a = 0.2$ is larger than for $a = 0.5$ (but this is not the case in figures 5(a) and (b)); although the capacity for $a = 0.1$ is zero for the networks presented in the figure, in larger networks the capacity is likely to be larger than for less sparse patterns. In series '$d = hxw(a)$', although for $N = 100$ neurons the capacity obtained in simulations is lower than the above theoretical prediction, for larger networks (panels (b), (c)) the capacity in simulations converges upon the theoretical prediction.

To summarize, this section shows that the number of stimuli for which the Hebbian model may discriminate familiarity with probability of error 1% is equal to 0.023 times the number of modifiable synapses in the network, and can be made independent of the sparseness of coding. This capacity is achieved when the number of active bits is equal in each input pattern representing stimuli. When this assumption is not satisfied, the capacity may decrease slightly (up to fourfold, for $a = 0.5$), but this decrease may be avoided by a potentially biologically

implementable modification of the weight change rule. For simplicity, in the following sections of this paper it will be assumed that the number of active inputs in each pattern is equal (i.e. as constrained by equation (2.4)).

3.2. Role of modifying weights of inactive neurons

To observe the importance of the modification of the synaptic weights of inactive novelty neurons, we will analyse what happens if only the weights of active novelty neurons are modified after presentation of a novel stimulus. Let us consider a modification of the Hebbian model, in which the weights are modified according to a rule of type mod_win. The synaptic weights after presentation of P stimuli in this modified version of the Hebbian model are given by

$$w_{ij} = \frac{1}{Na^2(1-a)^2} \sum_{\mu=1}^{P} y_i^\mu (x_j^\mu - a) = \frac{1}{Na^2(1-a)^2} \sum_{\mu=1}^{P} x_i^\mu (x_j^\mu - a). \quad (3.4)$$

Appendix B shows that if the above learning rule of type mod_win is used (with the standard decision function of type act_win; equation (2.2), the capacity decreases dramatically to

$$P_{\max} \approx \frac{0.046N^2}{Na^2 + (1-a)^2}. \quad (3.5)$$

For large N, the term Na^2 in the denominator is much larger than $(1-a)^2$ and the capacity becomes proportional to N rather than N^2. In practice, this means that the capacity is much lower than in the case of the standard weight modification rule of type mod_all (equation (2.6)), where the weights of inactive novelty neurons are also modified. The less sparse the representation, the larger this decrease. Sample results of simulations and theoretical predictions are shown in figure 6 (series labelled '$d = hx$'). For larger values of a (e.g. values of 0.5 and 0.2 in figure 6) the results of simulations match the theoretical predictions of equation (3.5).

When a approaches 0, then the capacity expressed by equation (3.5) converges to $0.046N^2$, a value twice as high as the capacity of the standard Hebbian model (equation (3.1)). The reason for this is that the calculations of appendix B assume that the weights are not symmetric, i.e. $w_{ij} \neq w_{ji}$. However for a approaching 0, the weight modification rule of equation (3.4) converges to the standard weight modification rule of the Hebbian model of equation (2.6), and the weights become symmetric. As explained by Bogacz *et al* (1999, 2001), the symmetry of the weights halves the capacity (intuitively, when the weights are symmetric, there are pairs of weights storing the same information, thus the amount of information that can be stored in a given number of weights is half that of the case where the weights do not need to be symmetric; for a formal explanation see Bogacz *et al* (1999, 2001). Hence, for a approaching 0, the capacity of the modification of the Hebbian model with the weight updating rule of equation (3.4) converges to $0.023N^2$ rather than $0.046N^2$. For small values of a, the predictions of capacity of equation (3.5) are higher than the results of simulations (e.g. value of 0.05 in figure 6).

To avoid the decrease in capacity due to using a learning rule of type mod_win, one can use the following decision function:

$$d(x) = \sum_{i=1}^{N} (x_i - a)h_i' = \sum_{\substack{i,j=1 \\ i \neq j}}^{N} (x_i - a)x_j w_{ij}. \quad (3.6)$$

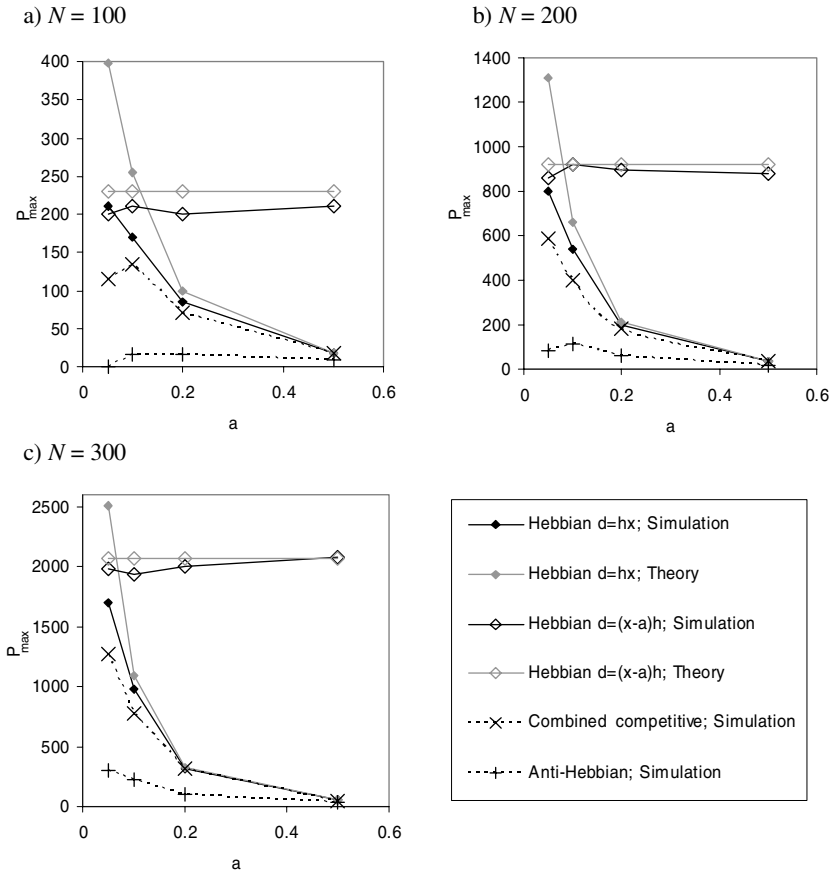

Figure 6. Capacity of the familiarity discrimination networks in which only the weights of active novelty neurons are modified during learning (i.e. for mod_win learning rules). Capacities are shown for fully connected networks of 100, 200 and 300 neurons (panels (a)–(c)). The sparseness of representation is shown on the x-axes; the capacity is shown on the y-axes. The method of simulation is as for figure 4 (see also section 3.3 for the method of simulation of the combined competitive and the anti-Hebbian models). Note that the reduction in capacity due to the mod_win learning rules can be prevented by using an act_dif decision function.

Let us note that equation (3.6) may be rewritten as

$$d(x) = \sum_{i=1}^{N} x_i h'_i - a \sum_{i=1}^{N} h'_i. \tag{3.7}$$

Hence the decision function of equation (3.6) is equal to the decision function of type act_win (the first summation in equation (3.7)) decreased by the sum of the membrane potentials of all the novelty neurons multiplied by the sparseness of coding (the second summation in equation (3.7)). Therefore the decision function of equation (3.6) may be calculated by a biologically plausible neural network. An example of such a network is given in Bogacz (2001).

Since the decision function of equation (3.6) is based on the difference between the activities of the most and the least active neurons, decision functions of this type will be called 'act_dif' in the remainder of this paper.

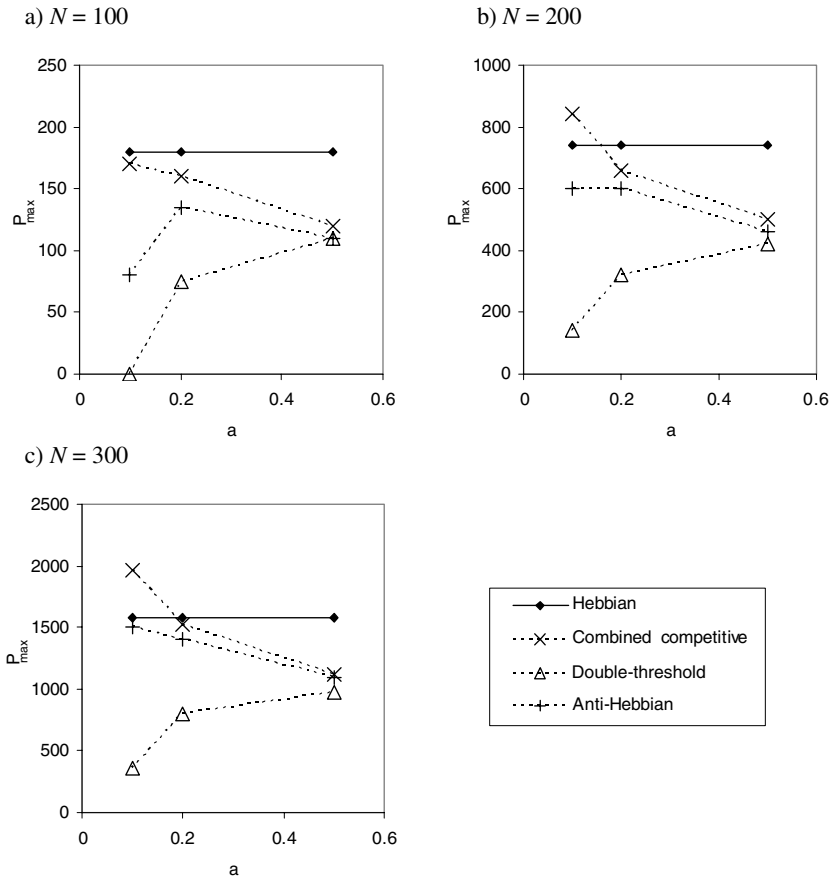

Figure 7. Comparison of the capacity of familiarity discrimination networks for random patterns. The combined competitive model is a simplified version of the neocortical part of the Norman and O'Reilly (2001) model, and the double-threshold model is a simplified version of the Sohal and Hasselmo (2000) model. Capacities are shown for fully connected networks of 100, 200 and 300 neurons (panels (a)–(c)). The sparseness of representation is shown on the x-axes; the capacity is shown on the y-axes. The method of simulation is as for figure 4; see also section 3.3. For comparison, the capacity for the Hebbian model is also shown. It is slightly lower than in figure 4, because of the use of a different weight modification rule and normalization (see section 3.3). In the results for the double-threshold network, for sparseness $a = 0.5$, the activation threshold was -0.4 (not $-a/2$; see appendix C) because it resulted in the highest capacity. Note that the sparseness of coding increases the capacity of the combined competitive and the anti-Hebbian models, while it decreases the capacity of the double-threshold model.

If a decision function of type act_dif (defined as in equation (3.6), and not act_win as in equation (2.2)) is used together with a learning rule of type mod_win, equation (3.4), the network achieves a capacity of $0.023N^2$, the same as for the standard Hebbian model. This result follows as both the signal and noise under these new rules have exactly the same values as in appendix A.1. Sample results of simulations for the decision function of equation (3.6) with the learning rule of equation (3.4) are shown in figure 6 (series labelled '$d = (x - a)h$').

Although it is difficult to show analytically, simulations confirm that the above principles apply to the other models as well as the Hebbian model. In particular, the combined competitive and the anti-Hebbian models use learning rules of type mod_win (see equations (2.7) and (2.9)).

Table 1. Relative capacity achieved by familiarity discrimination networks for different combinations of learning rule and decision function. 'High capacity' denotes combinations where the capacity is proportional to the number of synapses in the network; 'low capacity' is where the capacity is proportional to the number of neurons in the network (for large networks).

Decision function	Learning rule	
	mod_all	mod_win
act_win	High capacity	Low capacity
act_dif	High capacity	High capacity

When the combined competitive and the anti-Hebbian models use the decision function of equation (3.8), which is of type act_win, then these models also have only a very low storage capacity (as for the Hebbian model), as is shown in figure 6:

$$d(x) = \sum_{i=1}^{N} y_i h_i. \tag{3.8}$$

Therefore, in the combined competitive and the anti-Hebbian models analysed in this paper it is important to use the following decision function, which is of type act_dif:

$$d(x) = \sum_{i=1}^{N} (y_i - a) h_i. \tag{3.9}$$

The capacity of the combined competitive and the anti-Hebbian models, when using the decision function of equation (3.9) is analysed in the following sections.

To summarize, this section has calculated the capacity of familiarity discrimination networks when the weights of only active novelty neurons are modified (which may be necessary for a plausible implementation of the anti-Hebbian model; see section 2.4). In this case, the capacity becomes very low when the familiarity judgement is based on the population activity of novelty neurons (act_win). A high capacity may be achieved when the familiarity judgement is based on the difference between the responses of the more active and the responses of the less active neurons (act_dif).

Table 1 summarizes which types of learning rule and decision function allow familiarity discrimination networks to achieve a high storage capacity. The magnitudes of the effects upon capacity establish that only certain combinations provide plausible solutions.

3.3. Method of simulation of the models which use competition

The capacities of the combined competitive, the double-threshold and the anti-Hebbian models have been found in simulations. This section introduces the method used to generate these simulations.

In these simulations, the synaptic weights of the novelty neurons were initialized to random values. After each weight modification (following simulated stimulus presentations) the weights were normalized such that for each neuron the mean of its weight was 0 and the sum of the weights squared was 1, i.e. the following constraints are forced for each neuron i:

$$\sum_{j=1}^{N} w_{ij} = 0 \qquad \text{and} \qquad \sum_{j=1}^{N} w_{ij}^2 = 1. \tag{3.10}$$

This normalization ensures equal chances of activation for each neuron. However, the normalization also means that stimuli presented initially were not as well remembered as ones presented subsequently. To avoid this problem, the stimuli were presented twice during training in each simulation session, the second presentation being in the reverse order to the first presentation. The effect of such normalization and double presentation was also simulated for the Hebbian model (see series 'Hebbian' in figure 7): capacity was decreased by $\sim$20%.

The learning rate η (see equation (2.7)) influences the familiarity discrimination threshold of the decision functions (which also depends on N and a). The value of the threshold was taken in simulations as the average of the mean decision function values for novel and familiar stimuli.

3.4. Combined competitive model

The combined competitive model (simplified version of the neocortical part of the Norman and O'Reilly (2001) model) operates in a similar way to the Hebbian model, hence they have similar capacity when the inputs are independent (figure 7). The fact that in the combined competitive model the active neurons are chosen as those with the highest membrane potential (rather than those receiving stronger connections in the Hebbian model) affects the capacity in two ways.

First, a different subset of neurons may have the highest membrane potential, and hence be chosen to be active, for the first and second presentations of a stimulus (because of weight modifications produced by the occurrences of intervening novel stimuli). This decreases the capacity (because it reduces the signal term in the decision function; the signal term is defined in appendix A.1).

This effect of change in representation is less prominent for sparse representations, making the pattern of activity of the novelty neurons more stable for a given stimulus for such cases; i.e. for sparse representations, it is more likely that the same novelty neurons are activated during the first and the second presentations of a given stimulus. This may be explained by the following analogy with hetero-associative memories. The Hebbian learning rule of the combined competitive model strengthens the associations between x^μ and y^μ in a way similar to that in which the Hebbian learning rule of hetero-associative memories strengthens the associations between pairs of stored patterns. The associations are more stable in the hetero-associative memories for sparse coding (and they achieve higher capacity Amit (1989)) and, similarly, the associations between x^μ and y^μ, in the combined competitive model are more stable for sparse coding. Therefore, the effect of change in representation is less prominent for sparse coding, and the capacity of the combined competitive model is larger for sparse representations.

Second, for very sparse representations (e.g. $a = 0.1$), when only a small proportion of the novelty neurons are selected to have their weights modified, each neuron is selected to be active for relatively similar patterns (i.e. patterns that share some proportion of bits or share the same features); this follows because the novelty neurons are activated for the patterns most similar to their weight vectors (such patterns result in the highest membrane potentials for the neurons; see equation (2.1)). Consider the case when a novelty neuron is selected to be active for a given pattern A, and the weight modification resulted in its membrane potential for pattern A being increased. Then, it is likely that the weight modifications caused by other, subsequent patterns (which are somewhat similar to A) for which the neuron is active will result in a yet further increase in the membrane potential of the neuron during the next presentation of A. This follows as the weights of the neuron become more and more correlated with the features shared by the patterns for which the neuron was activated. The increase in the membrane

potential during the second presentation of A contributes to the increase in the value of the decision function. Therefore, the result of repeated activation of a novelty neuron by relatively similar patterns is to increase the storage capacity for uncorrelated input patterns.

For less sparse representations the first effect of a decrease in capacity due to the change in representation prevails. Hence the combined competitive model has a slightly lower capacity than the Hebbian model for less sparse representations (e.g. $a = 0.2$ or 0.5 in figure 7). However, for very sparse representations the second effect of an increase in capacity due to the repeated activation of novelty neurons by similar patterns prevails. Thus when neuronal activity is uncorrelated, the combined competitive model achieves a slightly higher capacity than the specialist Hebbian model for very sparse representations (e.g. $a = 0.1$ in figure 7). Also, the combined competitive network has a slightly greater capacity for sparse than for non-sparse representations.

3.5. Double-threshold model

Appendix C shows that the capacity of the double-threshold model (the simplified version of the Sohal and Hasselmo (2000) model) for random patterns is proportional to the number of synapses of the novelty neurons. Surprisingly (by contrast to any other memory networks), appendix C also shows that the capacity decreases for sparser representations. This finding is in agreement with the results of the simulations presented in figure 7. The figure shows that for uncorrelated patterns that are not sparse, the capacity of the double-threshold model is similar to the capacity of the combined competitive model, but lower than that of the Hebbian model. For sparse representations, the capacity of the double-threshold network is the lowest of the models.

3.6. Anti-Hebbian model

The capacity of the anti-Hebbian model is shown in figure 7 to be also proportional to the number of modifiable synapses in the network. Figure 7 indicates that the capacity of the anti-Hebbian model is slightly larger for sparse representations. Although in figure 7 the capacity for $a = 0.1$ is lower than that for $a = 0.2$ when $N = 100$, this is an artefact of the small number of neurons (i.e. $N = 100$). For larger networks (e.g. $N = 300$ in figure 7), the capacity of the anti-Hebbian model is increased by sparse representations.

The observation that the anti-Hebbian model has slightly higher capacity for sparser representations may come from the following fact. For small a, only a small fraction of the most active novelty neurons (those which contribute to the value of the decision function most) have their weights modified, while the weights of all other neurons are left unchanged. By contrast for higher a (i.e. less sparse representation), many more neurons (including those which do not contribute so much to the value of the decision function) have their weights modified. Thus for sparser representations the weight modifications are distributed in a more nearly optimal way, so increasing the capacity.

4. Discussion

This paper analyses how the sparseness of coding influences the capacity of familiarity discrimination networks. It shows that for uncorrelated input patterns the familiarity discrimination networks have approximately similar capacities, being proportional to the number of synapses in the network. The sparseness of coding has different effects on different familiarity discrimination networks. The capacity of an optimally implemented Hebbian model

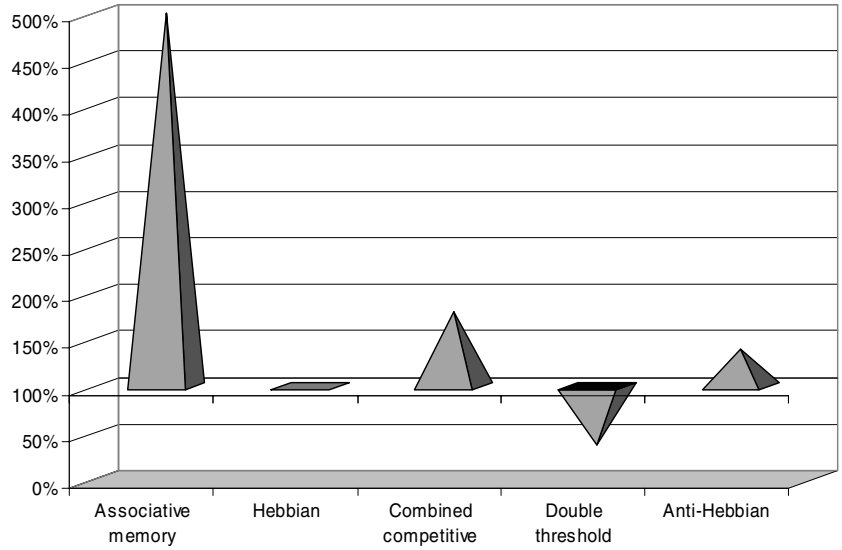

Figure 8. Schematic representation of the changes in the capacity achieved by different networks due to increasing the sparseness of coding (decreasing a from $a = 0.5$ to 0.1). The combined competitive model is a simplified version of the neocortical part of Norman and O'Reilly (2001) model, and the double-threshold model is a simplified version of the Sohal and Hasselmo (2000) model. The value of 100% represents the capacity of a network of $N = 300$ neurons with a sparseness of coding of $a = 50\%$. Each pyramid points toward the capacity achieved by the network of the same size with a sparseness of coding of 10%. The values of the capacities of familiarity discrimination networks were taken from figure 7(c).

does not depend on the sparseness of coding. Sparser representations increase the capacity of the combined competitive and the anti-Hebbian models, and decrease the capacity of the double-threshold model. In general, the influence of the sparseness of coding on the capacity of familiarity discrimination networks is far smaller than on the capacity of associative memory networks. The above observations are summarized in figure 8 which shows the changes in capacity achieved by different models due to changing the sparseness of coding from $a = 0.5$ to 0.1.

It is important to emphasize that the influence of the sparseness of coding on the capacity of familiarity discrimination networks is much smaller than on the capacity of associative memory networks. This fact may be explained intuitively in the following way. Sparse coding increases the capacity of associative memories because it reduces the statistical noise in the activities of individual neurons and hence reduces the probability of individual neurons making an error in recalling of a pattern (Amit 1989). In the case of familiarity discrimination, the decision of the network is based on the population activity of a group of neurons (not on the activity of individual neurons). Hence, although sparse coding increases the accuracy of the responses of individual novelty neurons, it decreases the number of novelty neurons active and hence the number of those contributing to the decision process. These two effects of sparse coding (i.e. increased accuracy of individual neurons, and decreased number of neurons involved) act to counterbalance each other and therefore the influence of the sparseness of coding on the capacities of familiarity discrimination networks is lower than for associative memories.

Sections 3.1 and 3.2 show that the capacity of the Hebbian model may be changed by making small modifications to the learning rule and the decision function. This highlights the

importance of the precise form of learning rules for the capacity of familiarity discrimination networks. The importance of the precise form of learning rules has been also shown for the capacity of associative memories (Willshaw and Dayan 1990, Dayan and Willshaw 1991).

Section 3.2 shows that when only the weights of active novelty neurons are modified, then a more complicated familiarity criterion (i.e. decision function act_dif) must be used to maximize the storage capacity of the network. In this case, in order to achieve high capacity while using a simple (thus biologically plausible) learning rule, the role of determining the required average characteristics of the input stimuli is shifted from the learning rule to the discrimination function. A similar approach has been used in modelling associative memories: Buckingham and Willshaw (1993) and Graham and Willshaw (1995) showed that more sophisticated strategies of setting the threshold for sparsely connected associative memory networks result in higher capacities than are obtained by using the simple learning rule proposed by Willshaw *et al* (1969).

This paper establishes how sparseness of coding influences the capacity of familiarity discrimination networks under the simplifying assumption that the responses of the neurons providing input to the network are uncorrelated. Analysis of the capacity of the familiarity discrimination networks when the activities of the input neurons are correlated is much more complex. It is not given here, but it is discussed by Bogacz (2001). Preliminary results indicate that the sparseness of coding has relatively little effect (in comparison to that for associative memories) on the capacity of familiarity discrimination networks also for correlated input patterns. In particular, Bogacz (2001) shows that the capacity of the Hebbian model for correlated inputs is the same for any sparseness of coding. Investigation of the influence of the sparseness of coding on the capacity of other familiarity discrimination networks for correlated input patterns will be the subject of future work.

The purpose of this paper was to analyse the influence of the sparseness of coding on the capacity of familiarity discrimination networks. All the published models of familiarity discrimination in the perirhinal cortex are compared in terms of the efficiency they are likely to achieve in the human brain and the consistency of their behaviour with experimental observations in Bogacz and Brown (2002).

There is evidence that stimulus representation in the rhinal cortex is sparse, i.e. for each stimulus fewer than half the neurons respond (see also Barnes *et al* (1990), Kreiman *et al* (2000)) but there are indications that the representation in the perirhinal cortex is less sparse than in the hippocampus (Barnes *et al* 1990, Brown *et al* 1996).

According to information theory, less sparse patterns may encode more information than more sparse patterns (Amit 1989). Since for familiarity discrimination the sparseness of coding does not influence the capacity very significantly, there is no efficiency gain in using sparse coding in such a network. The less sparse representation may lead to better discriminability between stimuli and so may improve accuracy at no cost to capacity. This conclusion may explain why the coding in the rhinal cortex seems to be less sparse than in the hippocampus (Barnes *et al* 1990, Brown *et al* 1996).

Acknowledgments

We thank Kenneth Norman and Christophe Giraud-Carrier for discussion and comments on an earlier version of the manuscript. This work was supported by the ORS, Wellcome Trust, and MRC.

Appendix A. Capacity of the Hebbian model

This appendix analyses the capacities of different variations of the Hebbian model.

A.1. Capacity for the standard case

In this appendix we will calculate the capacity of the Hebbian model using signal-to-noise analysis (for a clear introduction to this method see Hertz *et al* (1991)). We will calculate successively

(i) the mean value of the decision function d (defined in equation (2.2)) for familiar patterns,
(ii) the mean d for novel patterns,
(iii) the variance of d across patterns.

Using these three values, we will find the probability of discrimination error. Finally, having the expression for error probability, we will find the number of stored patterns $P_{\max}$ for which this error probability is 1%; $P_{\max}$ is then the capacity.

Let us therefore calculate the value of the decision function of equation (2.2) after presentation of a sample familiar stimulus (e.g. x^1):

$$\begin{aligned} d(x^1) &= \sum_{\substack{i,j=1\\ i\neq j}}^{N} x_i^1 w_{ij} x_j^1 = \frac{1}{Na^2(1-a)^2} \sum_{\substack{i,j=1\\ i\neq j}}^{N} \sum_{\mu=1}^{P} x_i^1 x_j^1 (x_i^\mu - a)(x_j^\mu - a) \\ &= \frac{1}{Na^2(1-a)^2} \sum_{\substack{i,j=1\\ i\neq j}}^{N} x_i^1 x_j^1 (x_i^1 - a)(x_j^1 - a) \\ &\quad + \frac{1}{Na^2(1-a)^2} \sum_{\substack{i,j=1\\ i\neq j}}^{N} \sum_{\mu=2}^{P} x_i^1 x_j^1 (x_i^\mu - a)(x_j^\mu - a). \end{aligned} \tag{A.1}$$

The first term in the last expression of equation (A.1) is called the signal in neural network literature (Amit 1989) and the second term is called noise. Since $x_j^\mu \in \{0, 1\}$, then $x_j^\mu = (x_j^\mu)^2$; hence the signal term is given by

$$\text{Signal} = \frac{1}{Na^2(1-a)^2} \sum_{\substack{i,j=1\\ i\neq j}}^{N} x_i^1 (1-a) x_j^1 (1-a) = \frac{1}{Na^2} a^2 N^2 = N. \tag{A.2}$$

The noise term may be rewritten as

$$\text{Noise} = \frac{1}{Na^2(1-a)^2} \sum_{\mu=2}^{P} \sum_{i=1}^{N} x_i^1 (x_i^\mu - a) \sum_{\substack{j=1\\ i\neq j}}^{N} x_j^1 (x_j^\mu - a). \tag{A.3}$$

The term $\sum_{i=1}^{N} x_i^1 (x_i^\mu - a)$ may be treated as a random variable with a hypergeometrical distribution, for the following reason. A hypergeometrical distribution is followed when we select n elements from a set of N elements, among which M elements have a certain feature. Here the elements are $(x_i^\mu - a)$; there are N of these elements, among which $M = aN$ have the feature that they are equal to $1 - a$. We select $n = aN$ elements—those for which $x_i^1 = 1$ (for the rest $x_i^1 = 0$). Therefore, it may be shown that this variable has mean 0 and variance $Na^2(1-a)^2$. The variable approximating the term $\sum_{j=1, i\neq j}^{N} x_j^1 (x_j^\mu - a)$ has the same mean and variance. Hence it might be expected that the variable approximating

$\sum_{i=1}^{N} x_i^1(x_i^\mu - a) \sum_{j=1, i\neq j}^{N} x_j^1(x_j^\mu - a)$ would have mean 0 and variance $N^2a^4(1-a)^4$. However, the variance of this variable is twice as large, because the weights are symmetrical ($w_{ij} = w_{ji}$; see equation (2.6), and hence in this summation there are pairs of the same elements, i.e. $x_i^1(x_i^\mu - a)x_j^1(x_j^\mu - a)$ and $x_j^1(x_j^\mu - a)x_i^1(x_i^\mu - a)$; they can be added once only and the result may be multiplied by 2, which will result in twice the variance (see Bogacz *et al* (1999, 2001)). Since the whole summation in equation (A.3) is a sum of P of these variables, the whole summation may be approximated by a normal distribution with mean 0 and variance $2PN^2a^4(1-a)^4$. Let $\theta(m, \sigma)$ denote a random variable with mean m and standard deviation σ. It follows that

$$\text{Noise} \approx \frac{1}{Na^2(1-a)^2}\theta\left(0, \sqrt{2P}Na^2(1-a)^2\right) = \theta\left(0, \sqrt{2P}\right). \tag{A.4}$$

After presentation of a novel stimulus, the signal is equal to 0, and the noise has the same mean and variance as for familiar stimuli. Hence, the average value of d for familiar stimuli is N, while for novel stimuli it is 0. Therefore, by taking as the threshold the middle value $N/2$, we can define a familiarity discrimination criterion, namely, if $d > N/2$ then the stimulus is considered familiar, otherwise it is novel.

This familiarity discrimination network works well when the noise θ is small. We consider the network 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:

$$\Pr\left(\theta(0, \sqrt{2P_{\max}}) < \frac{N}{2}\right) = 0.99. \tag{A.5}$$

In equation (A.5), Pr denotes probability. Equation (A.5) is equivalent to

$$\Pr\left(\theta(0, 1) < \frac{N}{\sqrt{8P_{\max}}}\right) = 0.99. \tag{A.6}$$

Since the noise may be estimated by a normal distribution, equation (A.6) may be solved by checking the value of the inverted standard normal cumulative distribution for 0.99:

$$\frac{N}{\sqrt{8P_{\max}}} \approx 2.33. \tag{A.7}$$

Solving equation (A.7) with respect to $P_{\max}$, we get $P_{\max} \approx 0.023N^2$ (Bogacz *et al* 1999).

A.2. Capacity for different number of active inputs

When the number of active bits in a pattern may differ between patterns and the decision function of equation (2.2) is used, the capacity of the Hebbian model decreases slightly due to additional noise which has two sources. First, the terms $\sum_{i=1}^{N} x_i^1(x_i^\mu - a)$ and $\sum_{j=1, i\neq j}^{N} x_j^1(x_j^\mu - a)$ in the noise equation (A.3) have a binomial rather than a hypergeometrical distribution. Hence, it may be shown that the variance of each of these terms is increased by a factor of $1/(1-a)$, thus becoming $Na^2(1-a)$. Therefore, the total variance of the noise increases by $1/(1-a)^2$. Secondly, the signal is no longer constant and so must be treated as a random variable. Let us calculate the value of the signal after presentation of the first pattern:

$$\text{Signal} = \frac{1}{Na^2(1-a)^2} \sum_{\substack{i,j=1 \\ i\neq j}}^{N} x_i^1(x_i^1 - a)x_j^1(x_j^1 - a) \approx \frac{1}{Na^2}\left(\sum_{i=1}^{N} x_i^1\right)^2. \tag{A.8}$$

In equation (A.8) x_i^μ may be treated as a random variable with mean a and variance $a(1-a)$. Hence, the summation inside the bracket in equation (A.8) has mean $\mu = Na$ and variance $V = Na(1-a)$. Therefore, the square of the summation has mean N^2a^2 and the following variance:

$$D^2 = 2\mu^2 V + V^2 = 2N^2a^2Na(1-a) + N^2a^2(1-a)^2. \tag{A.9}$$

For large N, the second term in equation (A.9) is much smaller than the first and may be discarded. Hence, the mean of the signal is equal to N and the variance of the signal is equal to $2N(1-a)/a$. Making calculations analogous to those of equations (A.5)–(A.7), it may be shown that the capacity of the Hebbian model for different numbers of active inputs between patterns is expressed by equation (3.2), i.e.

$$P = (1-a)^2\left(0.023N^2 - \frac{N(1-a)}{a}\right). \tag{3.2}$$

For larger N, the second term in the bracket in equation (3.2) becomes insignificantly small in comparison to the first, and may be discarded. Hence, for large N, when the number of active bits in patterns may differ, the capacity converges to $0.023N^2(1-a)^2$.

Appendix B. Capacity of the Hebbian model with a mod_win learning rule

To observe the importance of the modification of the synaptic weights of inactive novelty neurons, we will analyse what happens if only the weights of active novelty neurons are modified after presentation of a novel stimulus (a mod_win learning rule is used). For this case, the synaptic weights after presentation of P stimuli in the modified version of the Hebbian model are expressed by equation (3.4).

For simplicity, assume (as section 2.1.2, equation (2.4)) that the number of active inputs in each pattern is constant and equal to aN. Calculations analogous to those of equations (A.1), (A.2) show that after presentation of a familiar stimulus the signal in d is equal to $-N/(1-a)$. Let us calculate the value of the noise in d after presentation of familiar stimulus x^1. Calculations analogous to those of equations (A.1) and (A.3) give

$$\text{Noise} = \frac{1}{Na^2(1-a)^2}\sum_{\mu=2}^{P}\sum_{i=1}^{N} x_i^1 x_i^\mu \sum_{\substack{j=1\\ i\neq j}}^{N} x_j^1(x_j^\mu - a). \tag{B.1}$$

As in appendix (A.1), term $\sum_{j=1,\,i\neq j}^{N} x_j^1(x_j^\mu - a)$ has a hypergeometrical distribution with mean 0 and variance $V = Na^2(1-a)^2$. The term $\sum_{i=1}^{N} x_i^1 x_i^\mu$ has the same variance V but its mean is equal to $\mu = Na^2$. Hence, term $\sum_{i=1}^{N} x_i^1 x_i^\mu \sum_{j=1,\,i\neq j}^{N} x_j^1(x_j^\mu - a)$ has mean 0 and variance equal to

$$D^2 = \mu^2 V + V^2 = N^2a^4(1-a)^2(Na^2 + (1-a)^2). \tag{B.2}$$

The noise may be approximated by a normal distribution as in appendix (A.1):

$$\text{Noise} \approx \theta\left(0, \frac{\sqrt{P(Na^2 + (1-a)^2)}}{(1-a)}\right). \tag{B.3}$$

Function d may be used as a criterion of familiarity as in the Hebbian model, but this time the threshold is equal to $-N/(2(1-a))$. Finding the storage capacity analogously to equations (A.5)–(A.7), we get the capacity expressed by equation (3.5):

$$P_{\max} \approx \frac{0.046N^2}{Na^2 + (1-a)^2}. \tag{3.5}$$

For large N, the term Na^2 in the denominator is much larger than $(1-a)^2$ and the capacity of the Hebbian model with a mod_win learning rule becomes proportional to N rather than N^2.

Appendix C. Capacity of the double-threshold model

The behaviour of the double-threshold model (simplified version of the Sohal and Hasselmo (2000) model) is complex, and hence its capacity is calculated here using approximations. After delivery of familiar stimulus x^1, the membrane potential of a novelty neuron whose membrane potential was below the plasticity threshold when x^1 was presented for the first time ($y_i^1 = 0$) is

$$\begin{aligned} h_{i:y_i^\mu=0}(x^1) &= \sum_{j=1}^{N} w_{ij} x_j^1 \\ &\approx \frac{1}{Na(1-a)} \sum_{j=1}^{N} \sum_{\mu=1}^{P} (y_i^\mu - a)(x_j^\mu - a) x_j^1 \\ &= \frac{1}{Na(1-a)} \sum_{j=1}^{N} (0-a)(x_j^1 - a) x_j^1 + \frac{1}{Na(1-a)} \sum_{j=1}^{N} \sum_{\mu=2}^{P} (y_i^\mu - a)(x_j^\mu - a) x_j^1. \end{aligned} \tag{C.1}$$

Between the first and second lines there is only an approximate equality, because of the learning constant η and the normalization process (described in section 3.3). The last line of equation (C.1) contains signal and noise terms. It is easy to calculate that the signal term is equal to $-a$. It can also be shown that the elements of the summations in the noise term have mean 0 and variance $(1-a)^2 a^3$. Since in the noise term there is a sum of approximately NP such elements, the summation may be approximated by a normal distribution with variance $NP(1-a)^2 a^3$.

Hence the distribution of the membrane potentials of novelty neurons such that $y_i^\mu = 0$ after presentation of a familiar stimulus may be approximated by $\theta(-a, \sqrt{Pa/N})$, and after presentation of novel stimulus by $\theta(0, \sqrt{Pa/N})$—because there is no signal. Therefore, to maximize storage capacity the activation threshold can be taken as $-a/2$. (Since it was assumed for simplicity of calculation that the average value of w_{ij} is 0—see section 3.3—then h_i may be negative and hence, under these artificially simplified conditions, the activity threshold may also be negative. In reality, such negative values may be readily avoided by assuming that the mean background activity of the network is above zero.) For the purpose of calculating capacity, we can approximate the distribution of all novelty neurons after presentation of familiar stimulus μ by $\theta(-a, \sqrt{Pa/N})$ (the approximation is shown in figure C.1) because after this approximation the potentials of the neurons for which $y_i^\mu = 1$ remain above the activation threshold. The grey area in figure C.1(b) is equal to the probability of a neuron not responding for the familiar stimulus. Denote this probability by q and calculate it:

$$q = \Pr\left(\theta\left(-a, \sqrt{\frac{Pa}{N}}\right) < -\frac{a}{2}\right) = \Pr\left(\theta(0,1) < \sqrt{\frac{Na}{4P}}\right) = \Phi\left(\sqrt{\frac{Na}{4P}}\right). \tag{C.2}$$

In this model, the decision function is simply the number of active novelty neurons. Hence the decision function for familiar stimuli has mean Nq and variance $Nq(1-q)$, and for novel stimuli it has mean $N(1-q)$ and the same variance. Therefore, as a familiarity criterion we may take the number of active novelty neurons being lower than $N/2$ (because

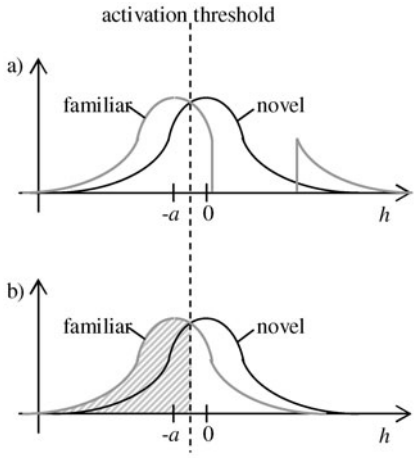

Figure C.1. Distribution of the membrane potentials of the novelty neurons in the double-threshold model (simplified version of the Sohal and Hasselmo (2000) model) for novel and familiar stimuli. Description as in figure 2. (a) Distributions of the membrane potentials before the approximation: the grey curve shows the distribution for familiar stimuli, the black for novel stimuli. (b) Distributions of the membrane potentials after approximation. The hatched area is equal to the probability of the neuron not responding to a familiar stimulus.

$N/2 = 1/2(Nq+N(1-q)))$. Following appendix A, the equation giving the network capacity is

$$\Pr\left(\theta(N(1-q), \sqrt{Nq(1-q)}) < \frac{N}{2}\right) = 0.99. \tag{C.3}$$

Solving equation (C.3) with respect to q as in appendix (A.1), we get

$$q = \frac{1}{2} + \sqrt{\frac{5.41}{4N}}. \tag{C.4}$$

According to equation (C.4), q is very close to $1/2$, and according to equation (C.2), q is equal to the cumulative normal distribution Φ. Therefore, we may approximate $\Phi(x) \approx 1/2+x/\sqrt{2\pi}$, and get

$$q = \frac{1}{2} + \sqrt{\frac{Na}{8\pi P_{\max}}} = \frac{1}{2} + \sqrt{\frac{5.41}{4N}}. \tag{C.5}$$

Solving equation (C.5) with respect to $P_{\max}$, we get $P_{\max} \approx 0.03N^2a$. This result shows that for uncorrelated input patterns, the double-threshold model achieves a capacity proportional to the number of synapses in the network, but that this capacity decreases with the sparseness of coding.

References

Aggleton J P and Brown M W 1999 Episodic memory, amnesia and the hippocampal-anterior thalamic axis *Behav. Brain Sci.* **22** 425–98

Aggleton J P and Shaw C 1996 Amnesia and recognition memory: a re-analysis of psychometric data *Neuropsychologia* **34** 51–62

Amit D J 1989 *Modelling Brain Function* (Cambridge: Cambridge University Press)

Barnes C A, McNaughton B L, Mizumori S J Y, Leonard B W and Lin L-H 1990 Comparison of spatial and temporal characteristics of neuronal activity in sequential stages of hippocampal processing *Prog. Brain Res.* **83** 287–300

Bliss T V P and Collingridge G L 1993 A synaptic model of memory: long-term potentiation in hippocampus *Nature* **361** 31–9

Bogacz R, Brown M W and Giraud-Carrier C 1999 High capacity neural networks for familiarity discrimination *Proc. Int. Conf. on Artificial Neural Networks '99 (Edinburgh)* pp 773–6

Bogacz R, Brown M W and Giraud-Carrier C 2001 Model of familiarity discrimination in the perirhinal cortex *J. Comput. Neurosci.* **10** 5–23

Bogacz R 2001 Computational models of familiarity discrimination in the perirhinal cortex *PhD Thesis* University of Bristol (http://www.math.princeton.edu/∼rbogacz)

Bogacz R and Brown M W 2002 Comparison of computational models of familiarity discrimination in the perirhinal cortex *Hippocampus* at press

Brown M W and Aggleton J P 2001 Recognition memory: what are the roles of the perirhinal cortex and hippocampus? *Nat. Rev. Neurosci.* **2** 51–62

Brown M W, Wilson F A W and Riches I P 1987 Neuronal evidence that inferotemporal cortex is more important than hippocampus in certain processes underlying recognition memory *Brain Res.* **409** 158–62

Brown M W and Xiang J Z 1998 Recognition memory: neuronal substrates of the judgement of prior occurrence *Prog. Neurobiol.* **55** 149–89

Brown M W, Fahy F L and Zhu X O 1996 Studies of the recognition memory system *Perception, Memory and Emotion: Frontiers in Neuroscience* ed T Ono, B L McNaughton, S Molotchnikoff, E T Rolls and H Nishijo (Oxford: Elsevier)

Buffalo E A, Reber P J and Squire L R 1998 The human perirhinal cortex and recognition memory *Hippocampus* **8** 330–9

Buckingham J and Willshaw D 1993 On setting unit thresholds in an incompletely connected associative net *Network: Comput. Neural Syst.* **4** 441–59

Buhmann J, Divko R and Schulten K 1989 Associative memory with high information content *Phys. Rev.* A **39** 2689–92

Dayan P S and Willshaw D 1991 Optimising synaptic learning rules in linear associative memories *Biol. Cybern.* **65** 253–65

Eichenbaum H, Otto T and Cohen N J 1994 Two functional components of the hippocampal memory system *Behav. Brain Sci.* **17** 449–518

Erickson C A, Jagadeesh B and Desimone R 2000 Clustering of perirhinal neurons with similar properties following visual experience in adult monkey *Nature Neurosci.* **3** 1143–8

Fahy F L, Riches I P and Brown M W 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** 457–72

Gerstner W 1998 Spiking neurons *Pulsed Neural Networks* ed W Maas and C Bishop (Cambridge, MA: MIT Press)

Graham B and Willshaw D 1995 Improving recall from an associative memory *Biol. Cybern.* **72** 337–46

Hertz J, Krogh A and Palmer R G 1991 *Introduction to the Theory of Neural Computations* (Redwood City, CA: Addison-Wesley)

Hopfield J J 1982 Neural networks and physical systems with emergent collective computational abilities *Proc. Natl Acad. Sci. USA* **79** 2554–8

Ito M 1989 Long-term depression *Ann. Rev. Neurosci.* **12** 85–102

Kemp N and Bashir Z I 2001 Long-term depression: a cascade of induction and expression mechanisms *Prog. Neurobiol.* **65** 339–65

Kohonen T, Oja E and Ruohonen M 1974 Adaptation of a linear system to a finite set of patterns occurring in an arbitrarily varying order *Acta Polytechn. Scand. Electr. Eng.* **25**

Kreiman G, Koch C and Fried I 2000 Category-specific visual responses of single neurons in the human medial temporal lobe *Nat. Neurosci.* **3** 946–53

Li L, Miller E K and Desimone R 1993 The representation of stimulus familiarity in anterior inferior temporal cortex *J. Neurophysiol.* **69** 1918–29

Marr D 1971 Simple memory: a theory of archicortex *Phil. Trans. R. Soc.* B **262** 23–81

McCulloch W S and Pitts W 1943 A logical calculus of ideas immanent in nervous activity *Bull. Math. Biophys.* **5** 115–33

Miller E K, Li L and Desimone R 1993 Activity of neurons in anterior inferior temporal cortex during short-term memory task *J. Neurosci.* **13** 1460–78

Murray E A 1996 What have ablation studies told us about the neural substrates of stimulus memory? *Semin. Neurosci.* **8** 13–22

Murray E A and Bussey T J 1999 Perceptual-mnemonic functions of the perirhinal cortex *Trends Cogn. Sci.* **3** 142–51

Norman K A and O'Reilly R 2001 Modelling hippocampal and neocortical contributions to recognition memory: a complementary learning systems approach *Technical Report* 01-02 University of Colorado, Boulder

Palm G and Sommer F T 1992 Information capacity in recurrent McCulloch–Pitts networks with sparsely coded memory states *Network: Comput. Neural Syst.* **3** 177–86

Riches I P, Wilson F A and Brown M W 1991 The effects of visual stimulation and memory on neurons of the hippocampal formation and neighbouring parahippocampal gyrus and inferior temporal cortex of the primate *J. Neurosci.* **11** 1763–79

Sobotka S and Ringo J L 1993 Investigation of long-term recognition and association memory in unit responses from inferotemporal cortex *Exp. Brain Res.* **96** 28–38

Sohal V S and Hasselmo M E 2000 A model for experience-dependent changes in the responses of infero-temporal neurons *Network: Comput. Neural Syst.* **11** 169–90

Treves A and Rolls E T 1994 A computational analysis of the role of the hippocampus in memory *Hippocampus* **4** 374–91

Willshaw D, Buneman O and Longuet-Higgins H 1969 Non-holographic associative memory *Nature* **222** 960–2

Willshaw D and Dayan P S 1990 Optimal plasticity from matrix memories: What goes up must come down *Neural Comput.* **1** 85–93

Xiang J Z and Brown M W 1998 Differential neuronal encoding of novelty, familiarity and recency in regions of the anterior temporal lobe *Neuropharmacology* **37** 657–76