Network: Computation in Neural Systems
September 2008; 19(3): 161–182

informa
healthcare

Computational models can replicate the capacity of human recognition memory

ZACHARIAS ANDROULIDAKIS[1], ANDREW LULHAM[1,2], RAFAL BOGACZ[1], & MALCOLM W BROWN[2]

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

(Received 16 April 2008; revised 1 August 2008; accepted 8 August 2008)

Abstract
The capacity of human recognition memory was investigated by Standing, who presented several groups of participants with different numbers of pictures (from 20 to 10 000), and subsequently tested their ability to distinguish between previously presented and novel pictures. The estimated number of pictures retained in recognition memory by different groups when plotted as a logarithmic function of the number of pictures presented formed a straight line, representing a power–law relationship. Here, we investigate if published models of familiarity discrimination can replicate Standing's results. We first consider a simplified assumption that visual stimuli are represented by uncorrelated patterns of firing of visual neurons providing input to the familiarity discrimination network. We show that for this case three models (Familiarity discrimination based on Energy (FamE), Anti-Hebbian and Info-max) can reproduce the observed power–law relationship when their synaptic weights are appropriately initialized. For more realistic assumptions on neural representation of stimuli, the FamE model is no longer able to reproduce the power–law relationship in simulations, while the Anti-Hebbian and Info-max can reproduce it. Nevertheless, the slopes of the power–law relationships produced by the models in all simulations differ from that observed by Standing. We discuss possible reasons for this difference, including separate contributions of familiarity and recollection processes, and describe experimentally testable predictions based on our analysis.

Keywords: *capacity, familiarity discrimination, recognition memory*

Correspondence: Rafal Bogacz, Department of Computer Science, University of Bristol, Bristol, BS8 1UB, UK. Tel: +44-117-954-5141. Fax: +44-117-954-5208. E-mail: R.Bogacz@bristol.ac.uk

ISSN 0954-898X print/ISSN 1361-6536 online/08/030161–182 © 2008 Informa UK Ltd.
DOI: 10.1080/09548980802412638

Introduction

Recognition memory is defined in psychology as a type of memory that allows us to judge if a stimulus has been encountered before. From everyday experience we know that our recognition memory has very high capacity, and we can often recognize someone as familiar even if we cannot recollect the details, such as the name, of that person. The question of the capacity of human recognition memory has been investigated in psychology for many decades (Nickerson 1965; Shepard 1967; Standing et al. 1970). But the most detailed and surprizing results were provided by Standing (1973). We focus on these results in this article.

Standing presented different numbers of natural images to different groups of participants. One group was presented with 10 000 pictures. Each picture was presented only once for 5 s. Two days after learning, each participant performed a recognition test. On each trial the participant was shown two pictures and had to decide which was novel and which was presented before. The participants who saw 10 000 pictures achieved an accuracy of 83%. Furthermore, Standing estimated the number of pictures retained in memory R from the following formula:

$$R = P(1 - 2E). \tag{1}$$

In Equation 1, P denotes the number of stimuli presented during learning, and E denotes the error rate on test (note that if participants guess, then $E = 0.5$, and Equation 1 gives $R = 0$). The solid line in Figure 1 shows the number of pictures

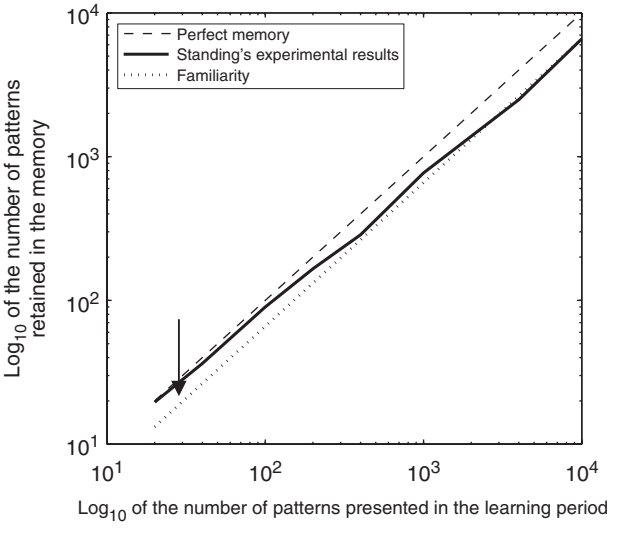

Figure 1. Number of items retained (R) in memory as a function of the number presented (P) during learning in Standing's experiment (Standing 1973). The dashed line corresponds to perfect recognition memory, where all of the presented stimuli are retained. The solid line corresponds to the experimental results found by Standing. The leftmost experimental data point overlaps with the perfect memory line, since for very few presented stimuli participants were able to discriminate familiarity perfectly. For larger numbers of stimuli, the proportion of presented stimuli that are retained is reduced. The dotted line is an example of a relation predicted by the FamE model with randomly initialized weights. The arrow indicates where the predictions do not match experimental data (see text).

retained (R), as a function of the number of pictures presented (P) to each group of participants. In Figure 1 the axes are logarithmic, and the relation between R and P forms a straight line, implying a power–law relation. The straight line in Figure 1 also suggests that there is no sign of saturation in human recognition memory even after seeing 10 000 pictures (Standing 1973).

Since Standing's study, great advances have been achieved in understanding recognition memory. First, it has been proposed that recognition involves two separate processes, recollection and familiarity (Mandler 1980), and this dual-process model is supported by a large number of behavioural studies (for review, see Yonelinas 2002). Furthermore, many experimental studies strongly indicate that the recollective component is dependent on the hippocampus, while the familiarity component is dependent on the perirhinal cortex (for reviews, see Eichenbaum et al. 1994; Brown and Xiang 1998; Murray and Bussey 1999; Brown and Aggleton 2001). Thus participants can use two strategies to recognize a stimulus as previously seen: they can recollect the episode of seeing the stimulus (dependent on the hippocampus), or have a feeling of familiarity (dependent on the perirhinal cortex).

Computational models have been developed both for the hippocampus and the perirhinal cortex, and their storage capacities for recollection and familiarity, respectively, have been calculated. In particular, a fully connected neural network with N neurons has capacity for recollection proportional to N (Amit 1989), while a network of the same size specialized for familiarity discrimination can perform it for an order of N^2 stimuli (Bogacz et al. 2001). This difference can be intuitively understood by noticing that familiarity is a much easier task than recollection. For example, if one wants to recollect some details of an encountered person, this information may be represented in an auto-associative memory by N neurons, so will contain an order of N bits of information. By contrast, if one instead wants to determine the familiarity of an encountered person, one only makes a binary decision (novel or familiar), and hence only one bit of information is required.

Estimates of the capacity of the human perirhinal cortex based on the computational models show that it could potentially discriminate familiarity for thousands of times more stimuli than the hippocampus could recollect (Bogacz and Brown 2003). The estimates of the capacity of the human perirhinal cortex suggest that humans should be able to discriminate familiarity even for numbers of stimuli orders of magnitude higher than those tested by Standing (Bogacz and Brown 2003). Besides its potentially greater capacity, familiarity discrimination is typically less effortful than recall, hence it is plausible to assume that the participants of Standing's experiment relied primarily on the familiarity process when discriminating 10 000 pictures. Given this assumption, it is interesting to investigate whether models of familiarity discrimination reproduce the power–law relation shown in Figure 1. This is the question addressed in this article.

The determination of whether models of familiarity discrimination can reproduce Standing's power–law relation, should verify (or falsify), constrain and help to distinguish between currently proposed models. So far, eight[1] such models have been published (Brown and Xiang 1998; Sohal and Hasselmo 2000; Bogacz et al. 2001; Bogacz and Brown 2003; Norman and O'Reilly 2003; Meeter et al. 2005; Norman et al. 2005; Lulham et al. 2006). These models have similar levels of detail of description (at the network level), but differ in the proposed rules for synaptic plasticity, and the way the familiarity signal is read out from the network.

Five of these models have been compared in detail with respect to their performance and consistency with experimental data, including mechanisms of synaptic plasticity and responses of perirhinal neurons (Bogacz and Brown 2003). One important factor that has proved a stumbling block for some of the models was that real input patterns to the perirhinal network are likely to have a correlation structure (Erickson et al. 2000). Introducing such correlation greatly reduces the capacity of some of the models. The ability to fit Standing's data has not previously been investigated and may further distinguish between the current models.

The analysis of this article makes a number of counterintuitive experimental predictions which are described in the Discussion. Since the following 'Methods' and 'Results' sections are technical we here summarize these sections, to allow Readers without a background in neural network modelling to follow the predictions in the 'Discussion'.

Summary of results

We simulated performance in Standing's experiment for three computational models: (i) FamE (Bogacz et al. 2001) – an abstract model whose simplicity allows mathematical analysis, (ii) Anti-Hebbian model (Kohonen et al. 1974; Brown and Xiang 1998; Bogacz and Brown 2003) – a model that uses synaptic weakening, and which achieved the best performance on realistic inputs in the comparison study of Bogacz and Brown (2003) and is consistent with experimental data (Brown and Bashir 2002), (iii) Info-max (Bell and Sejnowski 1995, 1997) – a well-known feature extraction model, that has been recently shown to also perform familiarity discrimination efficiently (Lulham et al. 2006). The Anti-Hebbian and Info-max models were chosen, as these models achieved the best performance in our previous studies, and hence have most potential to reproduce Standing's results, and were compared to the analytically tractable FamE model. In the first set of tests uncorrelated input patterns were used. The models were then tested using correlated patterns.

In initial simulations using uncorrelated patterns, the Anti-Hebbian and Info-max models, but not FamE, reproduced Standing's power law. We identified that the element of the FamE model that prevented it from reproducing the power law was its over-simplistic initialization of synaptic weights. When the weights in FamE were initialized as in other models, it also reproduced the power law.

The relation between the numbers of stimuli retained (R) and presented (P) for FamE with proper weight initialization is well approximated by:

$$R = Pf(\eta). \tag{2}$$

In Equation 2, η denotes the learning rate, i.e. the magnitude of synaptic weight modification after a presentation of a stimulus during learning, and f denotes a monotonic function ($f(\eta) = 0$ for $\eta = 0$, and then $f(\eta)$ increases towards 1, as η increases to infinity). Equation 2 is satisfied for larger numbers of neurons (N). In particular, it is satisfied for $N \geq 300$ for P up to 10 000. Note that N in the human perirhinal cortex is much larger.

Equation 2 implies that the number of stimuli retained is linearly proportional to the number of stimuli presented. This comes from the following property of the FamE model: for a given stimulus presented there is a probability that the weights

are modified sufficiently to recognize it as familiar, and this probability depends on η but not on P. The independence of this encoding probability from P implies that other stimuli presented during learning do not interfere with the memory of the given stimulus. Although the FamE model includes interference, its capacity for uncorrelated patterns is so large that the effect of the interference is negligible.

To obtain the relation between R and P in Figure 1 with logarithmic axes, we take the logarithm of Equation 2,

$$\log R = \log P + \log f(\eta). \tag{3}$$

Equation 3 implies the power law relation produced by the FamE model has a slope of 1, as shown by the dotted line in Figure 1. Furthermore, the position of the line is changed by the learning rate, i.e. it is shifted up by increasing η, and shifted down by decreasing η.

Despite this qualitative match between experimental and simulated results, the slope of the stimulated results produced by the FamE model could not be matched to that of standing's data. Compare the solid and the dotted lines in Figure 1, corresponding to experimental data and model predictions, respectively. Although the lines overlap for large P, for lower P experimental participants achieved better performance than predicted by the model, as indicated by an arrow in Figure 1. This is because the participants made fewer errors for low P, while in the FamE model the error rate is independent of P.

The error rate of the Anti-Hebbian and Info-max models does increase with P, but to a much lower extent than in the experimental data. Hence the interference between stimuli in these models, although small, cannot be ignored. Consequently, these models produce the power law with a slope slightly closer to the experimental data.

When tested on patterns with a realistic value of correlation between inputs (Erickson et al. 2000), the FamE model was unable to reproduce Standing's power law. By contrast, the performance of Anti-Hebbian and Info-max models was little affected by the patterns correlation.

Methods

Models of familiarity discrimination

In this section, we describe the three models simulated in this work.

FamE. It is an abstract model of familiarity discrimination whose basic description is given below. It does not reproduce known data on the neurobiology of perirhinal cortex, but equivalent computations can also be performed by a biologically plausible network (details given in Bogacz et al. 2001, not restated here). In this abstract model the familiar patterns are stored in a Hopfield (1982) network, and the discrimination of familiarity of a test pattern is performed by computing the value of the energy of the network for the test pattern, as we describe in detail below.

The Hopfield network is a fully connected recurrent neural net consisting of N neurons. The activations of these neurons are denoted by x_i and can take the values 1 or -1 for active or inactive states, respectively. It is assumed that the stimuli are represented by binary patterns of length N. The stored patterns (corresponding to

the familiar stimuli presented during learning) are denoted by ξ^{μ} and their number by P. Finally, the weight of a connection between neurons i and j is denoted by w_{ij} and computed from the Hebb rule (Hertz et al. 1991)[2]:

$$w_{ij} = \begin{cases} \frac{1}{N}\sum_{\mu=1}^{P} \xi_i^{\mu}\xi_j^{\mu}, & \text{for } i \neq j, \\ 0, & \text{for } i = j \end{cases}. \tag{4}$$

During familiarity discrimination of a test pattern the states of the Hopfield network x_i are set to this pattern and the energy function (Hopfield 1982) is computed:

$$H(x) = -\frac{1}{2}\sum_{i=1}^{N} x_i \sum_{j=1}^{N} x_j w_{ij}. \tag{5}$$

Note that in the FamE model (unlike in the Hopfield model) relaxation of the network is not performed.[3] The value of the energy function is lower for stored patterns than others (Hopfield 1982): in particular, its average is equal to $-N/2$ for stored and to 0 for novel patterns.

The Anti-Hebbian. This model was originally proposed as an engineering solution for the novelty detection problem (Kohonen et al. 1974), long before its potential neural bases were discovered. It was then proposed as a model for the perirhinal cortex (Brown and Xiang 1998) and later formalized (Bogacz and Brown 2002, 2003).

The Anti-Hebbian model is a fully connected feed-forward network, shown in Figure 2. During the weight initialization, all the weights w_{ij} are randomly generated from a uniform distribution between -0.5 and 0.5 and then normalized such that for each neuron the average is 0 and the Euclidian length of the vector of weights is 1. During the learning phase, input neurons x are set to the pattern being learnt and the membrane potentials of the novelty neurons are computed from:

$$h_i = \sum_{j=1}^{N} w_{ij} x_j. \tag{6}$$

The activities of the neurons are determined using the 'k-winners' method with $k = N/2$ (Bogacz and Brown 2003). In particular, the activation values y_i of the half

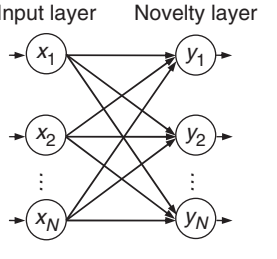

Figure 2. Architectures of the Anti-Hebbian and Info-max models. Circles denote neurons and arrows denote connections. The network is fully connected, with a layer containing novelty neurons receiving feed-forward projections from a layer of input neurons.

of the neurons with the highest membrane potential h_i are set to 1 and all other activation values are set to -1.

After computing activities, all the weights in the network are updated. The main feature of the Anti-Hebbian model is that the weights that connect two active neurons are decreased (and hence the name of the model). In particular, the weights are modified according to the following learning rule which makes the weights between two active neurons smaller and between one active output neuron and one inactive input neuron larger (if both of the neurons are inactive there is no change):

$$\Delta w_{ij} = -\frac{\eta}{2N}(y_i + 1)x_j. \tag{7}$$

Recall that η denotes the learning rate. After learning a pattern, all the weights are renormalized.

The weight modification of Equation 7 reduces the average response of the neurons during following presentations of pattern x, which is the main response property of perirhinal novelty neurons (Brown et al. 1987; Li et al. 1993; Sobotka and Ringo 1994; Brown and Xiang 1998). Thus the total activity level of the neurons can be used to determine if a presented pattern is novel or familiar. Accordingly, familiarity discrimination of a test pattern is achieved by setting the network input to this pattern, computing the activity of the neurons, as described above, and evaluating the following decision function:

$$d(x) = \sum_{i=1}^{N} y_i h_i. \tag{8}$$

This decision function is lower for familiar patterns than for novel ones.

Info-max. It was originally proposed as an iterative algorithm performing Independent Component Analysis (ICA) (Bell and Sejnowski 1995). ICA finds the underlying features of a set of mixed signals, which can then be used to unmix them. More recently, it has been shown that using an Info-max learning rule in a fully connected feed-forward network, as shown in Figure 2, its algorithm can also be used non-iteratively to perform familiarity discrimination (Lulham et al. 2006).

The weights of the network are initialized as for the Anti-Hebbian model. All the weights w_{ij} are randomly generated from a uniform distribution between -0.5 and 0.5 and then normalized such that for each neuron the standard deviation (SD) of the weights is 1 and the mean is 0. During the learning phase input neurons x are set to the pattern being learnt, the membrane potentials of the novelty neurons are computed from Equation 6, and the activities of the neurons are computed as $y_i = \tanh(h_i)$. Subsequently, the weights of the neurons are modified according to the following rule:

$$\Delta w_{ij} = \frac{\eta}{N}\left(\left[w_{ij}^T\right]^{-1} - 2y_i x_j\right). \tag{9}$$

The second term on the right-hand side is an 'Anti-Hebbian' term, since it causes connections between active input neurons and active output neurons to be weakened. It also causes connections between inactive input neurons and active novelty neurons to be strengthened, again as in the Anti-Hebbian model. Hence the

two models are similar. Info-max is additionally able to perform feature extraction, a process also believed to occur in the perirhinal cortex (Murray and Bussey 1999). The first term in the learning rule requires the computation of the inverse of a matrix containing all of the synaptic weight information for the network, which would be difficult to perform within a biologically plausible network. Nevertheless, several models have been proposed which converge to the same weights as Info-max, while having more biologically plausible learning rules (Olshausen and Field 1996; Amari and Cichocki 1998; Waydo and Koch 2008).

A decision on the familiarity of a given stimulus is based on the average activity of the novelty neurons (the decision function is equal to the sum of the absolute values of the h^i). Consistent with biological constraints, this value is large for novel stimuli and small for familiar stimuli.

Simulation methods

A MATLAB toolbox for performing the simulations of Standing's experiment has been developed, and is available online at http://www.cs.bris.ac.uk/home/lulham/toolbox/. Included with the toolbox are MATLAB implementations of all of the models of familiarity discrimination that have been tested here.

For each of the models described in the 'Models of Familiarity Discrimination' section we simulated the tests that Standing (1973) performed with human participants, making eight simulations corresponding to the eight groups of participants in the experiment.

In the learning phase, the number of input patterns P used during each simulation was the same as Standing used in his experiment (20, 40, 100, 200, 400, 1000, 4000 or 10 000 patterns). The learning phase consists of three stages, which are repeated for each pattern. First, the pattern is presented to a given network. Second, the membrane potentials of novelty neurons are computed, and from these the activities of neurons can be determined. Finally, the weights from inputs to novelty neurons are modified according to the learning rule.

In the test phase, the number of simulated test trials was the same as in Standing's experiment (20, 40, 80, 80, 80, 80, 160 and 160 patterns). At each trial a pattern from the learning phase and a novel one were used. The network computed the decision or energy function for each one of the patterns and the two were compared in order to decide which of the two patterns was more familiar. At the end of the test phase an error rate was computed.

To get a closer estimate of mean error rates, the above procedure was repeated 40 times for each of the models. Then for each model and for each simulated group of participants (indexed by t) we computed the average error rate in the test phase $\mu(E^t_{\text{Simulation}})$ and the SD of the error rates across repetitions $\sigma(E^t_{\text{Simulation}})$.

Methods of pattern generation

Two types of patterns were used when testing the recognition capacity of the models, which will herein be referred to as uncorrelated and correlated. Both types comprise of vectors of length N with entries equal to 1 or -1. However, the uncorrelated patterns are randomly and independently generated.

The correlated patterns are biased towards a randomly generated binary template pattern, in order to introduce correlation between input neurons. For each bit of a pattern, the probability of that bit equalling the corresponding template bit is $½ + ½b$, where b is the parameter controlling the bias. For patterns generated in this way, the correlation r_{ij} between a pair of inputs is equal to b^2 or $-b^2$ (Bogacz and Brown 2003). All correlated patterns used here have bias 0.2, corresponding to an absolute correlation value of 0.04. This correlation value was chosen because Bogacz and Brown (2003) estimate that it corresponds to the level of correlation between distant perirhinal neurons observed by Erickson et al. (2000).

Estimation of learning rate

The error rates produced by the Anti-Hebbian and Info-max models depend on the value of the learning rate parameter η. For each model the value of η chosen was that giving the best fit to Standing's data. In particular, η was chosen to minimize the following cost function (Bogacz and Cohen 2004):

$$\text{Cost} = \sum_{t=1}^{8} \left(\frac{E^t_{\text{Standing}} - \mu\left(E^t_{\text{Simulation}}\right)}{\sigma\left(E^t_{\text{Simulation}}\right)} \right)^2. \tag{10}$$

In Equation 10, E^t_{Standing} is the error rate in Standing's experiment for group t. Simulations that gave average error rate $\mu(E^t_{\text{Simulation}}) = 0$ for any $P > 20$ were not evaluated because they were considered implausible. Simulations that gave $\mu(E^t_{\text{Simulation}}) > 0$ for $P > 20$ and any $\mu(E^t_{\text{Simulation}})$ for $P = 20$ were accepted as plausible, since it is very easy for the network to obtain 100% accuracy in the $P = 20$ case. If a model achieved 100% accuracy for $P = 20$, we did not take the simulations for $P = 20$ into consideration in the cost function.

Results

FamE

In this section, we derive analytically the average number of items retained in memory by the FamE model when presented with uncorrelated patterns. To simplify calculation we define a new decision function $d(x) = -2H(x)$. Now we calculate the probability of correct discrimination on a single test trial of a simulated version of Standing's experiment.

Bogacz et al. (2001) have shown that after presentation of a sample familiar pattern, the decision function has an approximately normal distribution with mean N and SD equal to $\sqrt{2P}$, which we denote $d(\xi^1) = \theta(N, \sqrt{2P})$; for a novel pattern $d(x^{\text{new}}) = \theta(0, \sqrt{2P})$.

In each simulated test trial of Standing's experiment, the model has to decide which of the two pictures is familiar and which is novel. The probability (Pr) of the model making the correct choice is equal to the probability of the decision function for a novel pattern, which is sampled from $\theta(0, \sqrt{2P})$, being smaller

than the decision function for a familiar pattern, which is sampled from $\theta(N, \sqrt{2P})$:

$$\Pr(\text{correct}) = \Pr\Big(\theta\Big(0, \sqrt{2P}\Big) < \theta\Big(N, \sqrt{2P}\Big)\Big). \tag{11}$$

Using elementary properties of random variables, we obtain:

$$\begin{aligned}\Pr(\text{correct}) &= \Pr\Big(\theta\Big(0, \sqrt{2P}\Big) - \theta\Big(N, \sqrt{2P}\Big) < 0\Big) \\ &= \Pr\Big(\theta\Big(-N, \sqrt{4P}\Big) < 0\Big) \\ &= \Pr\Big(\theta\Big(0, \sqrt{4P}\Big) < N\Big) = \Pr\Big(\theta(0,1) < \frac{N}{\sqrt{4P}}\Big) \\ &= \text{normcdf}\Big(\frac{N}{\sqrt{4P}}\Big).\end{aligned} \tag{12}$$

In Equation 12, normcdf denotes the normal standard cumulative distribution function. Since the error rate, $E = 1 - \Pr(\text{correct})$, we can substitute Equation 12 into Equation 1 and obtain:

$$R = P\Big(2\text{normcdf}\Big(\frac{N}{\sqrt{4P}}\Big) - 1\Big). \tag{13}$$

Figure 3 plots the number of patterns retained by the FamE model computed from Equation 13 and from simulations. It shows that the predictions of Equation 13 match the simulations very closely. However, the relationship between R and P of this model differs qualitatively from Standing's results. For small P all patterns are retained in memory (note that the model curves initially overlap with perfect memory), before the memory saturates. As N increases, the number of patterns P

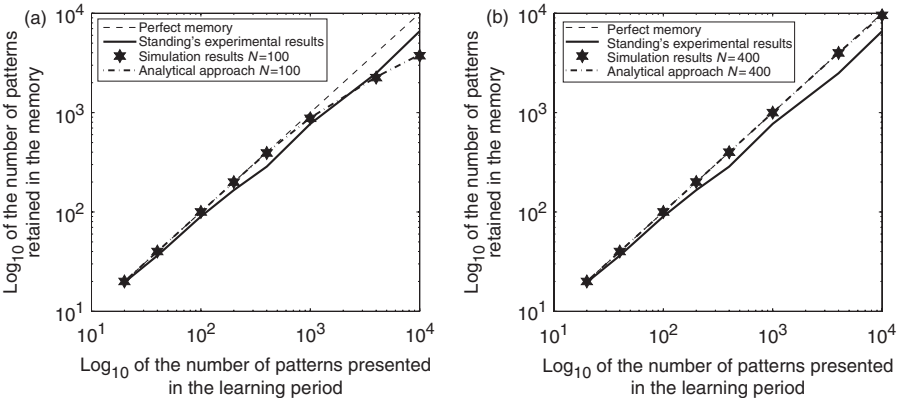

Figure 3. Number of items retained in the memory predicted by the FamE model. The model is trained and tested on uncorrelated patterns. Dashed-dotted lines show predictions of Equation 13, and stars show results of simulations (the predictions of Equation 13 match simulations very closely). For comparison, solid lines show Standing's result and dashed lines correspond to perfect recognition memory. (a) and (b) correspond to network of 100 and 400 neurons, respectively. As the size of the networks increases, the model converges to give perfect recognition memory.

after which the network starts to make mistakes also increases. For large N (the biologically more realistic case), the FamE model predicts that all of the 10 000 stimuli should be retained in memory (Figure 3b), and participants should not make any mistakes at all. This result is clearly inconsistent with Standing's data.

Anti-Hebbian

We could not derive R analytically for the Anti-Hebbian model, because Equation 7 describing the weight update is iterative (it includes term y_i which depends on previous learning iterations) and, consequently, it is not possible to derive a simple analytical formula for weights after multiple learning episodes (unlike for FamE, which is given by Equation 4). For this reason, Figure 4 only includes the numbers of stimuli retained by the Anti-Hebbian model as obtained from simulations.

Figure 4(a) shows that for uncorrelated patterns, Standing's results are not reproduced by networks with few neurons. However, the network with 500 neurons can reproduce the power law of Standing's experiment in an accurate way (Figure 4b), and the relationship between R and P produced by the Anti-Hebbian model forms a straight line.

It has also been shown that as the learning rate grows, the number of stimuli retained in memory increases and converges to the number of the items presented during learning (results not given here, but see Androulidakis 2007).

Info-max

As for the Anti-Hebbian model, it is not trivial to derive R analytically for the Info-max model since it also uses an iterative weight update equation.

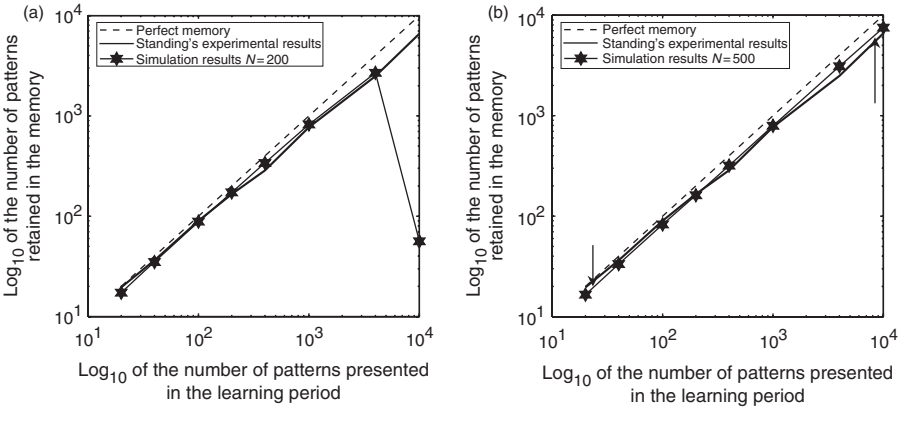

Figure 4. Number of items retained in the memory predicted by the Anti-Hebbian model. The model is trained and tested on uncorrelated patterns. Stars show results of simulations. For comparison, solid lines show Standing's result and dashed lines correspond to perfect recognition memory. (a) and (b) correspond to networks of 200, and 500 neurons, respectively. The learning rates for (a) and (b) were 0.21 and 0.10, respectively. (b) shows that for small P the model performs worse than experimental participants, but for large P the model performs better (indicated by arrows).

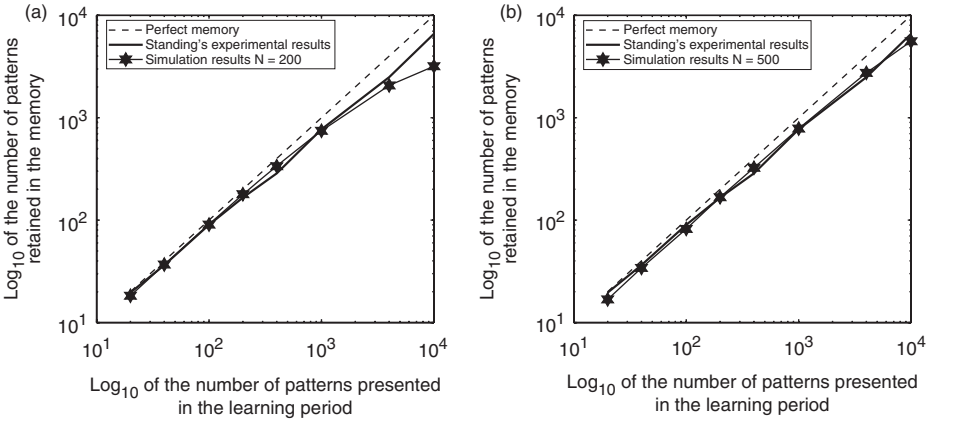

Figure 5. Number of items retained in the memory predicted by the Info-max model. The model is trained and tested on uncorrelated patterns. Stars show results of simulations. For comparison, solid lines show Standing's result and dashed lines correspond to perfect recognition memory. (a) and (b) correspond to networks of 200 and 500 neurons respectively. The learning rates for (a) and (b) were 0.0036 and 0.0012, respectively. The model performs worse than experimental participants for both small and large values of P.

Therefore, Figure 5 gives only simulation data and we draw our conclusions about the model from these.

As with the Anti-Hebbian model, we see that networks with few neurons are unable to reproduce Standing's power law. However, there is again a trend towards the power law as the number of neurons in the network is increased. For a network of 500 neurons, we see a straight line except for the final point in the graph. We surmize that for networks of biologically realistic size, the relationship between P and R will converge to a power law.

FamE with initialized weights

The FamE and Anti-Hebbian models differ in a number of aspects, including weight initialization, and weight renormalization after learning (present only in the Anti-Hebbian model). It is of interest to determine which aspect is critical for fitting Standing's data. The effect of the two aspects on the fit of the FamE model has been investigated (Androulidakis 2007). It was found that only one of them – weight initialization – improved the fit of the FamE model.

In its original description the weights of the FamE model are initialized to 0 (Equation 4), which is an implausible assumption as it would correspond to participants' recognition memory stores being empty at the start of the experiment. Here we analyse a modified version of the FamE model in which the initial weights are set to values sampled from a normal distribution with mean 0 and SD 1. Additionally, we assume that after presentation of each learning pattern, the weights are modified in proportion to a learning rate η. Thus, the weights w' in the modified FamE model are equal to:

$$w'_{ij} = \eta w_{ij} + \theta(0, 1). \tag{14}$$

In Equation 14, w denotes the weights of the original FamE model (Equation 4). In the modified model, the decision function after presentation of pattern x is given by:

$$\begin{aligned} d'(x) &= \sum_{i=1}^{N} x_i \sum_{j=1}^{N} x_j w'_{ij} \\ &= \eta \sum_{i=1}^{N} x_i \sum_{j=1}^{N} x_j w_{ij} + \sum_{i=1}^{N} \sum_{j=1}^{N} x_i x_j \theta(0,1). \end{aligned} \tag{15}$$

The bottom line of Equation 15 includes two terms. The first term is equal to the value of the decision function in the original FamE model scaled by the learning rate. The second term is a sum of N^2 random variables with mean 0 and SD 1 (note that $x_i x_j$ is equal to 1 or -1, so does not influence mean or SD). Hence according to the central limit theorem, the second term has normal distribution with mean 0 and SD N; thus:

$$d'(x) = \eta d(x) + \theta(0, N). \tag{16}$$

In Equation 16, $d(x)$ denotes the value of the decision function of the original FamE model. Recall from 'FamE', that $d(x)$ has distribution $\theta(N, \sqrt{2P})$ for familiar patterns and $\theta(0, \sqrt{2P})$ for novel. Thus, the decision function of the modified FamE model has the following distribution for a familiar pattern:

$$d\left(\xi^1\right) = \eta\theta\left(N, \sqrt{2P}\right) + \theta(0, N) = \theta\left(\eta N, \sqrt{2P\eta^2 + N^2}\right). \tag{17}$$

Analogously, for a novel pattern:

$$d(x^{\text{new}}) = \theta\left(0, \sqrt{2P\eta^2 + N^2}\right). \tag{18}$$

As in the 'FamE' section, we can compute the probability of correct discrimination at test:

$$\Pr(\text{correct}) = \Pr\left(\theta\left(0, \sqrt{2P\eta^2 + N^2}\right) < \theta\left(\eta \text{N}, \sqrt{2P\eta^2 + N^2}\right)\right). \tag{19}$$

Using the same manipulations as in Equation 12, we obtain:

$$\Pr(\text{correct}) = \text{normcdf}\left(\frac{\eta N}{\sqrt{4P\eta^2 + 2N^2}}\right). \tag{20}$$

Consequently, the number of items retained becomes:

$$R = P\left(2\text{normcdf}\left(\frac{\eta N}{\sqrt{4P\eta^2 + 2N^2}}\right) - 1\right). \tag{21}$$

Figure 6 plots the number of uncorrelated patterns retained by the modified FamE model computed from Equation 21 and from simulations. It shows that the simulations match the predictions of Equation 21 very closely. Furthermore, the relation between R and P of the modified FamE model matches that of Standing's experiment.

Moreover, for larger N, the relation between R and P becomes a straight line with slope 1 (i.e. parallel to the perfect memory line). This fact can be shown analytically.

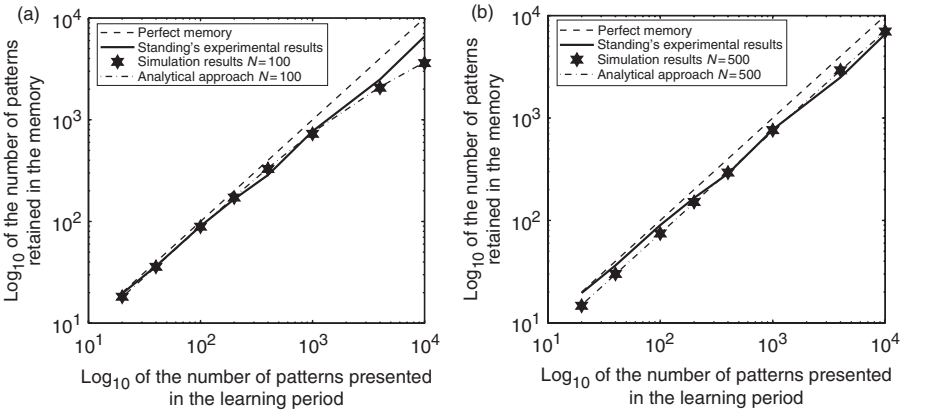

Figure 6. Number of items retained in the memory predicted by the FamE model with randomly initialized weights. The model is trained and tested on uncorrelated patterns. Dashed-dotted lines show predictions of Equation 21, and stars show results of simulations (the predictions of Equation 21 match simulations very closely). For comparison, solid lines show Standing's result and dashed lines correspond to perfect recognition memory. (a) and (b) correspond to networks of 100 and 500 neurons, respectively. The learning rates for (a) and (b) were 2.37 and 1.62, respectively. (b) shows that for small P the model performs worse than experimental participants, but for large P the model performs better.

For large N, the term $2N^2$ in the denominator of Equation 21 becomes much larger than $4P\eta^2$, and hence the latter may be ignored. Then, Equation 21 simplifies to:

$$R = P\left[2\text{normcdf}\left(\frac{\eta}{\sqrt{2}}\right) - 1\right]. \tag{22}$$

Thus, the relation between R and P has a general form given in Equation 2 which, as described in the Introduction, leads to a straight line with slope 1 on a plot with logarithmic axes (Equation 3).

Furthermore, as the learning rate increases, the line relating R and P moves up, and it converges to the line $R = P$ for large η. This happens because the content of the square brackets in Equation 22 is a monotonously increasing function of η, which converges towards 1 for large η.

Differences between model predictions and data

Although the Anti-Hebbian, the Info-max and the modified FamE models produce straight lines for larger numbers of neurons (Figures 4b, 5b and 6b), the performance of the models differs from that observed experimentally. In particular, the models retain lower numbers of stimuli than participants for low P, as indicated by an arrow in the bottom left corner of Figure 4(b). Conversely, the Anti-Hebbian and the modified FamE models retain more stimuli than participants for large P, as indicated by an arrow in the top right corner of Figure 4(b).

To investigate this difference in more detail, Figure 7(a) compares the error rates of the models with those observed in Standing's experiment. In the experiment, the error rate increased for larger P (see solid line in Figure 7a). By contrast, the error

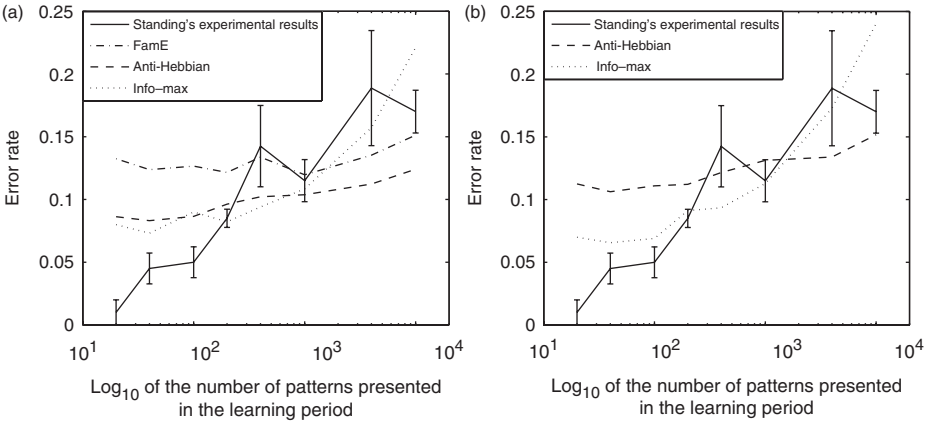

Figure 7. Comparison of error rates of participants of Standing's experiment with values predicted by the models with $N = 500$ neurons. Solid lines show Standing's experimental results, and error bars describe standard error. (a) Error rates of modified FamE, Anti-Hebbian and Info-max models for uncorrelated patterns. They are shown in non-solid lines and correspond to the data also visualized in Figures 6b, 4b, and 5b, respectively. (b) Error rates for the Anti-Hebbian and Info-max models for correlated patterns. The magnitude of the correlation used was $|r_{ij}| = 0.04$. The learning rates for the Anti-Hebbian and Info-max models were 0.098 and 0.0012, respectively.

rate of the modified FamE model did not significantly depend on P (the correlation between log P and E was not significant). As mentioned in 'Summary of results', this is a significant qualitative difference between predictions of the modified FamE model and experimental data.

The error rate of the Anti-Hebbian model does depend on P (the significance of correlation between log P and E: $p < 10^{-4}$). This difference between the FamE and the Anti-Hebbian models results from the fact that the patterns are stored in the FamE model independently from one another (note that Equation 4 implies that the order in which patterns are presented does not matter). By contrast, in the Anti-Hebbian model, after each presentation, the weights are renormalized which weakens memory traces of stimuli presented earlier. Although the predictions of Anti-Hebbian model are qualitatively more similar to the data, nevertheless the extent to which error rate varies with P is clearly much lower than observed experimentally.

The error rate of the Info-max model also varies with P (the significance of correlation between log P and E: $p < 0.01$). This variation seem to be larger than in the case of the Anti-Hebbian model, but as speculated in the end of the 'Info-max' section, the high error rate of Info-max for very large P may be a particular property of the network of the size tested (i.e. $N = 500$) and may be reduced for larger networks.

Capacity for correlated patterns

The number of correlated patterns that can be retained in memory according to the FamE model with initialized weights can also be derived analytically. Bogacz and Brown (2003) have shown that after training the FamE model on correlated

patterns, the value of the decision function for a familiar pattern has the distribution:

$$d(\xi^1) = \theta\left(N + Nr^2P, \sqrt{2P + 4Nr^3P^2}\right), \tag{23}$$

where r is the mean of the absolute value of correlation coefficients. For a novel pattern,

$$d(x^{\text{new}}) = \theta\left(Nr^2P, \sqrt{2P + 4Nr^3P^2}\right). \tag{24}$$

Following the same logic as before, the probability of correctly discriminating a familiar stimulus from a novel one becomes:

$$\begin{aligned}\Pr(\text{correct}) = \Pr\Big(&\theta\left(\eta Nr^2P, \sqrt{2P\eta^2 + 4NP^2\eta^2r^3 + N^2}\right) \\ &\leq \theta\left(\eta N + \eta Nr^2P, \sqrt{2P\eta^2 + 4NP^2\eta^2r^3 + N^2}\right)\Big).\end{aligned} \tag{25}$$

Using the same manipulations as in Equation 20, we obtain:

$$\Pr(\text{correct}) = \text{normcdf}\left(\frac{\eta N}{\sqrt{4P\eta^2 + 8NP^2\eta^2r^3 + 2N^2}}\right). \tag{26}$$

So the number of items retained for correlated patterns equals:

$$R = P\left(2\text{normcdf}\left(\frac{\eta N}{\sqrt{4P\eta^2 + 8NP^2\eta^2r^3 + 2N^2}}\right) - 1\right). \tag{27}$$

Figure 8 plots the number of correlated patterns retained by the modified FamE model computed from Equation 27 and from simulations. Again, the simulations match the predictions closely. It can also be seen that the modified version of FamE is not able to reproduce the power law. Equations 21 and 27 differ by just one denominator term, but for $r > 0$ this term is significant.

From the analytic formula for R (Equation 27), we can infer the relationship between R and P for the modified FamE model with a much larger number of neurons than we could simulate. Doing this shows that the FamE model predicts a line with slope 1 only for N larger than about 400 000. This value is only one order of magnitude lower than an estimated number of novelty neurons in the perirhinal cortex, i.e. $N = 4\,000\,000$ (there are ~40 000 000 neurons in the human perirhinal cortex (Insausti et al. 1998), but only ~10% of them are novelty neurons in monkeys (Brown and Xiang 1998), so we assume that there will be a similar proportion in humans). But note that we do not consider several biological properties of real familiarity discrimination networks, e.g. sparse connectivity between neurons, noise in neuronal processing and in synaptic plasticity, all of which decrease the performance of the networks (Bogacz and Brown 2003; Zhang 2007). Thus, it may be unlikely for a biologically realistic familiarity discrimination network of the size of human perirhinal cortex working according to the FamE model to achieve as high accuracy for large P as the participants of Standing's experiment.

Unlike the modified FamE model, the Anti-Hebbian and Info-max models were unaffected by the introduction of correlation to input patterns. The plots showing

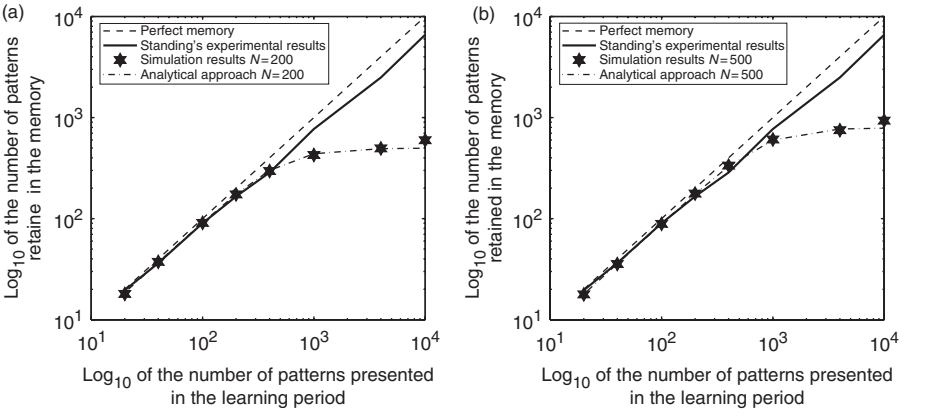

Figure 8. Number of items retained in the memory predicted by the FamE model with randomly initialized weights. The model is trained and tested on patterns produced by correlated inputs. The magnitude of the correlation used was $|r_{ij}| = 0.04$. The dashed-dotted line shows the prediction of Equation 27 and stars show results of simulations. For comparison, solid lines show Standing's result and dashed lines correspond to perfect recognition memory. (a) and (b) correspond to networks of 100 and 400 neurons, respectively. The learning rates for (a) and (b) were 2.56 and 2.30, respectively. For small P the performance of the model is very similar to that of the experimental participants. However, as P increases, the proportion of presented stimuli that are retained decreases.

the number of correlated patterns retained by these models look indistinguishable from Figures 4 and 5, and hence are not reproduced here. Instead, Figure 7(b) shows the error rates of the models. For the Anti-Hebbian model, the error rate for correlated patterns E_c does not vary with P to a significantly larger extent than the error rate for uncorrelated patterns E_u (the correlation between log P and $E_c - E_u$ was not significant). For the Info-max model, E_c seems to vary with P to a larger extent than E_u (significance of the correlation between log P and $E_c - E_u$: $p = 0.01$).

Discussion

In summary, we have shown that when uncorrelated patterns are used, three models of familiarity discrimination are able to reproduce the power law observed by Standing, as long as the synaptic weights are properly initialized. For larger numbers of neurons in the network, the models produce a relationship between R and P which forms a straight line in a graph with logarithmic axes. The intercept of this line is controlled by the learning rate, and the line moves up towards perfect recognition memory as the learning rate increases. The slope of this line is equal to 1 for the FamE model, while it is slightly closer to experimental data for Anti-Hebbian and Info-max models, but still all the models retain fewer stimuli than participants in Standing's experiment for low P. When correlated patterns are used, the FamE model is no longer able to reproduce the power law in simulations even when weights are properly initialized, while the performance of the other two models tested was affected very little by introducing correlation to input patterns.

In this section we discuss possible reasons why the slope in Standing's data is different from that predicted by the models of familiarity, why the participants in his experiment did not achieve perfect recognition memory, and relationships to other work on modelling familiarity discrimination.

Slope of the relationship

Figure 1 illustrates that the slope of the relationship between *R* and *P* produced by the models of familiarity is higher than in Standing's data. We discuss possible reasons for this difference.

First, according to dual-process models, recognition memory involves two processes: familiarity and recollection (Eichenbaum et al. 2007), and so far we have only considered the contribution of the familiarity process. Thus it is possible that familiarity indeed contributes the number of retained items shown by the dotted line in Figure 1, while the additional number of stimuli retained (corresponding to the area between the solid and dotted lines in Figure 1) is provided by the recollection process. This hypothesis is plausible as the recollection process is highly accurate but has limited storage capacity (Norman and O'Reilly 2003). Accordingly, the recollection process can retain almost all presented stimuli for low *P*, but not for high *P*. Thus the difference in slope as *P* increases would arise from the reducing contribution of an additional, recollective process used by real participants. This hypothesis predicts that for very large *P* the slope should change to that of the models. Indeed, the line between the last two points in Standing's data has the same slope as predicted by the models (Figure 1) but further experimental tests of this prediction with even larger *P* would be impractical.

The above hypothesis makes two testable predictions. First, patients with damage to the hippocampus, who have an impaired recollection process, should produce a relationship between *R* and *P* with slope close to 1 in Standing's paradigm. Or equivalently, their discrimination error rate *E* should be similar for low *P* and for high *P* (as models of familiarity have low memory interference due to their high capacity).

Second, for healthy participants the shapes of the ROC curves should be different for low *P* and high *P*. If the contribution of recollection becomes negligible for high *P*, the discrimination will be based solely on familiarity: for such decisions, participants produce a symmetric ROC curve (Yonelinas 2002). For low *P*, healthy participants make decisions based on both familiarity and recollection: this produces an asymmetric ROC curve.

Alternatively, one could assume that the difference in slope is caused by participants adaptively setting a higher learning rate for small *P* and a lower learning rate for high *P*. This hypothesis makes opposite predictions to the previous one. Namely, hippocampal patients should have different *Es* for low and for high *P*, and healthy participants should have ROC curves indicating similar contributions for recollection and familiarity for different *P*.

A further hypothesis presumes that the learning rate changes within the experiment, being higher for the first few study items, and then decreasing. The decrease in learning rate would result in a higher average learning rate for lower *P*, and hence higher accuracy at low *P*. This hypothesis makes similar predictions to

the immediately previous one, but in addition predicts a primacy effect, i.e. the first study items should have a higher probability of recognition at test.

Finally, it is also possible that the lower accuracy for higher P observed by Standing is a result of an aspect not simulated in any of our models, e.g. some kind of synaptic decay occurring during learning new stimuli.

Learning rate

We have shown that for large familiarity networks, as the learning rate increases, the number of items retained increases towards perfect memory. Thus one could ask why the participants of Standing's experiment did not have a higher learning rate that could have resulted in even better memory. We hypothesize three possible answers to this question.

First, it may not be optimal from an ecological point of view to retain all visual stimuli in recognition memory, as synaptic plasticity costs metabolic energy. This hypothesis is supported by the results of another condition in Standing's (1973) experiment, in which he presented up to 1000 'vivid' pictures (containing striking images). In this condition, he obtained even higher accuracies on test, which could suggest that participants adaptively set a higher learning rate for more relevant stimuli.

Second, it has been proposed that familiarity discrimination is carried out by the same network which also performs visual feature extraction (Li et al. 1993) (the perirhinal cortex is the last area in the ventral visual stream). This idea has been implemented in a number of models of familiarity discrimination (Sohal and Hasselmo 2000; Norman and O'Reilly 2003; Norman et al. 2005; Lulham et al. 2006). These models use the same learning rate for feature extraction and learning familiarity. But since feature extraction is a gradual process requiring learning over many trials, intuitively we can see that to extract features efficiently, the learning rate needs to be low.

Third, fatigue during the experiment may have caused the learning rate to drop off for human subjects for large P. If a high learning rate requires high attention but produces fatigue so that it cannot be sustained, this provides an explanation for having high learning rates for small sets and lower for larger.

Relationship to other work on modelling familiarity

We claim that there is little interference between memories stored in models of familiarity networks. This statement may seem to contradict the findings of Norman and O'Reilly (2003), who showed that familiarity networks are much more sensitive to interference with similar patterns than the hippocampal recollection network. However, Norman and O'Reilly tested interference for patterns that were very similar to one another, because they simulated the behavioural paradigm in which some of the test items are carefully chosen to overlap semantically with learning items (Roediger III and McDermott 1995). By contrast, Standing presented randomly chosen pictures; thus it is likely that their neural representations had a similar level of correlation as patterns used in our simulations, for which the level of

correlation was estimated from the experiment of Erickson et al. (2000) who also presented random images to monkeys.

As mentioned in the 'Introduction', several other models of familiarity discrimination have been proposed in addition to those tested in our study. We have also investigated how well two other models (Norman and O'Reilly 2003; Sohal and Hasselmo 2000) (simplified as described by Bogacz and Brown 2003) with $N = 500$ fit Standing's data, and we found they have a poor fit for correlated patterns. The simulations described in this article for two other proposed models (Meeter et al. 2005; Norman et al. 2005) would take an impractical amount of time due to the complexity of these models. Thus before simulating the performance of these models in Standing's experiment they need to be simplified, which is a current subject of our work.

Acknowledgements

Zacharias Androulidakis and Andrew Lulham contributed equally to this work. This work was supported by an MRC Capacity Building Fellowship held by Andrew W Lulham. We thank Tobias Larsen and Jiaxiang Zhang for discussion.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

Notes

[1] A computational model of a neural circuit involved in recognition memory has also been described by Hasselmo and Wyble (1997). But in this model the recognition is based on the activity in a hippocampal network, thus this model is more connected with the recollective component of recognition memory rather than the familiarity component, and hence we do not consider it further in this article.

[2] The weights of a single neuron given by Equation 4 may have both positive and negative values, which is biologically unrealistic. In the biologically plausible implementation of the FamE model (Bogacz et al. 2001) the weights defined in Equation 4 are increased by a constant (so all have positive values), and inhibitory neurons are introduced to balance this increase in neurons' excitability.

[3] The relaxation does not take place in the biologically plausible implementation of the FamE model (Bogacz et al. 2001) because the network has a feed-forward architecture (rather than recurrent).

References

Amari S, Cichocki A. 1998. Adaptive blind signal processing – neural network approaches. Proceedings of IEEE 86:2026–2048.

Amit DJ. 1989. Modeling brain function. Cambridge: Cambridge University Press.

Androulidakis Z. 2007. Explaining the power law describing the accuracy of familiarity judgements. MSc. thesis Bristol: University of Bristol.

Bell AJ, Sejnowski TJ. 1995. An information-maximization approach to blind separation and blind deconvolution. Neural Computation 7:1129–1159.

Bell AJ, Sejnowski TJ. 1997. The 'independent components' of natural scenes are edge filters. Vision Research 37:3327–3338.

Bogacz R, Brown MW. 2002. The restricted influence of the sparseness of coding on the capacity of the familiarity discrimination networks. Network: Computation in Neural Systems 13:457–485.

Bogacz R, Brown MW. 2003. Comparison of computational models of familiarity discrimination in the perirhinal cortex. Hippocampus 13:494–524.

Bogacz R, Brown MW, Giraud-Carrier C. 2001. Model of familiarity discrimination in the perirhinal cortex. Journal of Computational Neuroscience 10:5–23.

Bogacz R, Cohen JD. 2004. Parameterization of connectionist models. Behavioral Research Methods, Instruments, and Computers 36:732–741.

Brown MW, Aggleton JP. 2001. Recognition memory: What are the roles of the perirhinal cortex and hippocampus?. Nature Reviews Neuroscience 2:51–62.

Brown MW, Bashir ZI. 2002. Evidence concerning how neurons of the perirhinal cortex may effect familiarity discrimination. Philosphical Transactions of the Royal Society London. Series B, Biological Sciences 357:1083–1095.

Brown MW, Wilson FAW, Riches IP. 1987. Neuronal evidence that inferotemporal cortex is more important than hippocampus in certain processes underlying recognition memory. Brain Research 409:158–162.

Brown MW, Xiang JZ. 1998. Recognition memory: Neuronal substrates of the judgement of prior occurrence. Progress in Neurobiology 55:149–189.

Eichenbaum H, Otto T, Cohen NJ. 1994. Two functional components of the hippocampal memory system. Behavioral and Brain Sciences 17:449–518.

Eichenbaum H, Yonelinas AP, Ranganath C. 2007. The medial temporal lobe and recognition memory. Annual Review Neuroscience 30:123–152.

Erickson CA, Jagadeesh B, Desimone R. 2000. Clustering of perirhinal neurons with similar properties following visual experience in adult monkey. Nature Neuroscience 3:1143–1148.

Hasselmo ME, Wyble BP. 1997. Free recall and recognition in a network model of the hippocampus: Simulating effects of scopolamine on human memory function. Behavioural Brain Research 89:1–34.

Hertz J, Krogh A, Palmer RG. 1991. Introduction to the theory of neural computations. Redwood City, CA: Addison-Wesley.

Hopfield JJ. 1982. Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Science 79:2554–2558.

Insausti R, Juottonen K, Soininen H, Insausti AM, Partanen K, Vainio P, Laakso MP, Pitkanen A. 1998. MR volumetric analysis of the human entorhinal, perirhinal and temporopolar cortices. American Journal of Neuroradiology 19:659–671.

Kohonen T, Oja E, Ruohonen M. 1974. Adaptation of a linear system to finite set of patterns occurring in an arbitrarily varying order. Acta Polytechnica Scandinavica-Electrical Engineering Series 25:7–15.

Li L, Miller EK, Desimone R. 1993. The representation of stimulus familiarity in anterior inferior temporal cortex. Journal of Neurophysiology 69:1918–1929.

Lulham A, Vogt S, Bogacz R, Brown MW. 2006. Anti-Hebbian learning in the perirhinal cortex may underlie both familiarity discrimination and feature extraction. Computational Neuroscience Conference, Edinburgh.

Mandler G. 1980. Recognizing: The judgement of previous occurrence. Psychological Review 87:252–271.

Meeter M, Myers CE, Gluck MA. 2005. Integrating incremental learning and episodic memory models of the hippocampal region. Psychological Review 112:560–585.

Murray EA, Bussey TJ. 1999. Perceptual-mnemonic functions of the perirhinal cortex. Trends in Cognitive Sciences 3:142–151.

Nickerson RS. 1965. Short-term memory for complex meaningful visual configurations: A demonstration of capacity. Canadian Journal of Psychology 19:155–160.

Norman KA, Newman EL, Perotte AJ. 2005. Methods for reducing interference in the complementary learning systems model: Oscillating inhibition and autonomous memory rehearsal. Neural Networks 18:1212–1228.

Norman KA, O'Reilly RC. 2003. Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. Psychological Review 110:611–646.

Olshausen BA, Field DJ. 1996. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381:607–609.

Roediger III HL, McDermott KB. 1995. Creating false memories: Remembering words not present in list. Journal of Experimental Psychology: Learning, Memory and Cognition 21:803–814.

Shepard RN. 1967. Recognition memory for words sentences and pictures. Journal of Verbal Learning and Verbal Behavior 6:156–163.

Sobotka S, Ringo JL. 1994. Stimulus specific adaptation in excited but not in inhibited cells in inferotemporal cortex of Macaque. Brain Research 646:94–99.

Sohal VS, Hasselmo ME. 2000. A model for experience-dependent changes in the responses of inferotemporal neurons. Network: Computation in Neural Systems 11:169–190.

Standing L. 1973. Learning 10,000 Pictures. Quarterly Journal of Experimental Psychology 25:207–222.

Standing L, Conezio J, Haber RN. 1970. Perception of memory for pictures: Single trial learning of 2500 visual stimuli. Psychonomic Science 19:73–74.

Waydo S, Koch C. 2008. Unsupervised learning of individuals and categories from images. Neural Computation 20:1165–1178.

Yonelinas AP. 2002. The nature of recollection and familiarity: A review of 30 years of research. Journal of Memory and Language 46:441–517.

Zhang L. 2007. Investigating the capacity of models for familiarity discrimination with noise. MSc. thesis. Bristol: University of Bristol.