European Journal of Neuroscience, Vol. 15, pp. 355–364, 2002

The organization of visual object representations: a connectionist model of effects of lesions in perirhinal cortex

Timothy J. Bussey*,† and Lisa M. Saksida†
Section on the Neurobiology of Learning and Memory, Laboratory of Neuropsychology, National Institute of Mental Health, Bethesda, MD 20892, USA

Keywords: configural learning, discrimination learning, inferotemporal cortex, object recognition, rhesus macaque monkey, ventral visual stream

Abstract

We have developed a simple connectionist model based on the idea that perirhinal cortex has properties similar to other regions in the ventral visual stream, or 'what' pathway. The model is based on the assumption that representations in the ventral visual stream are organized hierarchically, such that representations of simple features of objects are stored in caudal regions of the ventral visual stream, and representations of the conjunctions of these features are stored in more rostral regions. We propose that a function of these feature conjunction representations is to help to resolve 'feature ambiguity', a property of visual discrimination problems that can emerge when features of an object predict a given outcome (e.g. reward) when part of one object, but predict a different outcome when part of another object. Several recently reported effects of lesions of perirhinal cortex in monkeys have provided key insights into the functions of this region. In the present study these effects were simulated by comparing the performance of connectionist networks before and after removal of a layer of units corresponding to perirhinal cortex. The results of these simulations suggest that effects of lesions in perirhinal cortex on visual discrimination may be due not to the impairment of a specific type of learning or memory, such as declarative or procedural, but to compromising the representations of visual stimuli. Furthermore, we propose that attempting to classify perirhinal cortex function as either 'perceptual' or 'mnemonic' may be misguided, as it seems unlikely that these broad constructs will map neatly onto anatomically defined regions of the brain.

Introduction

The perirhinal cortex, a polymodal region located in the ventromedial portion of the temporal lobe, is a site of convergence of brain systems thought to be important for normal memory and perception. It is unclear, however, whether the effects of lesions in perirhinal cortex can be best understood in terms of an impairment in a specific type of memory function, or in terms of a role for this region in visual information processing (Murray & Bussey, 1999). Resolving this issue has been difficult, in part due to the diverse and puzzling patterns of effects observed following lesions of perirhinal cortex. It is well established that perirhinal cortex has an important role in object recognition memory (e.g. Zola-Morgan *et al*., 1989; Meunier *et al*., 1993; Suzuki *et al*., 1993; Mumby & Pinel, 1994; Aggleton *et al*., 1997; Bussey *et al*., 1999a; Bussey *et al*., 2000b), thought to be the canonical example of 'declarative' memory. More recent research has shown, however, that the use of object recognition paradigms is not necessary to obtain deficits following perirhinal cortex lesions. Buckley & Gaffan (1997), for example, using a large set of object stimuli in a concurrent discrimination paradigm, reported that lesions of perirhinal cortex disrupted retention of preoperatively learned discriminations as well as the acquisition of new discriminations. These results suggest that a general function of perirhinal cortex might be the discrimination of visual stimuli. Visual discrimination learning can be preserved, however, following lesions in this region, provided the task requires the discrimination of only a small number of objects (e.g. Aggleton *et al*., 1997; Gaffan & Murray, 1992; Buckley & Gaffan, 1997; Thornton *et al*., 1997; Bussey *et al*., 1999a; Bussey *et al*., 1999b; Bussey *et al*., 2000a; Bussey *et al*., 2000b; Saksida *et al*., 2000; Bussey *et al*., 2001). At the same time Buckley & Gaffan (1998), using a relatively small set of object stimuli, found that perirhinal cortex lesions impaired a 'configural' discrimination learning task. This puzzling set of results requires explanation.

In the present study we investigated whether such effects of lesions in perirhinal cortex can be accounted for by assuming perirhinal cortex to have visual information processing properties similar to those in other regions of the ventral visual stream, or 'what' pathway (Ungerleider & Mishkin, 1982). We constructed a very simple connectionist model based on the assumption that visual representations in the ventral visual stream are organized hierarchically, such that representations of simple features of objects are stored in caudal regions of the ventral visual stream, and representations of the

Correspondence: Dr T. J. Bussey, at *present address below
E-mail: tjb1000@cam.ac.uk

**Present address*: Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, UK

†The authors contributed equally to this paper; the order of authorships is arbitrary.

Received 14 June 2001, revised 28 September 2001, accepted 21 November 2001

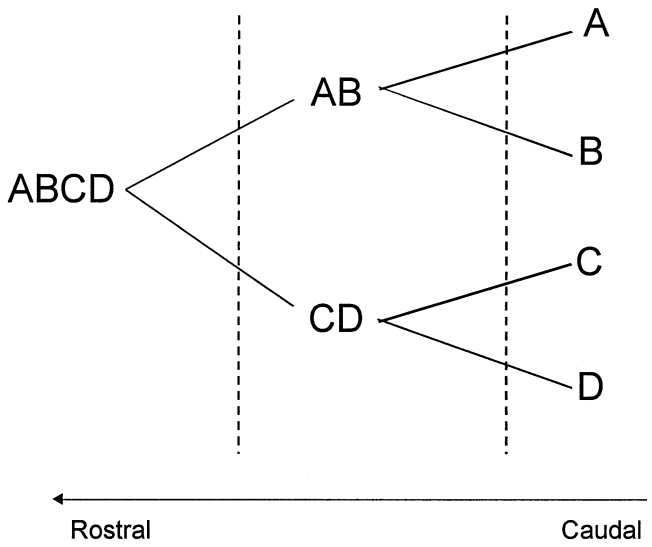

FIG. 1. The proposed organization of object feature representations in the ventral visual stream. A, B, C and D represent simple visual features encoded in caudal regions of the ventral visual stream. More complex representations of the conjunctions of these features are stored in more rostral regions. These conjunctive representations may reach maximum complexity in perirhinal cortex.

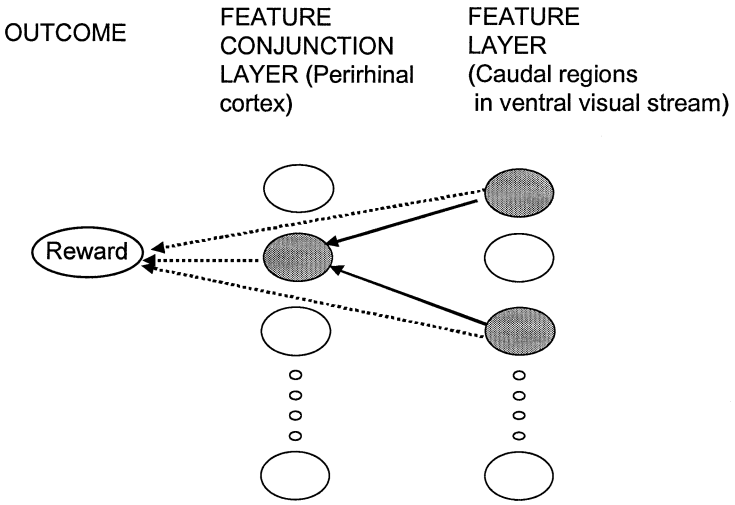

FIG. 2. Diagram of the connectionist model. The network consists of two layers of units, the feature layer and the feature conjunction layer, as well as an outcome node representing a consequent event (e.g. reward). Each node in the feature layer represents a single element of a stimulus. The feature layer is connected to the feature conjunction layer via a set of fixed weights. Active units are shown in grey. Note that, for simplicity, only two feature layer units are shown to be active, whereas in the simulations 10 units per stimulus are active. Both the feature layer and the feature conjunction layer are fully connected to the reward node. These weights are adjustable via an associative mechanism. The feature conjunction layer represents perirhinal cortex (PRh) and the feature layer represents more caudal regions of the ventral visual stream. See the Appendix for details.

conjunctions of these features are stored in more rostral regions. According to this model, a function of these rostral feature conjunction representations is to resolve 'feature ambiguity', a property of visual discrimination problems that can emerge when features of an object predict a given outcome (e.g. reward) when part of one object, but predict a different outcome when part of another object.

To test this model, lesions were made in the component of the network corresponding to perirhinal cortex, and the resulting effects compared with known effects of lesions in perirhinal cortex in monkeys. In the present study we focused on recently reported effects of lesions in perirhinal cortex in monkeys on visual discrimination learning and performance. Specifically, we tested three findings that have been reported in the neuropsychological literature: (i) that perirhinal cortex lesions disrupt acquisition of concurrent visual discriminations when the size of the stimulus set is large, but not when it is small; (ii) that perirhinal cortex lesions disrupt configural discrimination learning, even when the stimulus set size is small; and (iii) that perirhinal cortex lesions yield greater effects on retention of preoperatively learned discrimination problems than on acquisition of new problems of the same type. This was followed by studies in which novel predictions of the model were tested in lesion experiments in rhesus monkeys (Bussey *et al*., 1999b; Saksida *et al*., 2000; and the companion paper, Bussey *et al*., 2002).

General methods

In the present model we assume that the cortical regions of the ventral visual stream, including perirhinal cortex, store visual stimulus representations which can be subsequently associated with events such as rewards and responses. An intact perirhinal cortex will therefore be necessary for the performance of many visual tasks, including certain visual discrimination learning tasks. However, whilst lesions within perirhinal cortex often produce deficits in such tasks, equally often they do not. We propose that this can be explained by considering the putative hierarchical organization of representations in the ventral visual stream, illustrated schematically in Fig. 1. According to this model, representations of simple features of objects are stored in caudal regions of the ventral visual stream, and representations of the conjunctions of these features are stored in more rostral regions. This model is an extension of the well-known hierarchical model proposed by Hubel & Wiesel (1962, 1965), and is supported by more recent electrophysiological evidence suggesting that neurons in rostral regions of the ventral visual stream code more complex visual representations than do neurons in more caudal regions (e.g. Desimone & Ungerleider, 1989), and that complex objects may be represented by the combination of 'feature columns' (Tsunoda *et al*., 2001). Several models of object processing feature a hierarchical organization of visual representations (e.g. Fukushima, 1980; Perrett & Oram, 1993; Wallis & Rolls, 1997; Riesenhuber & Poggio, 1999). Although these models capture impressively the data from electrophysiological studies of inferotemporal cortex (IT), no such model has attempted to account for the effects of lesions in this region. In the present model perirhinal cortex is thought to contain representations of complex conjunctions of visual features, perhaps corresponding to a 'gestalt' representation of an object whole. Thus whilst lesions in perirhinal cortex will abolish these complex representations of objects, representations of the features that comprise those objects will be preserved in more caudal regions of the ventral visual stream. It has been suggested previously that parahippocampal regions may contain conjunctive representations of stimuli (e.g. Eichenbaum & Bunsey, 1995; Gluck & Myers, 1995; O'Reilly & Rudy, 2001). It is important to note that we do not view perirhinal cortex as the only region that contains conjunctive representations (see Fig. 1), but as a region that contains perhaps the most complex conjunctive representations in the ventral visual stream. To instantiate this model in connectionist form, we reduced

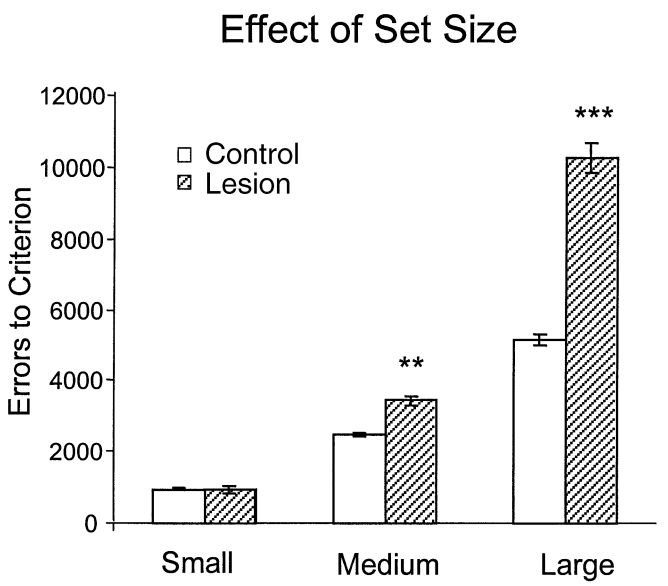

FIG. 3. Effect of set size on acquisition of pairwise concurrent discrimination. Error bars represent ± SEM. $^{**}P < 0.01$, $^{***}P < 0.001$ vs. Control.

the model shown in Fig. 1 to a two-layer model, in which the 'feature' layer (corresponding to regions caudal to perirhinal cortex) contains representations of simple features of objects, and the 'feature conjunction' layer (corresponding to perirhinal cortex) contains representations of conjunctions of these visual features (see Fig. 2).

The model we have constructed is formulated at a high level. Neurobiologically, the model is formulated at the anatomical systems level; psychologically we are interested in how representations are organized and associated in the modification of behaviour, much in keeping with the 'parallel distributed processing' school of computational modelling (e.g. McClelland & Rumelhart, 1986; Rumelhart & McClelland, 1986), and the 'associative' tradition of learning theory (e.g. Mackintosh, 1974; Dickinson, 1980). The units in the network thus represent visual representations and, although the model is influenced by electrophysiological studies of neurons, the units do not represent individual neurons and should not be interpreted as an endorsement of the idea of 'grandmother cells'. The neural entity that most probably corresponds to a unit in the network is a population of interconnected neurons that code for a given visual representation. No attempt is made to explain how exactly these representations are formed or coded at the neuronal level. Furthermore, the data for which we are attempting to account is constrained to visual discrimination learning and memory. Properties of neurons that are not necessary to achieve this aim (for example, back-projections or 'repetition suppression' phenomena) have not been included in the model, in keeping with the principle that a model should be no more complex than is absolutely necessary to account for the target data.

Finally, although we view perirhinal cortex as having properties similar to other regions in the ventral visual stream, it should be noted that perirhinal cortex also has connections not shared by these other regions. Most significantly, whereas other regions of the ventral visual stream are purely visual, perirhinal cortex has access to nonvisual sensory information, including auditory and somatosensory information (for review see box 3 of Murray & Bussey, 1999). It is conceivable these connections allow the binding of visual and nonvisual features in the service of object identification (Murray & Bussey, 1999). In the present study, however, we focus on the role of perirhinal cortex specifically in visual learning and memory. Accordingly, our model is based on the properties of perirhinal cortex thought to be shared with other visual areas in the ventral visual stream.

In each of the simulations in this article, we lesioned the component of the network corresponding to perirhinal cortex by removing it completely, and then compared the effects of this lesion on discrimination learning and memory in the model with the effects of lesions of perirhinal cortex in monkeys. The computational details of the connectionist network are provided in the Appendix.

Experiment 1: stimulus-set-size effects in concurrent discrimination learning

Much evidence points to a role for perirhinal cortex in the acquisition of visual discriminations. Lesions of perirhinal cortex have been reported to disrupt acquisition of visual pairwise discriminations when a large number of stimulus pairs must be discriminated concurrently (Buckley & Gaffan, 1997). Perirhinal cortex is not, however, always necessary for discrimination learning. Several studies using small stimulus-set sizes, in monkeys as well as in rats, have found preserved discrimination learning following lesions that include perirhinal cortex (e.g. Gaffan & Murray, 1992; Aggleton *et al*., 1997; Buckley & Gaffan, 1997; Thornton *et al*., 1997; Bussey *et al*., 1999a; Bussey *et al*., 1999b; Bussey *et al*., 2000a; Bussey *et al*., 2000b; Saksida *et al*., 2000; Bussey *et al*., 2001). Thus it appears that perirhinal cortex lesions have a greater effect on the acquisition of concurrent visual discriminations when the stimulus-set size is large, compared to when the set size is relatively small. In this experiment we investigated the effects of stimulus-set size in a manner similar to that of Buckley & Gaffan (1997).

Simulation

Three sets of stimuli were created by randomly selecting and activating 10 out of a possible 100 units for each stimulus. Set 'Small' consisted of two pairs of stimuli, set 'Medium' consisted of eight pairs of stimuli, and set 'Large' consisted of 16 pairs of stimuli. Two groups of four networks each were then initialized: Group 'Control' consisted of intact networks whereas group 'Lesion' consisted of networks that had been lesioned to simulate the effect of a lesion in perirhinal cortex. Each network was trained on a pairwise concurrent discrimination task using set 'Small' until it reached criterion. (Criterion was set at 80% correct on two consecutive blocks. A block is defined here and elsewhere as 10 presentations of each stimulus pair in the set.) The networks were then reinitialized. Each network was trained to criterion on the eight-pair concurrent discrimination (set 'Medium'). The networks were again reinitialized and trained to criterion on the 16-pair concurrent discrimination (set 'Large').

Results

As shown in Fig. 3, group Lesion was impaired relative to group Control on concurrent discrimination learning when the stimulus-set size was medium or large, but not when it was small. Two-way analysis of variance with Group and Set Size as factors revealed a significant main effect of Group, $F_{1,6} = 132.6$, $P < 0.0001$, a significant main effect of Set Size, $F_{2,12} = 1227.6$, $P < 0.0001$, and a significant Group × Set Size interaction, $F_{2,12} = 189.3$, $P < 0.0001$. Analysis of simple main effects revealed that group Lesion committed significantly more errors in reaching criterion than

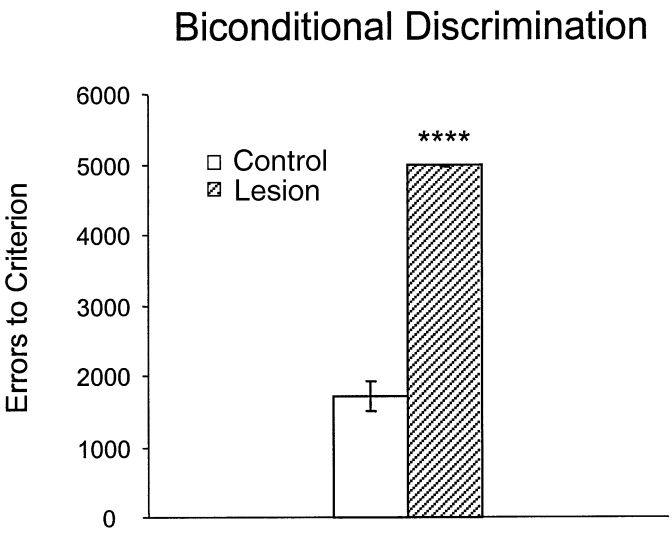

FIG. 4. Acquisition of a biconditional discrimination. Error bars represent ± SEM. ****$P < 0.0001$ vs. Control.

did group Control in the large-set-size condition, $P < 0.001$, and in the medium-set-size condition, $P < 0.01$, but not in the small-set-size condition, $P = 0.98$. Both groups committed more errors as the set size increased, as indicated by a significant effect of Set Size for both group Control, $P < 0.0001$, and group Lesion, $P < 0.0001$.

Discussion

This experiment shows that removing the feature conjunction layer of the network can reproduce stimulus-set-size effects similar to those observed following lesions of perirhinal cortex in monkeys (Buckley & Gaffan, 1997). Specifically, the larger the set size, the greater the effect of the lesion. This occurs because whereas the intact networks can represent individual features of a stimulus as well as their conjunction (i.e. they can represent the features A and B as well as the conjunction of these features, AB), the lesioned networks can represent only the individual stimulus features (A and B). The present model of perirhinal cortex is similar to the computational model of IT of Gaffan *et al.* (1986), in which lesions in IT are thought to reduce the number of 'stimulus attributes' available during visual discrimination learning. In the present model, however, after perirhinal cortex lesions attributes are not lost; rather, it is the ability to represent conjunctions of attributes of complex objects that is lost. As a result, due to their richer stimulus representation the intact networks can build several associative links (A + ; B + ; AB +) whereas the lesioned networks can only build associative links from the individual feature representations to the outcome (A + ; B +). As the stimulus-set size increases, the probability increases that a particular feature will be part of both a rewarded and a nonrewarded stimulus; this property, possessed by different discrimination problems to varying degrees, we refer to as 'feature ambiguity'. The probability of a specific conjunction of features being both rewarded and nonrewarded, however, is much smaller. Thus an intact network, and an animal with an intact perirhinal cortex, will have an advantage over their lesioned counterparts because they can represent the conjunction of complex stimulus features. In contrast the lesioned network, or animal, must rely on the spared representations of individual stimulus features to attempt to solve the discrimination. This advantage of the intact networks or animals increases as the degree of feature ambiguity increases. Because the probability of feature ambiguity increases with set size, so does the advantage of the intact group.

Experiment 2: configural learning

Feature ambiguity reaches a maximum in 'configural' tasks in which items cannot be discriminated according to simple features, but rather can only be discriminated through the use of representations of the conjunction of features. An example of such a task is the biconditional discrimination task. The biconditional discrimination task can be formalized as AB+, BC–, CD+, AD– (where features are represented by the letters A, B, C and D, objects by the conjunction of two of these features (e.g. AB), reward by '+', and nonreward by '–'). It can be seen that this task cannot be solved by assigning reward or nonreward to any of the features A, B, C or D, as each of these features is associated equally with both reward and nonreward. Only by associating the conjunctions of features (e.g. AB) with reward can the problem be solved. For this reason it is predicted that removing the feature conjunction layer of the network could have a devastating effect on the acquisition of such configural tasks. Furthermore, because of the high degree of feature ambiguity in such tasks, deficits should be obtainable using a smaller set size than is necessary to produce deficits on a concurrent learning task. Precisely this finding has recently been reported by Buckley & Gaffan (1998) using an extended version of the biconditional discrimination task. The present experiment tests these predictions using the standard version of the task, AB+, BC–, CD+, AD–.

Simulation

Four stimuli were created: A (consisting of units 1–10, of a possible 100), B (units 11–20), C (units 21–30), and D (units 31–40). Pairs of these stimuli were then presented together in a biconditional discrimination as follows:

Pair 1: AB → +, BC → –

Pair 2: CD → +, AD → –

Two groups of four networks each were then initialized: Group 'Control' consisted of intact networks whereas group 'Lesion' consisted of networks that had been lesioned to simulate the effect of a perirhinal cortex lesion. Each network was trained on the biconditional pairwise concurrent discrimination task until it either reached criterion (80% correct on two consecutive blocks) or exceeded 500 blocks, at which point it was deemed to have 'failed' the discrimination.

Results

Removing the feature conjunction layer of the network had a devastating effect on the acquisition of the biconditional discrimination (see Fig. 4). Analysis of variance revealed a highly significant effect of Group, $F_{1,6} = 322.9$, $P < 0.0001$. All networks in group Lesion failed to attain criterion levels of performance within 500 blocks.

Discussion

Lesioned networks were severely impaired on the biconditional discrimination. Although the biconditional discrimination set size was smaller than that used for the large set concurrent discrimination in Experiment 1, the lesioned networks showed a relatively greater impairment in acquisition of the biconditional discrimination task. This is because the biconditional discrimination task shifts the probability of feature ambiguity to one: each feature is both rewarded and nonrewarded, thereby eliminating the need for a large set size to

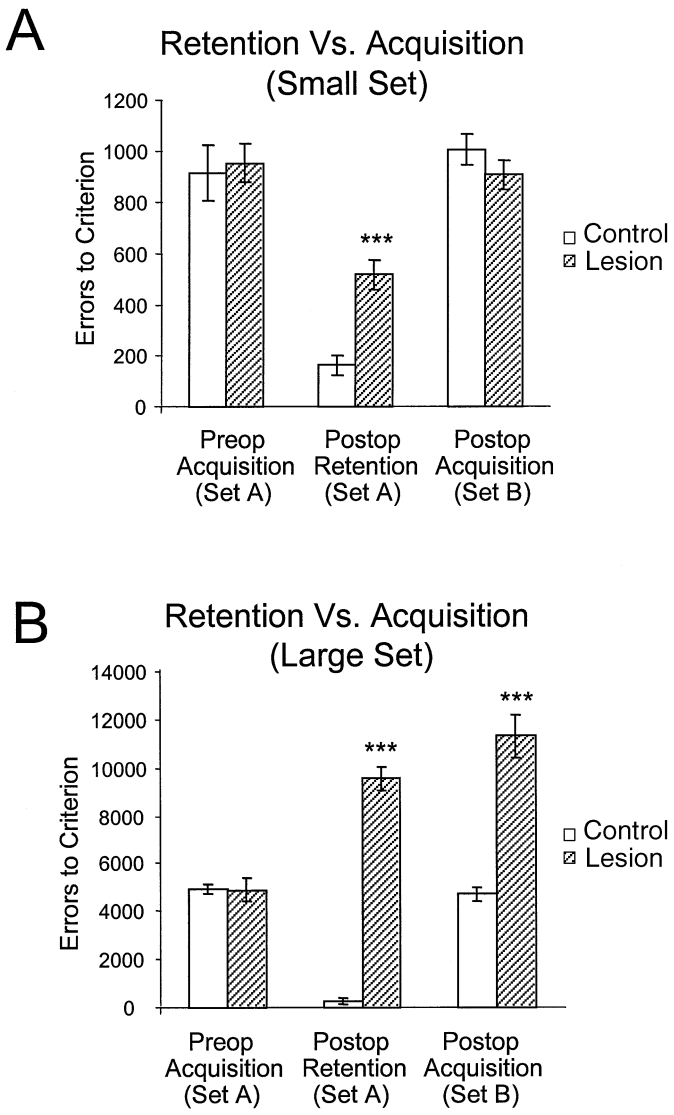

FIG. 5. (A) Effect of removing the feature conjunction layer on retention and acquisition of a small-set-size pairwise concurrent discrimination. (B) Effect of removing the feature conjunction layer on retention and acquisition of a large-set-size pairwise concurrent discrimination. Error bars represent ± SEM. ***$P < 0.001$ vs. Control.

bring out a deficit in the lesioned networks. Concurrent discriminations are, however, very difficult for the lesioned networks and for monkeys with perirhinal cortex lesions when the feature ambiguity is high (i.e. when the stimulus-set size is large; Experiment 1 of the present study; Buckley & Gaffan, 1997). In this sense a large-set-size concurrent discrimination can be thought of as a 'partial' configural task.

In the present model, perirhinal cortex is represented by the feature conjunction layer of the network, which contains configural representations of stimuli constructed from features in the feature layer. This allows the network to simulate lesion effects on the biconditional discrimination task, and on concurrent discriminations with a large set size (Experiment 1). However, it is important to emphasize that we do not view perirhinal cortex as specialized for configural learning as opposed to 'elemental' learning, in a manner akin to that proposed for the hippocampus (Sutherland & Rudy, 1989), nor do we view perirhinal cortex as the only region containing configural representations. In our view the representations in a particular region of the ventral visual stream are configural only relative to more caudal regions, which could be thought of as containing 'relatively' elemental representations (see Fig. 1). Such elemental layers, however, would serve as feature conjunction layers relative to regions caudal to them, and so no region of the ventral visual stream can be considered to be exclusively configural or elemental. The visual configural representations thought to reside in perirhinal cortex are different from those in more caudal regions of the ventral visual stream only by virtue of their complexity, explaining why configural learning with very simple stimuli may not be impaired following perirhinal cortex lesions (Bussey *et al*., 2000a). The model is thus consistent with the previous suggestion that 'elemental' and 'configural' learning circuits are anatomically coextensive (Bussey *et al*., 1998; Bussey *et al*., 2000a).

Experiment 3: greater effects of lesions in perirhinal cortex on retention vs. acquisition of visual discriminations

A consistent yet puzzling effect of perirhinal cortex lesions is that retention of preoperatively learned discriminations can be disrupted while the acquisition of new discriminations can be spared (e.g. Gaffan & Murray, 1992; Thornton *et al*., 1997). Only under certain challenging conditions is acquisition of concurrent discriminations impaired (Buckley & Gaffan, 1997; Buckley & Gaffan, 1998). Somehow, reacquisition and new learning are differentially affected by perirhinal cortex lesions (Buckley & Gaffan, 1997). It is tempting to explain this effect by appealing to separate learning systems in the brain, for example object-knowledge (perirhinal cortex) vs. procedural (nonperirhinal cortex) systems (e.g. Thornton *et al*., 1997). We predicted, however, that the simple properties intrinsic to the present model could provide a parsimonious explanation of this result.

In this experiment we tested the prediction that lesions of the feature conjunction layer of the network should yield greater effects on retention of preoperatively learned concurrent discrimination problems than on postoperative acquisition of new discriminations.

Simulation

Four sets of stimuli were created by randomly selecting and activating 10 out of a possible 100 units for each stimulus. Two sets consisted of two pairs of stimuli and the other two sets consisted of 16 pairs of stimuli. Two groups of four networks each were then initialized. For the first phase of this experiment, both group 'Control' and group 'Lesion' consisted of intact networks. Each network was trained on a pairwise discrimination task using the first small stimulus set until it reached criterion (80% correct on two consecutive blocks). For the second phase of the experiment, the feature conjunction layer was removed from the networks in the Lesion group, and then both groups were trained again to criterion on the first stimulus set. Finally, for the third phase of the experiment both groups were run on a new pairwise discrimination using the stimuli from the second small set. The above procedure was then repeated using the two large stimulus sets.

Results

As shown in Fig. 5A, group Lesion was impaired relative to group Control on the retention, but not the acquisition, of a concurrent discrimination when the stimulus-set size was small. Two-way analysis of variance with Group and Test as factors revealed a

significant overall main effect of Group, $F_{1,6} = 9.8$, $P = 0.02$, a significant main effect of Test, $F_{2,12} = 50.97$, $P < 0.0001$, and a significant Group × Test interaction, $F_{2,12} = 5.74$, $P = 0.018$. Analysis of simple main effects revealed that group Lesion committed significantly more errors than did group Control in reattaining criterion on the discrimination learned prior to lesioning, $P = 0.001$, but the groups did not differ in the acquisition of a discrimination with a new stimulus set, $P = 0.27$. There was no difference between the groups in the acquisition of the discrimination learned prior to lesioning, $P = 0.69$. Figure 5A can be compared with data from monkeys shown in fig. 5 of Gaffan & Murray (1992).

As shown in Fig. 5B, group Lesion was impaired relative to group Control on the retention *and* acquisition of a concurrent discrimination when the stimulus-set size was large. Two-way analysis of variance with Group and Test as factors revealed a significant main effect of Group, $F_{1,6} = 181.5$, $P < 0.0001$, a significant main effect of Test, $F_{2,12} = 43.4$, $P < 0.0001$, and a significant Group × Test interaction, $F_{2,12} = 78.15$, $P < 0.0001$. Analysis of simple main effects revealed that group Lesion committed significantly more errors than did group Control in reattaining criterion on the discrimination learned prior to lesioning, $P < 0.001$, and in the acquisition of a discrimination with a new stimulus set, $P < 0.001$. There was no difference between the groups in the acquisition of the discrimination learned prior to lesioning, $P = 0.90$.

Discussion

The present experiment reproduced the finding that perirhinal cortex lesions in monkeys lead more readily to impairments in the retention of preoperatively learned discriminations, compared with the postoperative acquisition of novel discrimination problems. Using a relatively small stimulus-set size, the lesioned networks were impaired in the retention of a preoperatively learned discrimination, but were unimpaired in learning new discriminations. When a large set size was used, however, the lesioned networks were impaired in both retention of a preoperatively learned discrimination and in learning new discriminations. Thus lesions in the network were able to produce deficits both in retention and in acquisition, but the former occurred more readily, requiring only a small set size, whereas larger set sizes were needed to produce deficits in acquisition. This pattern of results parallels those seen following perirhinal cortex lesions in monkeys.

The network produces the above pattern of results as follows. During acquisition of a discrimination by an intact network, two types of representation of the stimulus will have been formed: a representation of each stimulus feature in the feature layer and a representation of the conjunction of these features in the feature conjunction layer. As a result, associative links will be constructed between the outcome node and each of these types of stimulus representation. During training, the associative learning rule (see Equations 9 and 10 in the Appendix) operates such that the sum of the weights of the associative links, modulated by the degree of activation of the corresponding stimulus representation units in the feature layer and the feature conjunction layer, asymptotes at a value of one. This value is output as the response of the network. After the network has been trained and then lesioned, however, the stimulus conjunction representation in the feature conjunction layer will no longer exist, and the associative links from the feature conjunction layer will no longer contribute to the response of the network. As a result, the performance of the network on the previously learned discriminations will be impaired. Lesioned networks can still, however, acquire a small-set-size discrimination at the same rate as control networks, as described in Experiment 1.

This account can be recast in terms of the effects of lesions in monkeys. A control monkey can solve a simple small-set-size discrimination using either representations of the features of the stimulus or the conjunctions of those features. Normally the monkey will use both: the learning is distributed across these two types of representation. Thus when perirhinal cortex is lesioned following acquisition of the discrimination, a major contributor to the associative connections leading to the response is removed. The result is an impairment in the performance of the discrimination. If, however, the lesioned animal is then presented with a new small-set-size discrimination, learning can be accomplished using the feature representations alone, and there will be no difference between lesioned and intact animals in their ability to solve the discrimination. The lesioned animal may still, however, be impaired in the acquisition of a large-set-size concurrent discrimination, as described in Experiment 1.

General discussion

In the present study we tested the hypothesis that effects of lesions in perirhinal cortex on visual discrimination learning can be accounted for by assuming that perirhinal cortex has visual information processing properties similar to those of other regions of the ventral visual stream. We constructed a connectionist model that is based on the simple assumption that visual representations in the ventral visual stream are organized hierarchically, such that representations of simple features of objects are stored in caudal regions, and representations of the conjunctions of these features are stored in more rostral regions. The feature conjunction layer of the connectionist network, which corresponds to perirhinal cortex, was lesioned, and lesioned and intact networks tested on discrimination paradigms that have revealed perirhinal cortex lesion-induced deficits in monkeys. It was found that the foregoing simple assumptions are able to account for effects of perirhinal cortex lesions on visual discrimination tasks, including set-size effects, impairments in configural learning, and greater effects on retention vs. acquisition of visual discriminations.

Why redundant stimulus representations?

If objects are already fully specified by the representations of their features in caudal regions of the ventral visual stream, what then is the advantage conferred by the complex conjunctive representations in rostral regions such as perirhinal cortex? The results of the present study suggest that at least one role for complex conjunctive representations is to disambiguate stimuli when the behavioural significance of individual object features is ambiguous (a situation we have referred to as 'feature ambiguity'). These conjunctive representations provide the additional information that the features present in the stimulus 'belong together': the whole is more than merely the sum of the parts. In this way, these complex conjunctive representations serve to reduce interitem interference. When attempting to solve complex visual discriminations, an animal with a dysfunctional perirhinal cortex will rely more on simple features to attempt to distinguish the stimuli. For such animals, discriminations between stimuli will be difficult in certain cases in which visual features are rewarded when part of one object, but not when part of another.

In companion studies, this proposed critical role for feature ambiguity has been tested explicitly. Novel predictions made by the model were made concrete by simulations using the present connectionist network, and subsequent experiments in rhesus monkeys carried out to test these predictions. For example, in the present

study it was argued that to reveal an impairment in monkeys with perirhinal cortex lesions, a larger set size is required when using a concurrent as opposed to a configural discrimination paradigm, because of the high feature ambiguity in configural tasks relative to the probabilistic 'partial' configural properties of concurrent discriminations. A prediction that follows is that if set size were held constant, and feature ambiguity were varied systematically, perirhinal cortex lesions would lead to impairments under conditions of high feature ambiguity, irrespective of set size. This prediction was confirmed in the companion paper (Bussey *et al*., 2002). In addition, we have found that perirhinal cortex lesions impair the acquisition of single-pair visual discriminations having a high, but not a low, degree of feature ambiguity (Saksida *et al*., 2000). Furthermore, when discriminations are previously acquired, increasing the degree of feature ambiguity in the discriminanda can lead to perirhinal cortex lesion-induced impairments (Bussey *et al*., 1999b). These findings provide strong support for the assumptions of the present model.

According to the present model, perirhinal cortex contains conjunctive representations. In this narrow sense, the model is similar to the 'configural association theory' of hippocampal function of Sutherland & Rudy (1989). There are, however, critical differences. For example, configural association theory views spatial learning as a special case of configural learning. We do not; indeed, perirhinal cortex lesions, which can impair configural learning, can spare or even facilitate spatial learning (e.g. Bussey *et al*., 1999a, 2000a; b; 2001). Second, configural association theory posits distinct configural and elemental (nonconfigural) learning systems in the brain. According to the present view, however, the cortical fields of the ventral visual stream are neither exclusively configural nor exclusively elemental; the representations in a particular region of the ventral visual stream are configural only relative to more caudal regions, which could be thought of as containing elemental representations from which the more rostral conjunctive representations are formed (see Fig. 1). These 'elemental' representations, however, are themselves 'configural', being formed from 'elements' stored in still more caudal regions. This view is not compatible with the idea of anatomically separate stores for configural and elemental representations, and instead is more consistent with the view that 'configural' and 'elemental' learning circuits are anatomically coextensive (Bussey *et al*., 2000a; Bussey *et al*., 1998).

Perceptual and mnemonic accounts of perirhinal cortex function

That lesion effects in perirhinal cortex can be accurately simulated by assuming that perirhinal cortex has visual information processing properties similar to those of other regions of the ventral visual stream leads us to adopt quite a different view of perirhinal cortex function to that proposed by other authors. According to the prevailing 'declarative' view, perirhinal cortex, the hippocampus and other medial temporal lobe structures operate together in the service of declarative memory, playing little or no role in other functions such as visual analysis or perception (e.g. Squire, 1992; Sakai & Miyashita, 1993; Zola-Morgan *et al*., 1994; Buffalo *et al*., 1998a; Buffalo *et al*., 1998b; Buffalo *et al*., 1999; Buffalo *et al*., 2000). Evidence taken to support this view has included reports that perirhinal cortex lesions can have similar effects to lesions of the hippocampus. It is also true, however, that the effects of perirhinal cortex lesions can be dissociated from those of the hippocampus (e.g. Aggleton *et al*., 1997; Rothblat & Kromer, 1991; Gaffan, 1994; Glenn & Mumby, 1998; Murray & Mishkin, 1998; Bussey *et al*., 1999a; Bussey *et al*., 2000b; Murray, 2000; Baxter & Murray, 2001). Furthermore, perirhinal cortex lesions can lead to similar effects to lesions in other regions of the ventral visual stream on visual discrimination tasks, including concurrent discrimination learning (Cowey & Gross, 1970; Iwai & Mishkin, 1968; Phillips *et al*., 1988). Thus perirhinal cortex may not be best conceptualized as possessing functions similar to the putative mnemonic functions of the hippocampus but distinct from the putative perceptual functions of the ventral visual stream. According to the present model, the effects of lesions in perirhinal cortex on visual discriminations are due to the lesion compromising the representations of visual stimuli. When the lesion is made prior to acquisition, the substrate upon which the representations will be laid down is damaged, thus leading to impaired encoding of the representation. When the lesion is made subsequent to learning, the conjunctive representations stored there are lost. These disruptions of the memory for visual stimulus features lead to visual discrimination deficits that can be construed as perceptual. Thus according to the present model, it does not seem particularly useful to attempt to classify perirhinal cortex as either exclusively perceptual or exclusively mnemonic. Indeed, it seems to us highly unlikely that concepts such as perception and memory will map neatly onto anatomically distinct regions of the brain. The foregoing proposals thus comprise what we refer to as a 'perceptual–mnemonic/feature conjunction' (PMFC in the companion paper, Bussey *et al*., 2002) model of perirhinal cortex function (see also Murray & Bussey, 1999).

Conclusion

To summarize, we have constructed a model of perirhinal cortex function that is built on the assumption that perirhinal cortex is a rostral component of the ventral visual stream, which contains complex conjunctive representations of visual stimulus features. This model can reproduce the effects of perirhinal cortex lesions on visual discrimination tasks. The effects of lesions in perirhinal cortex, according to the present model, are due not to the impairment of a particular type of learning or memory (for example, stimulus–reward or stimulus–response, declarative or procedural) but to compromising the representations of visual stimuli. To the extent to which a lesion in the network mimics the effects of perirhinal cortex lesions in animals, there is no need to invoke special functions for perirhinal cortex, distinct from those of other regions of the ventral visual stream. In domains other than visual discrimination, however, it is likely that the unique connectivity of perirhinal cortex will endow it with properties distinct from those of other regions of the ventral visual stream. For example, perirhinal cortex may be ideally positioned to interact with the hippocampus in the service of object/place learning or episodic memory (Gaffan & Parker, 1996; Murray, 2000; Bussey *et al*., 2001; Bussey & Aggleton, in press). Indeed, in general it appears that perirhinal cortex receives many more polymodal inputs than other areas of the ventral visual stream (Suzuki & Amaral, 1994), and such polymodal information may allow the binding of visual and nonvisual features of objects (Parker & Gaffan, 1998; Murray & Bussey, 1999; Murray, 2000). How our model fits into a more comprehensive theory of perirhinal cortex function encompassing, for example, object recognition and polymodal learning and memory, is a target for future research.

Acknowledgements

The authors would like to thank John Aggleton, David Gaffan, Jay McClelland, Mortimer Mishkin and Betsy Murray for very helpful comments on an earlier version of this manuscript, and Betsy Murray for support and encouragement in all phases of this study.

Abbreviations

CR, conditioned response; IT, inferotemporal cortex.

References

Aggleton, J.P., Keen, S., Warburton, E.C. & Bussey, T.J. (1997) Extensive cytotoxic lesions involving both the rhinal cortices and area TE impair recognition but spare spatial alternation in the rat. *Brain Res. Bull.*, **43**, 279–287.

Baxter, M. & Murray, E.A. (2001) Opposite relationship of hippocampal and rhinal cortex damage to delayed nonmatching-to-sample deficits in monkeys. *Hippocampus*, **11**, 61–71.

Buckley, M.J. & Gaffan, D. (1997) Impairment of visual object-discrimination learning after perirhinal cortex ablation. *Behav. Neurosci.*, **111**, 467–475.

Buckley, M.J. & Gaffan, D. (1998) Perirhinal cortex ablation impairs configural learning and paired-associate learning equally. *Neuropsychologia*, **36**, 535–546.

Buffalo, E.A., Ramus, S.J., Clark, R.E., Teng, E., Squire, L.R. & Zola, S.M. (1999) Dissociation between the effects of damage to perirhinal cortex and area TE. *Learn. Mem.*, **6**, 572–599.

Buffalo, E.A., Ramus, S.J., Squire, L.R. & Zola, S.M. (2000) Perception and recognition memory in monkeys following lesions of area TE and perirhinal cortex. *Learn. Mem.*, **7**, 375–382.

Buffalo, E.A., Reber, P.J. & Squire, L.R. (1998a) The human perirhinal cortex and recognition memory. *Hippocampus*, **8**, 330–339.

Buffalo, E.A., Stefanacci, L., Squire, L.R. & Zola, S.M. (1998b) A reexamination of the concurrent discrimination learning task: The importance of anterior inferotemporal cortex, area TE. *Behav. Neurosci.*, **112**, 3–14.

Bussey, T.J. & Aggleton, J.P. (in press) The 'what' and 'where' of event memory: Independence and interactivity within the medial temporal lobe. In Parker, A., Wilding, E. & Bussey, T.J. (eds.), *The Cognitive Neuroscience of Memory: Encoding and Retrieval*. Psychology Press, London.

Bussey, T.J., Dias, R., Amin, E., Muir, J.L. & Aggleton, J.P. (2001) Perirhinal cortex and place–object conditional learning in the rat. *Behav. Neurosci.*, **115**, 776–785.

Bussey, T.J., Dias, R., Redhead, E.S., Pearce, J.M., Muir, J.L. & Aggleton, J.P. (2000a) Intact negative patterning in rats with fornix or combined perirhinal and postrhinal cortex lesions. *Exp. Brain Res.*, **134**, 506–519.

Bussey, T.J., Duck, J., Muir, J.L. & Aggleton, J.P. (2000b) Distinct patterns of behavioural impairments resulting from fornix transection or neurotoxic lesions of the perirhinal and postrhinal cortices in the rat. *Behav. Brain Res.*, **111**, 187–202.

Bussey, T.J., Muir, J.L. & Aggleton, J.P. (1999a) Functionally dissociating aspects of event memory: The effects of combined perirhinal and postrhinal cortex lesions on object and place memory in the rat. *J. Neurosci.*, **19**, 495–502.

Bussey, T.J., Saksida, L.M. & Murray, E.A. (1999b) Overgeneralization in monkeys with perirhinal cortex lesions. *Soc. Neurosci. Abstr.*, **25**, 789.

Bussey, T.J., Saksida, L.M. & Murray, E.A. (2002) Perirhinal cortex resolves feature ambiguity in complex visual discriminations. *Eur. J. Neurosci.*, **15**, 365–374.

Bussey, T.J., Warburton, E.C., Aggleton, J.P. & Muir, J.L. (1998) Fornix lesions can facilitate acquisition of the transverse patterning task: a challenge for 'configural' theories of hippocampal function. *J. Neurosci.*, **18**, 1622–1631.

Cowey, A. & Gross, C.G. (1970) Effects of foveal prestriate and inferotemporal lesions on visual discrimination by rhesus monkeys. *Exp. Brain Res.*, **11**, 128–144.

Desimone, R. & Ungerleider, L.G. (1989) Neural mechanisms of visual processing in monkeys. In Boller, F. & Grafman, J. (eds), *Handbook of Neuropsychology*, Vol 2. Elsevier Science, New York, pp. 267–299.

Dickinson, A. (1980) *Contemporary Animal Learning Theory* (1st edn.). Cambridge University Press, Cambridge.

Eichenbaum, H. & Bunsey, M. (1995) On the binding of associations in memory: Clues from studies on the role of the hippocampal region in paired-associate learning. *Curr. Dir. Psychol. Sci.*, **4**, 19–23.

Fukushima, K. (1980) Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cybern.*, **36**, 193–202.

Gaffan, D. (1994) Dissociated effects of perirhinal cortex ablation, fornix transection and amygdalectomy: Evidence for multiple memory systems in the primate temporal lobe. *Exp. Brain Res.*, **99**, 411–422.

Gaffan, D., Harrison, S. & Gaffan, E.A. (1986) Visual identification following inferotemporal ablation in the monkey. *Q. J. Exp. Psychol.*, **38B**, 5–30.

Gaffan, D. & Murray, E.A. (1992) Monkeys (*Macaca fascicularis*) with rhinal cortex ablations succeed in object discrimination learning despite 24-hr intertrial intervals and fail at matching to sample despite double sample presentations. *Behav. Neurosci.*, **106**, 30–38.

Gaffan, D. & Parker, A. (1996) Interaction of perirhinal cortex with the fornix-fimbria: Memory for objects and 'object-in-place' memory. *J. Neurosci.*, **16**, 5864–5869.

Glenn, M.J. & Mumby, D.G. (1998) Place memory is intact in rats with perirhinal cortex lesions. *Behav. Neurosci.*, **112**, 1353–1365.

Gluck, M.A. & Myers, C.E. (1995) Representation and association in memory: a neurocomputational view of hippocampal function. *Curr. Dir. Psychol. Sci.*, **4**, 23–29.

Hubel, D. & Wiesel, T. (1962) Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *J. Physiol. (Lond.)*, **160**, 106–154.

Hubel, D. & Wiesel, T. (1965) Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *J. Neurophysiol.*, **28**, 229–289.

Iwai, E. & Mishkin, M. (1968) Two visual foci in the temporal lobe of monkeys. In Yoshii, N. & Buchwald, N. (eds), *Neurophysiological Basis of Learning and Behavior.* Osaka University Press, Japan, pp. 1–11.

Mackintosh, N.J. (1974). *The Psychology of Animal Learning*. Academic Press, London.

McClelland, J.L. & Rumelhart, D.E. (eds). (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. **Vol. 2**. MIT Press, Cambridge, MA.

Meunier, M., Bachevalier, J., Mishkin, M. & Murray, E.A. (1993) Effects on visual recognition of combined and separate ablations of the entorhinal and perirhinal cortex in rhesus monkeys. *J. Neurosci.*, **13**, 5418–5432.

Mumby, D.G. & Pinel, J.P.J. (1994) Rhinal cortex lesions and object recognition in rats. *Behav. Neurosci.*, **108**, 11–18.

Murray, E.A. (2000) Memory for objects in nonhuman primates. In Gazzaniga, M.S. (ed.), *The New Cognitive Neurosciences*. The MIT Press, London, pp. 753–763.

Murray, E.A. & Bussey, T.J. (1999) Perceptual-mnemonic functions of perirhinal cortex. *Trends Cogn. Sci.*, **3**, 142–151.

Murray, E.A. & Mishkin, M. (1998) Object recognition and location memory in monkeys with excitotoxic lesions of the amygdala and hippocampus. *J. Neurosci.*, **18**, 6568–6582.

O'Reilly, R.C. & Rudy, J.W. (2001) Conjunctive representations in learning and memory: principles of cortical and hippocampal function. *Psychol. Rev.*, **108**, 311–345.

Parker, A. & Gaffan, D. (1998) Memory systems in primates: episodic, semantic, and perceptual learning. In Milner, A.D. (ed.), *Comparative Neuropsychology*. Oxford University Press, Oxford, pp. 109–126.

Perrett, D. & Oram, M. (1993) Neurophysiology of shape processing. *Imaging Vis. Comput.*, **11**, 317–333.

Phillips, R.R., Malamut, B.L., Bachevalier, J. & Mishkin, M. (1988) Dissociation of the effects of inferior temporal and limbic lesions on object discrimination learning with 24-h intertrial intervals. *Behav. Brain Res.*, **27**, 99–107.

Rescorla, R.A. & Wagner, A.R. (1972) A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In Black, A.H. & Prokasy, W.F. (eds), *Classical conditioning II: current research and theory*. Appleton-Century-Crofts, New York.

Riesenhuber, M. & Poggio, T. (1999) Hierarchical models of object recognition in cortex. *Nature Neurosci.*, **2**, 1019–1025.

Rothblat, L.A. & Kromer, L.F. (1991) Object recognition memory in the rat: the role of the hippocampus. *Behav. Brain Res.*, **42**, 25–32.

Rumelhart, D.E. & McClelland, J.L., eds. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. **Vol. 1**. MIT Press, Cambridge MA.

Sakai, S. & Miyashita, Y. (1993) Memory and imagery in the temporal lobe. *Curr. Opin. Neurobiol.*, **3**, 166–170.

Saksida, L.M., Bussey, T.J. & Murray, E.A. (2000) One-pair discrimination learning in monkeys with perirhinial cortex lesions. *Soc. Neurosci. Abstr.*, **26**, 546.

Shepard, R.N. (1987) Toward a universal law of generalization for psychological science. *Science*, **237**, 1317–1323.

Squire, L.R. (1992) Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans. *Psychol. Rev.*, **99**, 195–231.

Sutherland, R.J. & Rudy, J.W. (1989) Configural association theory: The role

of the hippocampal formation in learning, memory, and amnesia. *Psychobiology*, **17**, 129–144.
Suzuki, W. & Amaral, D.G. (1994) Perirhinal and parahippocampal cortices of the macaque monkey: Cortical afferents. *J. Comp. Neurol.*, **350**, 497–533.
Suzuki, W.A., Zola-Morgan, S., Squire, L.R. & Amaral, D.G. (1993) Lesions of the perirhinal and parahippocampal cortices in the monkey produce long-lasting memory impairment in the visual and tactual modalities. *J. Neurosci.*, **13**, 2430–2451.
Thornton, J.A., Rothblat, L.A. & Murray, E.A. (1997) Rhinal cortex removal produces amnesia for preoperatively learned discrimination problems but fails to disrupt postoperative acquisition and retention in rhesus monkeys. *J. Neurosci.*, **17**, 8536–8549.
Tsunoda, K., Yamane, Y., Nishizaki, M. & Tanifugi, M. (2001) Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nature Neurosci.*, **4**, 832–838.
Ungerleider, L.G. & Mishkin, M. (1982) Two cortical visual systems. In Ingle, D.J., Goodale, M.A. & Mansfield, R.J.W. (eds), *Analysis of Visual Behavior*. MIT Press, Cambridge, MA, pp. 549–586.
Wallis, G. & Rolls, E.T. (1997) Invariant face and object recognition in the visual system. *Prog. Neurobiol.*, **51**, 167–194.
Zola-Morgan, S., Squire, L.R., Amaral, D.G. & Suzuki, W.A. (1989) Lesions of perirhinal and parahippocampal cortex that spare the amygdala and hippocampal formation produce severe memory impairment. *J. Neurosci.*, **9**, 4355–4370.
Zola-Morgan, S., Squire, L.R. & Ramus, S.J. (1994) Severity of memory impairment in monkeys as a function of locus and extent of damage within the medial temporal lobe memory system. *Hippocampus*, **4**, 483–495.

Appendix: computational details

Network overview

The network consists of three layers of continuous units: a 'feature' layer, a 'feature conjunction' layer, and an 'outcome' node (see Fig. 2). Weights on the links between the feature layer and each feature conjunction layer unit are set upon presentation of a novel stimulus such that a new feature conjunction layer unit comes to represent the input pattern. The feature layer units and the feature conjunction layer units are each linked to an outcome representation, and the weight on the link reflects the strength of association between a stimulus representation and a given outcome. When an input is presented, the feature layer units are activated depending on the particular features present in the stimulus, and each feature conjunction layer unit is activated according to the similarity of the input to the stimulus it was originally recruited to represent. If no feature conjunction layer units are activated above a threshold, a new configural unit is recruited to represent the input and the weights are set such that presentation of the input causes the configural unit to fire maximally. Each layer computes an independent response, for each layer the sum of unit activations multiplied by their associative strengths determines the strength of the response (R). All associative weights are updated in proportion both to the degree of activation of the feature or feature conjunction layer unit and to the discrepancy between the response and the outcome.

Recruitment of feature conjunction units

A stimulus **S** is represented as a vector of elements ranging in value between 0 and 1, a set of which represents the value of the stimulus on a particular 'feature'. The feature layer **F** is in a one-to-one relationship with **S**. On a training trial, presentation of a stimulus leads to a pattern of activation on **F** that consists of the normalized stimulus vector.

$$a_j = \frac{a_i}{\sqrt{\sum_n a_n^2}} \tag{1}$$

where a_j is the activation of feature layer unit j, a_i is the activation of stimulus element i, and i is in the same relative location in **S** as j is in **F**.

Every unit in **F** is connected to each unit in the subsequent feature conjunction layer **C**. The link between one feature layer unit j and one feature conjunction layer unit k is called w_{jk}. Thus for feature conjunction layer unit k there exists a corresponding set of weighted links, $\mathbf{w}_k$, that consists of the set of links between each unit in the feature layer and unit k.

When a stimulus is presented to the network, the Euclidean distance between $\mathbf{w}_k$ and the pattern of activation on the feature layer are calculated (see Equation 3) and compared (see Equation 2) for each feature conjunction layer unit.

$$k^* = \arg\min(\mathbf{d}) \tag{2}$$

$$d_k = \sqrt{\sum_j (a_j - w_{jk})^2} \tag{3}$$

The feature conjunction layer unit k^* whose weights are the smallest Euclidean distance from the pattern of activation on the feature layer is called the 'winner'. Its activation level (a_{k^*}) is

$$a_{k^*} = 1 - d_{k^*} \tag{4}$$

If a_{k^*} is not above a threshold t, then a new feature conjunction layer unit (i.e. one with $\mathbf{w}_k$ equal to zero) is randomly selected and called the winner (k^*). The weights between the feature layer and the new k^* ($\mathbf{w}_{k^*}$) are set to be equivalent to the pattern of activation on the input layer (**a**). This ensures that when the same input stimulus is shown again, k^* will fire maximally.

Activation of the remaining feature conjunction layer units is determined by an exponential function (Shepard, 1987) as follows:

$$a_k = \exp[-q\,(1 - d_k)] \tag{5}$$

This steep generalization gradient means that a given feature conjunction layer unit strongly prefers a close match between the specific features comprising the conjunction and the features active on the feature layer; in other words, 'the whole is more than the sum of the parts'.

Associative learning

After activation in the feature and feature conjunction layers has been determined, all active feature layer units and feature conjunction layer units become associated with the outcome node to an extent dependent on their degree of activation. The weights on these links between feature layer units or feature conjunction layer units and the outcome unit are known as associative strengths. Subsequent presentations of the stimulus will then activate the same units and lead to performance of a CR.

The experiments in the present study all consist of simultaneous discrimination learning tasks. When an animal is presented with this type of task, two stimuli are presented simultaneously and the subject must indicate a choice by selecting one of the two stimuli. The subject is then either rewarded or not rewarded, depending on the choice. A number of such pairs can be presented concurrently during a session. In the simulations, it is not possible to present two stimuli simultaneously. Thus, on a given trial a stimulus **A** is presented and a CR is computed, then stimulus **B** is presented and a second CR is computed. The response of the network on the trial is then determined probabilistically depending on the relative CRs accorded to the stimuli (see Behaviour section).

Independent CRs are calculated for the feature layer and for the feature conjunction layer. CR_F consists of the sum of the associative strengths V_j of all of the feature layer units scaled by their activation, as shown in Equation 6.

$$CR_F = \sum_j V_j \times a_j \tag{6}$$

where a_j is the normalized activation level of feature layer unit j, as determined by Equation 1.

CR_C consists of the sum of the associative strengths V_k of all of the feature conjunction layer units scaled by their activation, as shown in Equation 7.

$$CR_c = \sum_k V_k \times n_k \tag{7}$$

where n_k is the normalized activation level of feature conjunction layer unit k, as determined by Equation 8.

$$n_k = \frac{a_k}{\sqrt{\sum_n a_n^2}} \tag{8}$$

where a_k is the original level of activation of feature conjunction layer unit k.

The output of the network, *CR*, is the sum of the CR_F and CR_C.

Once a *CR* is produced, all associative weights are updated, via the delta rule (Rescorla & Wagner, 1972), in proportion to the degree of activation of the corresponding feature or feature conjunction layer unit. This is shown in Equations 9 and 10, in which α and β are constant learning rates (as is conventional in the Rescorla–Wagner model), a_j is the activation of feature layer unit j, n_k is the normalized activation of feature conjunction layer unit k, λ refers to the value of the reinforcer, and *CR* is the conditioned response.

$$\Delta V_j = \alpha \beta a_j |\lambda - CR| \quad (9)$$

$$\Delta V_k = \alpha \beta n_k |\lambda - CR| \quad (10)$$

Behaviour

The initial output produced by the model consists of the *CR* to each of the presented stimuli. The actual response is then chosen stochastically as a function of *CR*(A) and *CR*(B). To choose a response, a random number between 0 and 1 is generated. If learning is not yet at asymptote and the number is greater than the sum of *CR*(A) and *CR*(B) then a choice is selected randomly. If *CR*(A) is greater than the random number, then action A is chosen. If *CR*(A) is less than the random number, then action B is chosen. Thus, if A is the correct stimulus, as *CR*(A) increases, so will the likelihood that *CR*(A) is greater than the random number and that a correct response will be made.

Parameters

Input stimuli each consisted of a vector of 100 elements (each representing a feature) with 10 randomly selected elements set to an activation level of one and the remaining 90 elements set to an activation level of zero. The number of possible feature conjunction layer units was also 100. Identical network parameters were used for all simulations presented in this paper: $\lambda = 1.0$ or 0.0, $\alpha = 0.005$, $\beta = 0.20$, $t = 0.75$, $q = 20$.