

Functional Dissociations within the Ventral Object Processing Pathway: Cognitive Modules or a Hierarchical Continuum?

Rosemary A. Cowell¹, Timothy J. Bussey², and Lisa M. Saksida²

Abstract

■ We examined the organization and function of the ventral object processing pathway. The prevailing theoretical approach in this field holds that the ventral object processing stream has a modular organization, in which visual perception is carried out in posterior regions and visual memory is carried out, independently, in the anterior temporal lobe. In contrast, recent work has argued against this modular framework, favoring instead a continuous, hierarchical account of cognitive processing in these regions. We join the latter group and illustrate our view with simulations from a computational model that extends the perceptual-mnemonic feature-conjunction model of visual discrimination proposed by Bussey and Saksida [Bussey, T. J., & Saksida, L. M. The organization of visual object representations: A connectionist model of effects of lesions in perirhinal cortex. *European Journal of Neuroscience*, 15, 355–364, 2002]. We use the extended model to revisit early data from Iwai and Mishkin [Iwai, E., & Mishkin, M. Two visual foci in the temporal lobe of monkeys.

In N. Yoshii & N. Buchwald (Eds.), *Neurophysiological basis of learning and behavior* (pp. 1–11). Japan: Osaka University Press, 1968]; this seminal study was interpreted as evidence for the modularity of visual perception and visual memory. The model accounts for a double dissociation in monkeys' visual discrimination performance following lesions to different regions of the ventral visual stream. This double dissociation is frequently cited as evidence for separate systems for perception and memory. However, the model provides a parsimonious, mechanistic, single-system account of the double dissociation data. We propose that the effects of lesions in ventral visual stream on visual discrimination are due to compromised representations within a hierarchical representational continuum rather than impairment in a specific type of learning, memory, or perception. We argue that consideration of the nature of stimulus representations and their processing in cortex is a more fruitful approach than attempting to map cognition onto functional modules. ■

INTRODUCTION

In the first empirical studies of the vastly complex processes underlying human thought and behavior, pioneering 19th century psychologists sought to carve up the nebulous subject matter of cognition into tractable portions. Folk psychology and introspection suggested, not unreasonably, the use of constructs like memory, perception, emotion, and attention. Much investigation of the mind and brain, since the time of William James, has followed a course that assumes functional modularity according to these boundaries within cognition.

In no branch of cognitive neuroscience has the influence of a modular approach been greater than in the study of memory. Assumptions of functional modularity found early support in the discovery of amnesic patient, H.M., whose bilateral removal of the medial-temporal lobes rendered him unable to store new information about facts and events (Bussey, Saksida, & Murray, 2002; Scoville & Milner, 1957). This memory deficit was selective, however, because H.M. did not appear to have any gross percep-

tual deficits and he even showed "perceptual priming" (Warrington & Weiskrantz, 1968), which suggested preservation of his perceptual capacities. Thus, Scoville and Milner's (1957) study of H.M. reinforced the already popular idea that the brain is organized into separable functional systems, in this case for perception and memory. This fascinating case study sparked much experimental work aimed at investigating animal models of human amnesia (e.g., Squire & Zola-Morgan, 1983; Mishkin, 1982). For the most part, experimental investigation was carried out on the assumption that memory processes could be, indeed should be, studied independently of perceptual processes. Spelling out this assumption was barely considered necessary because the notion of a distinction between perception and memory went unchallenged.

More recently, some researchers have made this functional distinction an explicit and a central aspect of theories of memory. Squire and Zola-Morgan (1991), for example, claimed a functional and an anatomical distinction between visual perception and declarative memory. In this highly influential article, the authors proposed that human declarative memory is mediated by a group of brain structures in the medial-temporal lobe (MTL). They also claimed there

¹University of California, San Diego, ²University of Cambridge, UK

is a functional dissociation between areas in inferotemporal cortex (IT) thought to mediate perception and structures within MTL that are said to support declarative memory (Shrager, Gold, Hopkins, & Squire, 2006; Levy, Shrager, & Squire, 2005; Buffalo, Ramus, Squire, & Zola, 2000; Stark & Squire, 2000; Buffalo et al., 1999; Suzuki, Zola-Morgan, Squire, & Amaral, 1993). The MTL memory system rapidly became the dominant theoretical construct in memory research (Squire, Stark, & Clark, 2004). Through the wide acceptance of this theory, the assumption of dissociable systems for perception and memory continues to prevail and has found support outside of neuropsychology in the domain of neurophysiology.

A Modular Approach to Visual Perception and Visual Memory

The work by Warrington and Weiskrantz (1968), Scoville and Milner (1957), and Squire et al. (2004) was interpreted as evidence for separate systems underlying visual perception and “declarative” or “episodic” memory. In this article, we examine the specific case of visual memory and perception, which are widely assumed to be functionally and anatomically distinct.

The empirical origins of this assumption lie in a substantial literature from the 1960s and 1970s, in which monkey researchers used a task known as “visual discrimination learning” and investigated the effects of lesions in both anterior and posterior areas of the ventral visual stream (VVS). Many studies revealed a dissociation between the behavioral effects of damage to the two regions, anterior and posterior. This dissociation was repeatedly interpreted as support for a functional distinction, with memory mediated in anterior areas and perception in posterior areas.

Some of the visual discrimination learning studies interpreted as evidence for such a functional distinction found only single dissociations. In some cases, this was because authors included in their experimental design lesions of only one area instead of several; in other cases, only one type of discrimination task was used rather than two (Kikuchi & Iwai, 1980; Dean, 1974; Butter, 1972; Iversen & Humphrey, 1971; Manning, 1971a, 1971b; Wilson & Kaufman, 1969). Notably, even these authors interpreted their data in terms of the (widely assumed) functional distinction between “perception” and “associative memory.” However, critical to any account of cognitive function in which different brain regions are said to mediate distinct functions is the demonstration of a double dissociation within one experiment. Several authors did find double dissociations within one study, by using two or more lesion groups with ablations at different points in the VVS and two or more discrimination tasks or experimental manipulations (Blake, Jarvis, & Mishkin, 1977; Wilson, Kaufman, Zieler, & Lieb, 1972; Gross, Cowey, & Manning, 1971; Cowey & Gross, 1970; Iwai & Mishkin, 1968). This work well characterizes the theoretical approach that led to the modular view of memory and perception that continues to prevail

in modern neuroscience and psychology (e.g., Buffalo et al., 1999; Sakai & Miyashita, 1993; Tulving & Schacter, 1990).

THE PERCEPTUAL-MNEMONIC FEATURE-CONJUNCTION MODEL OF PERIRHINAL CORTEX FUNCTION

Recently, some have begun to question the modular approach to visual cognition in favor of continuous accounts of temporal lobe function, in which memory and perception interact (e.g., Palmeri & Tarr, 2008; Bussey & Saksida, 2007; Palmeri & Gauthier, 2004; Gaffan, 2002).

Bussey and Saksida (2002) and Murray and Bussey (1999) have presented a new theoretical framework for visual object processing and visual recognition memory. They have focused on perirhinal cortex (PRh), which lies adjacent to anterior IT and receives most of its input from visual areas (Suzuki & Amaral, 1994). The role of PRh in recognition memory is well established (Eacott, Gaffan, & Murray, 1994; Meunier, Bachevalier, Mishkin, & Murray, 1993; Suzuki et al., 1993; Gaffan & Murray, 1992). However, Bussey and Saksida and Murray and Bussey proposed that PRh can be thought of as part of the VVS, suggesting an additional role for PRh in perception. The proposed system contains hierarchically organized representations of visual objects (Desimone & Ungerleider, 1989). Progressing through the hierarchy from posterior to anterior regions, simple features are combined into complex conjunctions (Figure 1), with the most complex representations—at the level of complexity corresponding to real-world objects—contained in PRh. Deficits in perception arising from PRh lesions are assumed to be due to the loss of conjunctive representations. Bussey and Saksida instantiated this idea in a connectionist model (Figure 2); PRh lesions were simulated by removing the component of the network corresponding to PRh. The perceptual-mnemonic feature-conjunction (PMFC) model accounted for a puzzling set of findings concerning the effects of PRh lesions on visual discrimination.

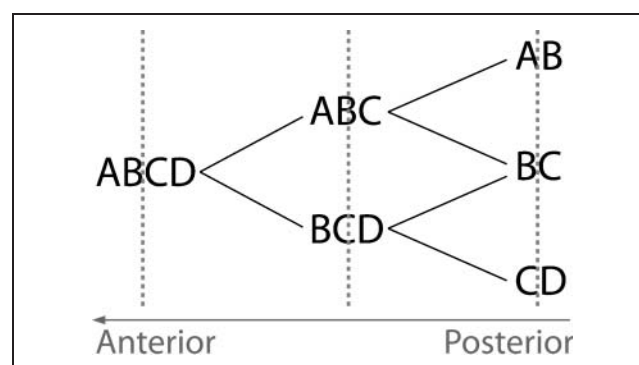


Figure 1. The hierarchical organization of stimulus representations in VVS. Each letter represents a visual feature. The number of features combined conjunctively into a single representation increases with progression through the hierarchy: simple patterns are represented in posterior regions and complex objects in anterior areas.

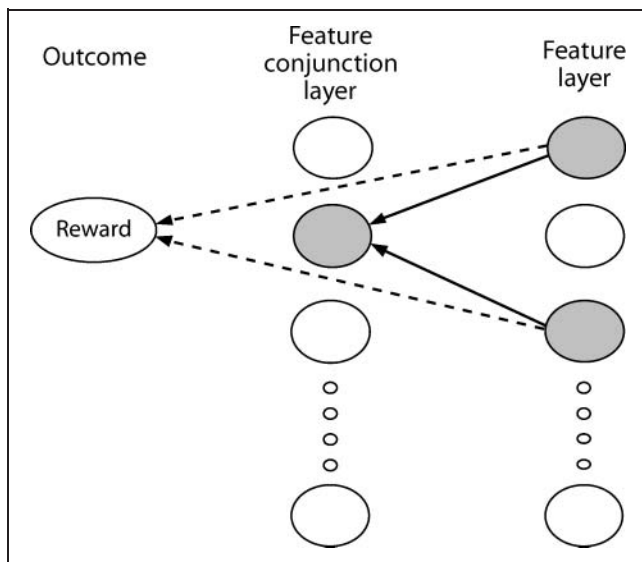


Figure 2. Architecture of the PMFC model. The network consists of two layers of nodes and a single outcome node whose activation represents an event, for example, reward. Active units are shown in gray. Feature layer units are connected to units in the feature-conjunction layer via fixed weights. Both layers are connected to the outcome node via weights that are adjustable by an associative mechanism. For simplicity, only four nodes are shown in each layer, although each layer has up to 100 nodes in total, and only two feature layer nodes are shown as active, although stimuli contained at least 10 active nodes.

Further, it made novel predictions that were supported by visual discrimination learning experiments in monkeys (Bussey, Saksida, & Murray, 2003; Bussey et al., 2002). Recently, Cowell, Bussey, and Saksida (2006) have extended this view to account for deficits in recognition memory following lesions to PRh.

Since publication, the PMFC model of PRh function has been the center of much controversy. It has been challenged by proponents of the modular view (Hampton, 2005; Squire et al., 2004). A considerable body of empirical research has emerged on the basis of the PMFC model, over and above experimental tests of the model authored by Bussey and Saksida (2002). Many authors have either tested the predictions of the PMFC model directly or examined the validity of the general framework advocated by the PMFC model. These studies have included work with rodents (Norman & Eacott, 2004; Gilbert & Kesner, 2003), nonhuman primates (Baker, Behrmann, & Olson, 2002), and human subjects (Preston & Gabrieli, 2008; van Strien, Scholte, & Witter, 2008; Barense, Gaffan, & Graham, 2007; Devlin & Price, 2007; Hartley et al., 2007; Lee, Bandelow, Schwarzbauer, Henson, & Graham, 2006; Lee, Buckley, et al., 2006; Shrager et al., 2006; Levy et al., 2005; Moss, Rodd, Stamatakis, Bright, & Tyler, 2005; Tyler et al., 2004; Stark & Squire, 2000). Of these studies, several support the modular view (Shrager et al., 2006; Levy et al., 2005; Stark & Squire, 2000) and the remainder are either in favor of or consistent with the PMFC model. Because of the challenge these ideas and findings have presented to traditional modular theory, three re-

cent articles debate the pros and the cons of this new view (Baxter, 2009; Suzuki, 2009; Suzuki & Baxter, 2009).

Extending the PMFC Model Account

The PMFC model proposes that PRh plays an important role in perception by providing complex conjunctive representations of stimuli that are necessary for visual discriminations in which the stimuli possess ambiguous features. These high-level representations are thought to help resolve what has been referred to as *feature ambiguity*. However, we do not claim that PRh is the only region of the brain in which conjunctive representations exist and play a role in cognitive function. We argue that there is an important role for conjunctive representations in other areas, such as those of lesser complexity than those in PRh, residing in regions upstream of the VVS. In this article, to investigate the contributions of these simpler conjunctive representations, we extend the PMFC model in a posterior direction. We use the extended model to account for data from studies in which posterior VVS has been lesioned.

The focus of the present article is the function of the whole of VVS and not just PRh. The critical aim of extending the model is to demonstrate that the *relative* difference in the complexity of representations in any two brain regions can be mapped onto the *relative* difference in the functional contributions of those two regions. That is, if the first brain region contains more complex representations than the second, it will play a more important role in discriminating more complex objects than the second region. In contrast, the second region will play a more critical role in discriminating simpler objects than the first region. In all simulations, the difference in functional role of the anterior “complex” layer relative to posterior “simple” layers remains the same, whether the anterior layer is intended to correspond to area TE (e.g., Simulation 1) or PRh (e.g., Simulations 2 and 3).

Empirical Data: Iwai and Mishkin (1968)

The optimal empirical data against which to test the continuous processing account would come from groups of human patients with brain damage in sequential regions of the VVS. The account predicts that a continuous pattern of deficits in discrimination and memory performance would be revealed across the groups by careful choice of stimulus material. There is a paucity of patients with clean focal damage in these areas; however, a rich source of data of exactly the type required is available in the monkey literature. Of the research cited above, Iwai and Mishkin (1968) is the ideal example: a comprehensive study in which both anterior and posterior lesions were tested on two different tasks and a double dissociation found.

Iwai and Mishkin (1968) took five groups of monkeys, retaining one group as unoperated controls and giving lesions to four other groups. One of the four operated groups

received sham lesions as a control; the second group were lesioned in anterior VVS, starting 10 mm anterior to the inferior occipital sulcus and extending 10 mm anteriorly from this point (Group III + IV); a third group received a more posterior lesion extending from the inferior occipital sulcus to a line 10 mm anterior to it (Group I + II); a fourth group received a posterior lesion overlapping with the last, which included the convexity posterior to the inferior occipital sulcus, and the region of cortex between the inferior occipital sulcus and a line 5 mm anterior to it (Group O + I).

The monkeys were trained on several different visual discrimination tasks, two of which are of interest here: a task termed “pattern relearning” and another named “concurrent object discrimination learning.” Pattern learning was first presented to all groups preoperatively. Monkeys were trained on a single-pair discrimination of two white patterns—a plus sign and the outline of a square—on a gray background, until attainment of criterion. After surgery, animals were retrained on the same problem to a criterion of 90 correct responses in 100 trials, yielding a “relearning” score. In the concurrent object discrimination learning task, subjects were trained on eight object-pair discriminations concurrently, with five randomly ordered daily presentations of each pair, until a criterion of 39 correct responses in a session of 40 trials was attained.

The results of this study are shown in Figure 3, which indicates that the two posterior groups (O + I and I + II) relearned the pattern discrimination task more slowly than the anterior group (III + IV). Statistical tests revealed that all group differences in the “trials to criterion” score

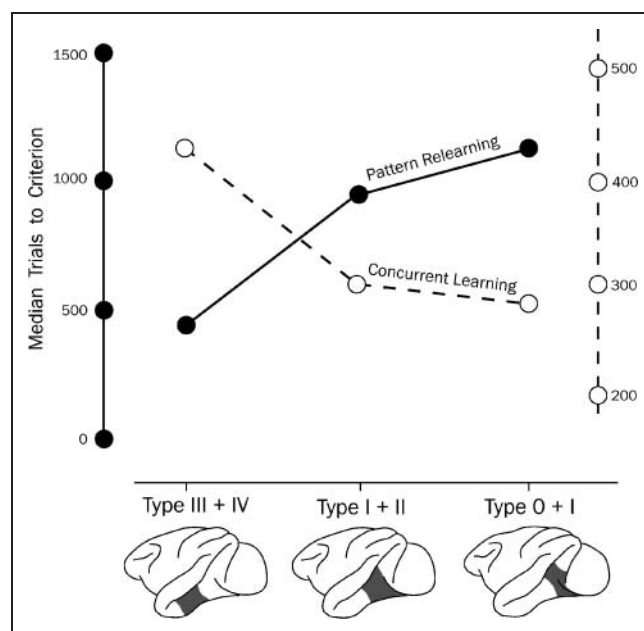


Figure 3. Data from Iwai and Mishkin (1968). Median trials to criterion made by monkeys in each of the three lesion groups on two visual discrimination learning tasks. Note that data for the two tasks are not shown on the same axis.

were significant, except those between the two posterior groups (O + I and I + II) and between the two control groups. On the concurrent object discrimination task, the trend was reversed: monkeys with anterior lesions learned more slowly than animals in groups I + II and O + I. The difference between the most anterior (III + IV) and the most posterior (O + I) groups was significant but that between the anterior group and the adjacent posterior group (I + II) was not.

The authors interpreted this double dissociation in terms of “two qualitatively different disorders: (1) a sensory, perceptual, or attentional loss and (2) a defect in memory or associative learning.” As discussed above, this modular interpretation has prevailed, and a version of it is currently the textbook view. However, does this modular model provide the only account of these data? In the present article, we explore the possibility that the effects of VVS lesions on visual discrimination learning performance are due to the animals possessing compromised representations of visual stimuli rather than an impairment of a specific type of learning, memory, or perception.

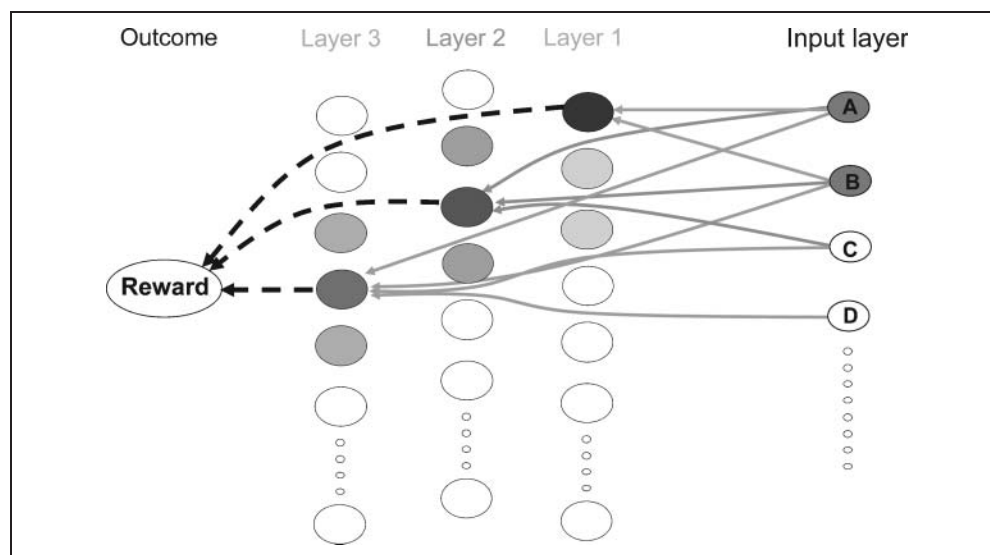
THE EXTENDED PMFC MODEL OF VVS FUNCTION

Our goal was to test whether the function of not only PRh but also the whole of VVS can be understood in terms of conjunctive representations and the resolution of feature ambiguity. We have extended the PMFC model to allow simulation of the effects of lesions in both anterior and posterior VVS. In the original PMFC model, there is only one layer containing conjunctive representations, so the model cannot account for the effects of using stimuli with a lesser level of complexity. In the PMFC model simulations, it was simply assumed that stimuli for which “perirhinal” layer lesions induced discrimination impairments were of a complex level. In the extended PMFC model, additional layers representing different levels of stimulus complexity allow explicit simulation of lesion effects on discrimination of different stimulus types.

The extended PMFC model continues to assume a hierarchical organization of representations in VVS, with simple visual features represented in posterior regions and more complex conjunctions of those features represented in anterior areas. However, the new model has three important alterations. First, we have added a further layer so that there are representations of intermediate complexity as well as simple and complex representations. Second, we have added an input layer to the front of the model allowing convergence of visual features in even the simplest layer. Third, the lateral inhibition function that was applied only to the conjunction layer of the PMFC model is now applied to all three layers.

At all layers in the network, visual features converge into conjunctions; the number of features integrated into a conjunction increases through the layers. In each layer,

Figure 4. Architecture of the extended PMFC model. The network consists of an input layer, three layers of units and a single outcome node. Layer 1 units represent the conjunction of two visual features, Layer 2 units the conjunction of three visual features, and Layer 3 units the conjunction of four visual features. Active units are shown in gray; darker shading indicates higher activation. A “simple,” two-featured stimulus is shown. Weights from the input layer to Layers 1, 2, and 3 are fixed. Layers 1, 2, and 3 are connected to the outcome node via weights that are adjustable by an associative mechanism.



conjunctive combinations of the 32 input features are represented on the basis of the assumption that an adult animal has learned about conjunctions of features occurring commonly in real-world patterns and objects. Weights connecting Layers 1, 2, and 3 to the outcome node are learned, reflecting an animal's ability to learn associations of real-world objects with events such as reward.

In Layer 1, input stimuli are represented as conjunctions of two stimulus dimensions or “visual features” (Figure 4). These correspond to representations of simple stimulus properties found in posterior regions of the VVS (Hubel & Wiesel, 1962). In Layer 2, the stimulus representations combine three visual features into a conjunction, and in Layer 3 they reach a maximum of complexity with four visual features brought together. This organization reflects the increasing complexity of neural representations found in VVS (Desimone & Ungerleider, 1989; Desimone & Schein, 1987; Desimone, Albright, Gross, & Bruce, 1984). In the input layer of the model, there are 32 input units, hence 32 available “visual features.” Because the number of possible feature combinations in each successive layer increases exponentially and yet there does not exist a combinatorial explosion of the number of neurons at successive stations of the VVS, we chose to keep the number of units in each layer approximately equal. Specifically, in Layer 1, we created a unit for all possible two-feature conjunctions of the 32 input stimulus dimensions, giving a tractable number of 496 units. In Layers 2 and 3, to keep the number of units in each layer equal, we initialized units for only a subset of the possible permutations of feature conjunctions at the level of complexity assigned to that layer. That is, in Layer 2, we again created 496 units, each one representing a three-feature conjunction; these conjunctions were selected at random from the total 4960 possible ways of combining 32 features into a three-feature conjunction. In Layer 3, we created 496 units—approximately 1% of the 35,960 possible four-feature combinations of 32 features. This initiali-

zation was based on the assumption that an adult animal possesses a set of well-established visual representations in VVS commensurate with its visual experience. Because simpler conjunctions occur frequently in visual objects, representations of all such conjunctions exist in the adult brain (and in Layer 1 of the model). However, as the complexity of a conjunction increases, the likelihood of an animal having seen that particular conjunction enough for its representation to become established in visual cortex reduces. In line with evidence for the establishment of new visual representations in cortex during visual discrimination learning (Baker et al., 2002), we allowed a unit for a particular conjunction of features to be “recruited” in Layer 2 or Layer 3 on presentation of that exact conjunction during training.

We noted that constraining all layers in the model to contain approximately the same number of units has no significance for the “computational” or “algorithmic” levels (to use Marr's terms) of our theory. This constraint has some influence on the equations used to calculate the lateral inhibition function. However, we could have implemented the lateral inhibition function in any one of several different ways, each of which would have been in line with the neurobiological data; we are not wedded to the implementational details that we present here. We used this constraint because the lack of a combinatorial explosion of the number of units in each layer reflects the properties of the brain.

The VVS Hypothesis

The extended connectionist network now has three layers of stimulus representations; each can resolve feature ambiguity at a certain level of stimulus complexity. The central tenet of our account of VVS function is that representations at a given stage of processing in the VVS provide the optimal solution for a given discrimination problem, according to

the level of complexity of the stimuli in that problem (Zhang & Cottrell, 2005; Ullman, Vidal-Naquet, & Sali, 2002). However, other processing stages contained in regions outside of the optimal processing stage may also provide suboptimal solutions for the discrimination; the closer a region lies to the optimal stage, the better its solution will be. Thus, there is a continuous gradation in the ability of different VVS regions to solve a particular discrimination, reflecting the continuous gradation in the complexity of stimulus representations along the VVS (Tanaka, Saito, Fukada, & Moriya, 1991).

Lesion effects can be explained according to this scheme as follows. If, when a monkey performs a particular discrimination, the “optimal” stage in the VVS for solving the discrimination lies within the animal’s brain lesion, discrimination performance will be severely impaired. In general, the closer the lesion falls to the processing stage best able to represent unambiguously the discriminanda, the worse an animal’s performance will be. This property allows us to simulate the data from Iwai and Mishkin (1968). These authors reported not an “all-or-nothing” double dissociation but an increasing and graded deficit in visual discrimination learning in both the anterior-to-posterior and the posterior-to-anterior directions.

Biological Plausibility

The present model is constrained and inspired by anatomical and physiological properties of visual cortex. For example, the existence of the “whole-preferring” conjunctive representations critical to the model’s mechanism was supported by data from Baker et al. (2002). In addition, lateral inhibition is known to occur in visual cortex and can be critical to the selective responses of visual neurons (Eysel, Worgotter, & Pape, 1987; Sillito, Kemp, Milson, & Berardi, 1980; Sillito, 1979).

Neurophysiological work revealing properties of individual neurons in the VVS has been the key to the notion of hierarchical organization of representations in visual cortex (e.g., Desimone & Ungerleider, 1989; Desimone & Schein, 1987; Desimone et al., 1984; Hubel & Wiesel, 1962). Anatomical data on the connectivity of the VVS also provide support for the architecture of the model. Desimone and Ungerleider (1989) summarized the interconnections between cortical visual areas in the macaque. Neurons project from occipital V1 directly to V2 and to V3; in addition, there are interconnections from V2 to V3. From V2 and V3, projections carry visual information further forward still, to V4, from where projections to both anterior and occipital TE are reported. Thus, visual information travels through a series of cortical stations in a posterior-to-anterior direction from the occipital to the temporal lobe. In addition, there is evidence that “jumping” connections in the VVS are also abundant, for example, between area V1 and V3, bypassing area V2 (Lennie, 1998), from V2 to TEO (Nakamura, Gattass, Desimone, &

Ungerleider, 1993), and from V4 to posterior TE (Saleem, Tanaka, & Rockland, 1992). Given the evidence for jumping projections, which convey information in the absence of intermediate cortical stages, we chose to model only the jumping projections rather than the serial ones because it vastly simplifies the mechanism of the model. This simplification allows us to focus on our central hypothesis concerning the consequences of damage to stimulus representations. Because we can account for the data with a simple, parallel architecture and because cortical stations are connected to some extent in parallel, we opt for parsimony.

There is neurobiological evidence to justify the association of representations at each layer of the model, including the earlier layers, with reward. It seems likely that representations throughout extrastriate visual cortex can influence behavior directly because the primate extrastriate visual cortex projects directly to the tail of the caudate nucleus (Wilson, 1995), which in turn projects to regions involved in executive control and motor responding (e.g., the prefrontal and premotor cortices via the globus pallidus and thalamus). However, we made no explicit claims about the brain location of the “outcome node” in our model.

Although the network architecture is inspired by anatomy and neurophysiology, our account of visual cognition remains abstract. The organization of representations and the learning mechanisms are putatively mediated at the level of anatomical systems. Activation of a unit does not represent firing in an individual neuron nor is the idea of a “grandmother cell” endorsed. The neural entity most probably corresponding to a unit in the network is a population of interconnected neurons that code for a given visual representation.

We make no attempt to include low-level properties of neurons that are not necessary to the proposed mechanism for visual discrimination, for example, back projections or “repetition suppression” in IT neurons. This is intentional. We aim first to define our target data (i.e., our problem space) and then to account for those data using the simplest possible model and the appropriate level of analysis. We define our problem space as the data from neuropsychological studies investigating the role of VVS in visual discrimination. In accordance with Occam’s razor, we use the simplest model possible to explain the data. That is, we have not included details of top-down processing from higher cortical areas because they are not necessary to capture the important patterns in the target data. Further, because our account of the target data hinges on the complexity of stimulus representations at different points in the brain, we have chosen the level of analysis most appropriate to this: a simple connectionist network in which the properties of the representations are clearly defined and play a critical role in determining discrimination performance. To include details of top-down processing known to exist in the brain would expand our problem space. This is, of course, a legitimate step, but we choose not to take it because it would decrease the parsimony of

the account and obscure the key mechanism at work in the model.

Other Computational Models of the Object Processing Pathway

Gaffan, Harrison, and Gaffan (1986) presented a computational model that simulated monkeys' acquisition of visual discrimination problems. Attributes of visual stimuli were schematically represented and associated with "good" or "bad" outcomes to examine the hypothesis that elementary visual features rather than an integrated object whole may have an important role in visual discrimination performance. This model has parallels with the extended PMFC model. However, an important difference is that Gaffan et al. did not include conjunctive representations—a crucial aspect of the present model. Further, they did not attempt to map the representations of visual attributes to different subregions of temporal cortex.

Hubel and Wiesel's (1962, 1965) widely accepted hierarchical model of VVS organization forms the basis of many model of object recognition in visual cortex (e.g., Wallis & Rolls, 1997; Perrett & Oram, 1993; Fukushima, 1980). These models often address electrophysiological data, capturing properties such as the translation-invariant responses of IT cells. However, Riesenhuber and Poggio (1999, 2000) proposed a model of processing in VVS that also addresses high-level processes. In their model, object representations are built up in a hierarchical, convergent fashion and used as input to task modules that learn to perform identification and categorization. There are many similarities of Riesenhuber and Poggio's approach to the extended PMFC model, but there are two key differences. First, the "perceptual" and "higher order" components of their model operate in serial fashion: only the final layer of object representations provides output to the task modules. In the extended PMFC model, representations in every layer are associated with an outcome, so that later stages of perceptual processing need not be complete before elementary perceptual representations can be used in a higher-order process. Thus, perception and learning are integrated rather than independent and serial. Second, the two models address different data sets. Riesenhuber and Poggio draw on electrophysiological data to solve the problems of view and location invariance. Our network addresses data from lesion studies to investigate the hypothesis that cognitive systems operate in a unitary, continuous fashion.

The present model also resembles the configural-cue model of Gluck and Bower (1988). Both models assume that all possible combinations of the elements of the input stimuli are represented in the sensory layer(s) of the model. Both models allow all sensory representations to become associated directly with an outcome. Where the extended PMFC model diverges from Gluck and Bower's network is in the use of a lateral inhibition function on each layer of the model. Thus, a "winner-take-all" process op-

erates independently on each subset (layer) of stimulus representations, according to their complexity. This property also allowed us to account for critical findings in the area of brain damage and object recognition memory (Cowell et al., 2006). A more important divergence between the two models lies in the cognitive phenomena they attempt to explain. Gluck and Bower used simple visual stimuli (which all possess the same number of elements) and category structures defined by rules that range in complexity from very simple to very complex, because the goal is to investigate the processes underlying category learning. The extended PMFC model uses simple discrimination problems (which tend to possess a similar degree of overlap) and a range of complexity of visual stimuli, because the goal is to investigate how impairments in visual discrimination arise from brain lesions at different points in visual cortex.

SIMULATION 1: IWAI AND MISHKIN REVISITED

We simulated Iwai and Mishkin (1968), a paradigm case of the nonhuman primate literature on visual discrimination learning. To simulate the pattern relearning task, networks were trained to discriminate a single pair of similar simple patterns. To simulate concurrent discrimination learning, networks learned to discriminate eight pairs of complex objects.

Methods

In all simulations, lesions of the network were effected by completely removing the layer corresponding to the cortical area lesioned in Iwai and Mishkin (1968). Computational details of the connectionist model are given in the Appendix.

Pattern Relearning

Stimuli. Simple visual patterns were represented by stimuli possessing two "visual features": 2 of the 32 units in the input layer were activated. We used stimuli with an overlapping feature so that the discrimination problem took the form AB+ vs. BC− (where a letter denotes an individual visual feature or active unit, two letters represent a whole stimulus, "+" indicates that the stimulus was rewarded, and "−" that it was nonrewarded). Because the stimuli used by Iwai and Mishkin (1968)—a plus sign and a square—possessed some distinct features and some common features, these inputs seem reasonable representations of the stimuli they used.

Training. Networks were trained to discriminate between a single pair of input patterns, as in Iwai and Mishkin (1968). Training involved two phases: prelesion and postlesion acquisition of the discrimination. In "preoperative" Phase 1, 24 networks were initialized and trained on a

single pair discrimination using the pattern relearning stimuli until discrimination performance reached criterion. At the end of Phase 1, the networks were divided into three groups of eight networks and lesioned: in group “posterior,” Layer 1 was removed; in group “middle,” Layer 2 was removed; and in group “anterior,” Layer 3 was removed. In “postoperative” Phase 2, networks were not reinitialized before retraining on the same single pair discrimination that had been acquired before removal of a layer of units. Training in Phase 2 also proceeded until criterion was reached. In Phases 1 and 2, criterion was reached when 9 of 10 correct responses were made in two consecutive blocks of 10 trials.

Concurrent Discrimination Learning

Stimuli. Complex, three-dimensional objects were represented by activating 4 out of a possible 32 units in the input layer for each stimulus. Thus, in the full set of 16 stimuli (or eight pairs), there were many visual features that were possessed by more than one stimulus, reflecting the occurrence of overlapping features across a set of real-world objects.

Training. In this task, training occurred in a single, postoperative phase, as in Iwai and Mishkin (1968). Three groups of eight networks—groups posterior, middle, and anterior, as in the pattern relearning task—were initialized and trained to criterion on the concurrent discrimination of eight pairs of stimuli. Criterion was reached when 9 of 10 correct responses were made in two consecutive blocks of 10 trials.

Results

The simulations bear a striking similarity to the data of Iwai and Mishkin (1968). As is shown in Figure 5, networks lacking Layer 1 (group posterior) were impaired relative to networks lacking Layer 3 (group anterior) on the pattern relearning task. However, on the concurrent discrimination learning task, the group anterior was impaired relative to the group posterior. Networks lacking Layer 2 (group middle) showed performance levels between those of groups posterior and anterior on both tasks.

In this simulation and all others in the present article, any data points lying further than two standard deviations from the mean were deemed outliers and removed from the analysis. We assessed the statistical significance of group differences in Simulation 1 with a one-way ANOVA with factor Lesion Group for each task. The mean scores of each group differed significantly from one another on both Pattern Relearning, $F(2, 20) = 18.09, p < .001$, and Concurrent Discrimination Learning, $F(2, 20) = 8.16, p < .01$. Sidak multiple comparisons showed that the group posterior performed significantly worse than the group anterior on pattern relearning ($p < .001$), whereas the group anterior performed significantly worse than the group posterior on concurrent discrimination learning ($p < .01$).

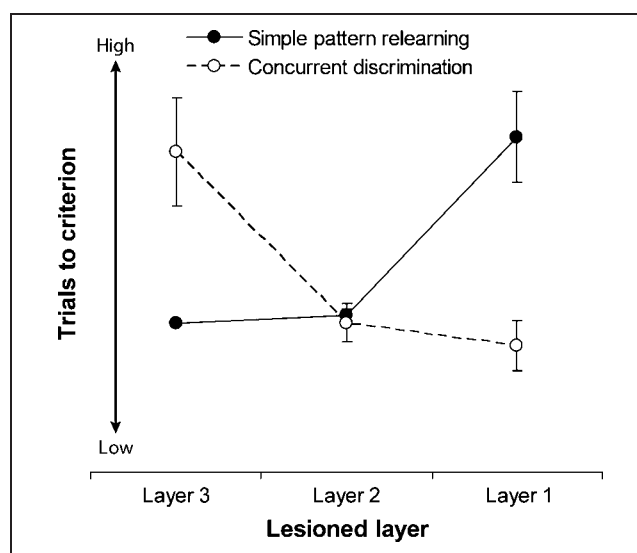


Figure 5. Performance of the extended PMFC model in simulations of the pattern relearning and concurrent learning tasks of Iwai and Mishkin (1968). Scores on the two tasks are shown on different scales. Error bars indicate $\pm SEM$.

Discussion

The simulations match very closely the double dissociations reported in studies such as Iwai and Mishkin (1968) and show how a completely different explanation from the modular view can account for these data. According to the extended PMFC model, the area of VVS critical for a given visual discrimination task depends on the level of complexity of conjunctive representations required to disambiguate the stimuli used in that task. If animals are required to discriminate simple patterns possessing simple conjunctions of few visual features—as in the pattern relearning task—the conjunctive representations in posterior regions are critical for good performance. Conversely, if they must discriminate complex objects possessing complex conjunctions of many visual features—as in the concurrent object discrimination task—representations in anterior regions are needed to solve the task efficiently.

SIMULATION 2: RESIMULATING THE FEATURE AMBIGUITY EFFECT

To verify that the extended PMFC model can still account for data simulated by the original PMFC model, we resimulated the visual discrimination studies in which PRh was lesioned (Bussey et al., 2002, 2003). In the original feature ambiguity study (Bussey et al., 2002), the PMFC model predicted, and an experiment in monkeys confirmed, that lesions of PRh should disrupt complex visual discriminations with a high degree of feature ambiguity, a property of visual discrimination problems that can emerge when features of an object are rewarded when they are part of one object but not when part of another.

Methods

The simulation was identical to the feature ambiguity simulation of the original PMFC model, except that stimulus input vectors possessed 32 elements (rather than 100) and each stimulus was composed of 4 active elements (and 28 inactive elements) rather than 20 active elements (and 80 inactive elements). There were three experimental conditions: maximum feature ambiguity, in which all stimulus features were explicitly ambiguous; intermediate feature ambiguity, in which half of the features were ambiguous; and minimum feature ambiguity, with no ambiguous features. To simulate PRh lesions, we removed Layer 3.

Results

The results of the simulation with the extended PMFC model are shown in Figure 6. A one-way ANOVA with Lesion Group as between-subjects factor and Feature Ambiguity as within-subjects factor revealed a significant main effect of Lesion Group, $F(1, 6) = 488.5, p < .001$, a significant main effect of Feature Ambiguity, $F(2, 12) = 63.39, p < .001$, and a significant Feature Ambiguity \times Lesion Group interaction, $F(2, 12) = 32.78, p < .001$. Analysis of simple main effects revealed differences between the two groups in the intermediate ($p < .005$) and maximum ($p < .001$) conditions but not in the minimum condition ($p = .145$).

Discussion

As the degree of feature ambiguity in concurrent discrimination of complex objects was increased, networks with

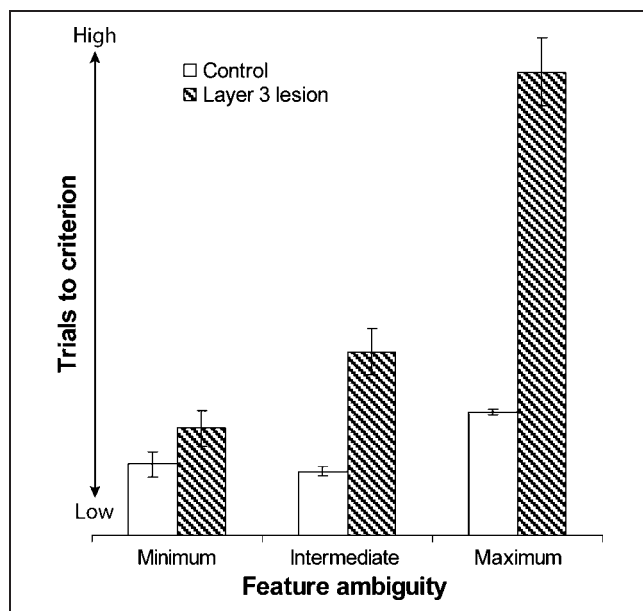


Figure 6. Performance of the extended PMFC model on the simulated feature ambiguity task; the results replicate the simulation data of Bussey et al. (2002). Error bars indicate $\pm SEM$.

a lesion of the most anterior layer became increasingly impaired. This replicates the findings of Bussey et al. (2002) for both connectionist networks and monkeys and demonstrates that the most anterior end of the extended PMFC model can be mapped onto anterior structures such as PRh, like the “conjunction layer” of the original PMFC model.

SIMULATION 3: RESIMULATING THE EFFECT OF MORPHED STIMULI

Next, we used the extended PMFC model to simulate the “morphed stimuli” experiment of Bussey et al. (2003). In this study, the perceptual difficulty of single-pair discriminations was manipulated by blending together pairs of gray-scale picture stimuli to create discriminanda that shared many features. The PMFC model predicted that lesion of PRh would cause impairments when there was a high degree of morphing between the to-be-discriminated stimuli but not with a lesser degree of morphing; the predictions were once more confirmed in monkeys.

Methods

Simulations using the extended PMFC model were almost identical to those with the original PMFC model. In the low-ambiguity stimulus pair of Bussey et al. (2003), 10 elements took a value of 1 in the first stimulus (with the remaining 90 elements set to 0), and 10 different elements took a value of 1 in the second stimulus (see Figure 7). In the high-ambiguity pair, the first stimulus possessed 10 elements with a value of 0.8 and a further 10 with a value of 0.2; in the second stimulus, the 10 elements that had been set to 0.8 in the first stimulus were set to 0.2, and the 10 elements set to 0.2 in the first stimulus were set to 0.8. In the present simulation, we used stimulus input vectors with 32 elements rather than 100. There were four elements (rather than 10) with a value of 1 in each stimulus of the low ambiguity pair; no elements were shared by the two stimuli. In the high ambiguity pair, there were four elements with a high value (0.75) plus four elements with a low value (0.25) in each stimulus; as in Bussey et al. (2003), the high-valued elements in the first stimulus served as low-valued elements in the second stimulus and vice versa, creating overlap. To simulate PRh lesions, we again removed Layer 3.

Results

Simulation data from the extended PMFC model are shown in Figure 8. In the low feature ambiguity condition, there was no significant main effect of Group, $F(1, 6) = 2.051, p = .202$, a significant effect of Block, $F(11, 66) = 11.18, p < .001$, and no Group \times Block interaction, $F(11, 66) = 0.729, p = .707$. In the high feature ambiguity condition,

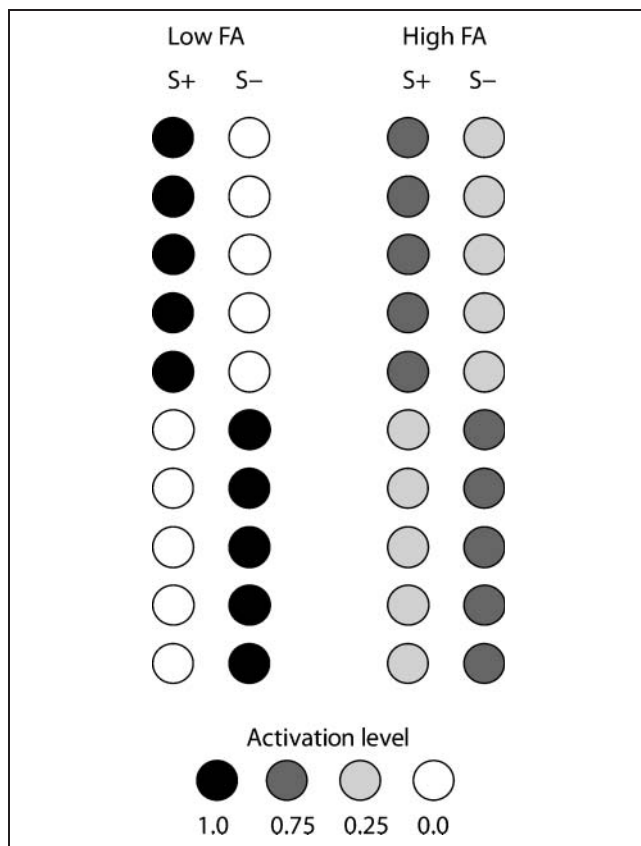


Figure 7. Schematic illustration of how stimulus inputs were “morphed” or blended by Bussey et al. (2003) and in Simulation 3 of the present article. The left panel shows a pair of stimuli with low feature ambiguity, in which all feature units are either fully activated (1.0) or fully off (0.0) and no features are common to the two stimuli. The right panel shows a pair of stimuli with high feature ambiguity, in which units that were fully activated in the low ambiguity pair have become less active (0.75) and units that were previously inactive have become partially active (0.25). The units in one stimulus that become partially active are the same units that are highly active in the other stimulus of the pair. This process parallels the morphing process used on photographs shown to monkeys in Bussey et al. (2003).

there was a significant effect of Group, $F(1, 6) = 32.74, p < .05$, a significant effect of Block, $F(11, 66) = 3.29, p < .01$, and a significant Group \times Block interaction, $F(11, 66) = 2.81, p < .01$.

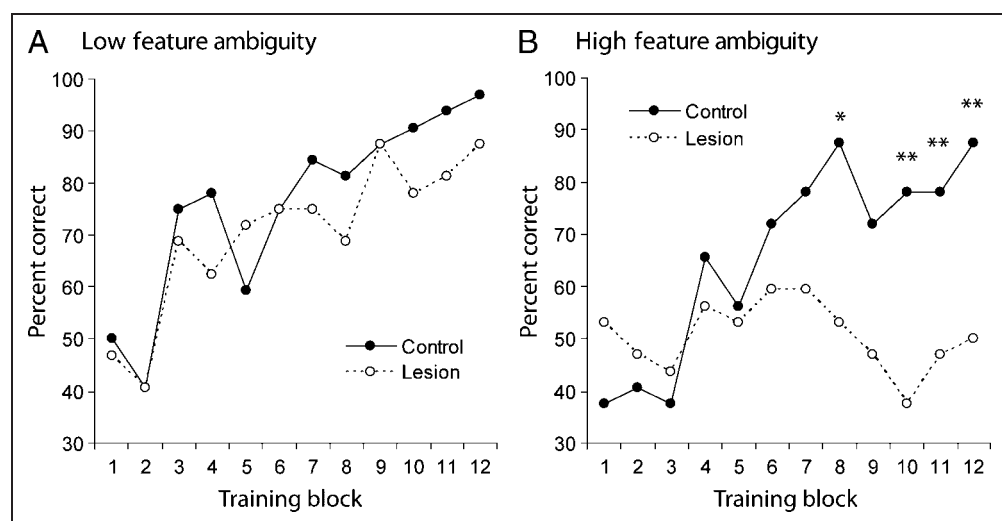
Discussion

As reported for both networks and monkeys in Bussey et al. (2003), introducing feature overlap between two complex stimuli in a visual discrimination caused impairments in networks lacking the most anterior layer, where the optimal representations of complex stimuli reside.

SIMULATION 4: SINGLE-PAIR DISCRIMINATIONS

Faithful replication of Iwai and Mishkin’s (1968) two tasks necessitated the use of different stimulus set sizes: a single pair of stimuli for pattern relearning and eight pairs in the concurrent discrimination task. Increasing set size is thought to increase the level of feature ambiguity between stimuli (Bussey & Saksida, 2002), which, in the case of the Concurrent Discrimination task, would create a greater need for whole-preferring conjunctive stimulus representations on Layer 3. Thus, our simulations of Iwai and Mishkin’s concurrent discrimination task demonstrate that dependence of performance on Layer 3 arises as a consequence of the cumulative feature ambiguity produced by a large set size. Would performance on the concurrent discrimination task also depend on Layer 3 if feature ambiguity were created not by cumulative feature ambiguity but by feature ambiguity between a single pair of complex stimuli? Our theory predicts that it would; we made that prediction computationally explicit in the present simulation.

Figure 8. Acquisition of discriminations between morphed stimuli by the extended PMFC model. Circles show percent correct in each successive block of eight training trials. (A) Low feature ambiguity condition. (B) High feature ambiguity condition. Asterisks indicate significant differences between the groups: $*p < .05$, $**p < .01$.



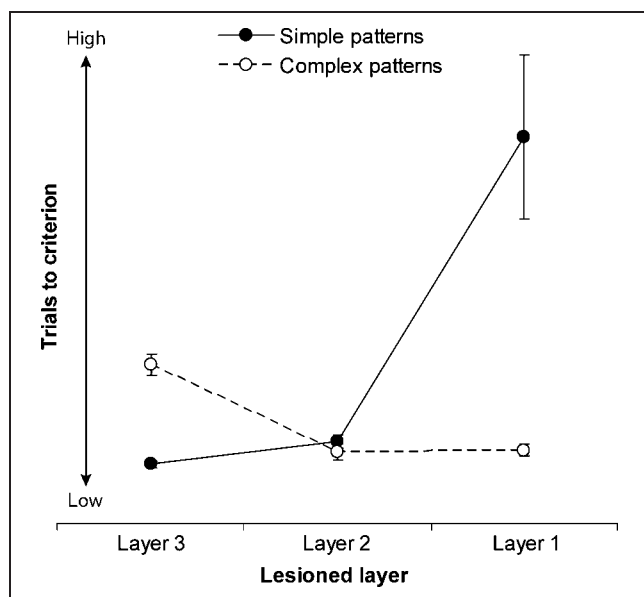


Figure 9. Performance of the extended PMFC model in simulating “simple” and “complex” single pair discriminations. Acquisition data for simple stimuli are shown with filled circles and solid lines; data for complex stimuli are shown with open circles and dotted lines.

Methods

Stimuli

We constructed a pair of complex stimuli that mirrored the single pair of simple stimuli used in the pattern relearning task, in that they possessed 50% ambiguous features. We simulated single pair discriminations with both simple and complex stimuli (of the form BC+ vs. CD– and BCDE+ vs. DEFG– for “simple” and “complex,” respectively). Both simple and complex tasks were learned by lesioned networks postoperatively, that is, we did not perform preoperative learning followed by postoperative reacquisition, as in pattern relearning above.

Training

Three groups of eight networks were initialized and trained as in the concurrent discrimination learning simulations described above, except that only one pair of stimuli was used for each of the two tasks, simple and complex. Networks were trained to a criterion of two consecutive blocks of 10 trials in which nine or more responses were correct.

Results

It can be seen from Figure 9 that the model predicts a Layer 3 lesion deficit for a single-pair discrimination between complex, feature-ambiguous stimuli, and a Layer 1 lesion deficit for a single-pair discrimination between simple, feature-ambiguous stimuli. A two-way ANOVA with

Task (simple, complex) as within-subjects factor and Lesion Group (posterior, middle, anterior) as between-subjects factor revealed a significant main effect of Lesion Group, $F(2, 21) = 11.19, p < .001$, a significant main effect of Task, $F(1, 21) = 7.24, p < .05$, and a significant Task \times Lesion Group interaction, $F(2, 21) = 19.68, p < .001$.

Discussion

This result shows that concurrent object discriminations are not necessary to reveal impairments with Layer 3 lesions. There is no fundamental difference between the way in which feature ambiguity in discriminations of complex stimuli induces dependence on Layer 3 and feature ambiguity in discriminations of simple stimuli induces dependence on Layer 1.

There are some extant experimental data that speak to these predictions. First, we can extrapolate from the pattern relearning task of Iwai and Mishkin (1968) that initial learning (i.e., without the *relearning* paradigm) of a discrimination between two simple patterns is likely to be impaired by posterior VVS lesions. Second, Bussey et al. (2003) trained monkeys with PRh lesions to discriminate a single pair of complex stimuli and found impairments.

To illustrate the extended PMFC model’s mechanism for single pair discriminations, Figure 10 shows the patterns of activation elicited by two simple stimuli across the three layers of the network. The patterns of activation elicited by two complex stimuli in the network are illustrated in Figure 11. In the simple single pair discrimination, at each

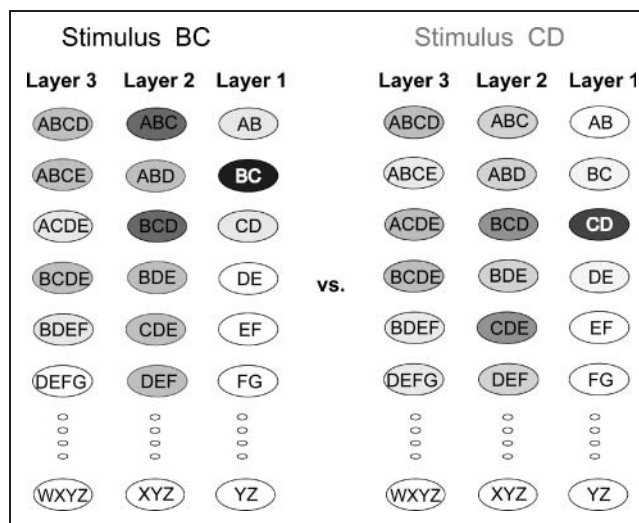


Figure 10. Schematic illustration of the representations in the model of stimuli used in the simple single pair discrimination task. Active units are shown in color; darker color indicates higher activity. At each of the three layers, the representations of BC and CD activate an overlapping set of units. However, in Layer 1, there is greater differentiation in activity amongst the units, yielding the least overlapping activation patterns: this layer is the most useful for solving the discrimination.

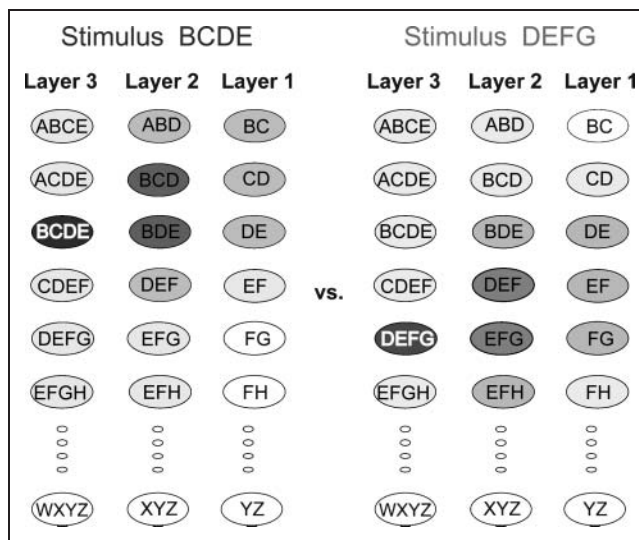


Figure 11. Schematic illustration of the representations of stimuli used in the complex single pair discrimination task. In this task, Layer 3 provides the least overlapping activation patterns and the best discrimination between the stimuli.

of the three layers, representations of BC and CD activate an overlapping set of units (Figure 10). Within each Layer 1 stimulus representation, the most active unit is relatively more active compared with the least active unit than is the case in Layer 2 or Layer 3 representations. This enhanced differentiation of units' activities in Layer 1 effectively reduces overlap between the two different stimulus representations for BC and CD. The enhanced differentiation of units' activities is due to inhibition of collateral units by the "winning unit," as described above. In Layers 2 and 3, several units share the status of "winner," leading to a more distributed pattern of activity for each stimulus, which results in poorer discrimination. Removal of Layer 1, therefore, has the most devastating effect on the discrimination of simple patterns.

In the complex single pair discrimination, the representations of BCDE and DEFG again activate an overlapping set of units (Figure 11). In this case, Layer 3 provides the most highly discriminable activation patterns representing stimuli BCDE and DEFG because Layer 3 contains the two units that represent the exact four-featured conjunctions BCDE and DEFG. Removing Layer 3, thus, has the most profound effect on the discrimination of patterns with four visual features. It is the same mechanism that accounts for data from the concurrent discrimination learning task of Iwai and Mishkin (1968).

SIMULATION 5: STIMULI OF INTERMEDIATE COMPLEXITY

In the present account, the degree of discrimination impairment caused by a lesion should reflect the extent to which the to-be-discriminated stimuli are "optimally" represented in the brain region that is lesioned. To complete

our demonstration of the model's predictions, we report a final simulation, in which the discriminanda are stimuli of complexity intermediate between those of the simple pattern and concurrent discrimination tasks of Iwai and Mishkin (1968). In the spectrum of complexity we have defined, these intermediate stimuli possess three features. In this discrimination, we expected to see the greatest impairment in networks that have lesions in the middle layer—Layer 2.

Methods

Stimuli

We constructed eight pairs of complex stimuli in the same way that stimuli were constructed for the concurrent discrimination task, except that we activated 3 out of a possible 32 units in the input layer, rather than 4, for each stimulus.

Training

Three groups of eight networks were initialized and trained as in the concurrent discrimination learning simulations described above. Networks were trained to a criterion of two consecutive blocks of 10 trials in which nine or more responses were correct.

Results

It can be seen from Figure 12 that the model predicts a Layer 2 lesion deficit for a concurrent discrimination between eight pairs of stimuli of intermediate complexity. A one-way ANOVA with Lesion Group (posterior, middle, anterior) as between-subjects factor revealed a significant

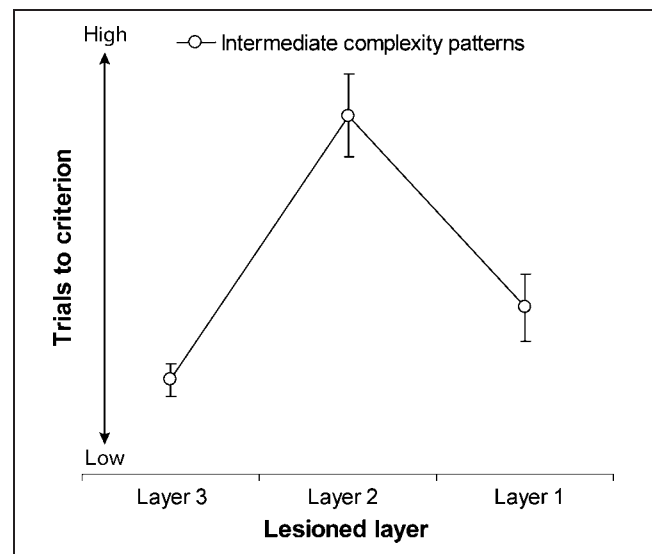


Figure 12. Performance of the extended PMFC model on concurrently learning to discriminate eight pairs of stimuli of intermediate level complexity. Error bars indicate $\pm SEM$.

effect of Lesion Group, $F(2, 20) = 18.00, p < .001$. Post hoc comparisons of group differences, corrected by Sidak adjustment, revealed that the group middle differed from the group posterior ($p < .005$) and from the group anterior ($p < .001$) but that there was no significant difference between the group anterior and the group posterior ($p = .30$).

Discussion

This simulation result gives a concrete demonstration of a central aspect of our theoretical account; namely, that a lesion at any point in the VVS will cause impairments in visual discrimination learning, if the to-be-discriminated stimuli possess a level of complexity best represented by the neurons in the lesioned area. Here, the stimuli possessed an intermediate level of complexity and it was the middle group—in which Layer 2 of the networks had been removed—that showed the greatest degree of impairment in contrast to the impairment of the posterior group in the case of simple pattern discrimination and the impairment of the anterior group in the concurrent discrimination of complex objects.

GENERAL DISCUSSION

Our simulations demonstrate that dissociable effects of VVS lesions on visual discrimination learning may reflect compromised stimulus representations within a single system rather than separate modules for perception and memory.

Accounting for Iwai and Mishkin (1968)

We assumed that the visual discriminations used by Iwai and Mishkin (1968), like any visual discriminations, possess feature ambiguity. This property of the stimuli may occur regardless of the level of complexity of those visual stimuli. The discrimination of stimuli that possess ambiguous features is facilitated by conjunctive representations of those stimuli, for which the whole stimulus is greater (more “activating”) than the sum of the parts. Such “whole-preferring” conjunctive representations exist for a given stimulus only at the point in the VVS where the level of complexity of representations matches that of the stimulus. Thus, any discrimination problem is solved most effectively by a particular “optimal” stage in the VVS.

In the pattern relearning simulation, the two simple patterns used by Iwai and Mishkin (1968) were represented by just two features each, AB and BC. The two stimulus representations in Layer 1 are more discriminable than those in Layers 2 and 3 because lateral inhibition greatly enhances the activity of the conjunctive units exactly matching the stimuli AB and BC. Removing Layer 1, therefore, causes the greatest impairments in discrimination. In the concurrent discrimination task, representations of similar, complex stimuli are most discriminable in Layer 3 because

lateral inhibition selectively enhances the activation of the units exactly representing the conjunction of the four features present in each stimulus. So removal of Layer 3 causes the greatest impairments. Exactly the same mechanism accounts for the results of the pattern relearning and concurrent discrimination tasks; only the level of stimulus complexity is different.

We note that the trials to criterion measure for networks with a lesion of Layer 2 is not significantly different from that of networks with a lesion of Layer 1 in the concurrent discrimination task or from that of networks without Layer 3 in the simple pattern task. Monkeys in the Iwai and Mishkin (1968) experiment exhibited a more gradual increase in impairment with position of lesion, at least in the pattern relearning task. The less graded performance seen in the model than in lesioned monkeys can be explained in terms of a certain simplification that we have assumed. In the model, each layer contains units of only one level of complexity: in Layer 1, all units combined exactly two features into a conjunction; in Layer 2, all units combined exactly three features into a conjunction, and so on. In cortex, the gradation in complexity of neural representations along VVS is surely much less discrete than in our simple model, that is, each “layer” in VVS is somewhat heterogeneous in terms of the level of complexity to which the individual neurons respond. In our model, all units in Layer 2 respond maximally and most selectively to a stimulus with exactly three features; in the brain, at the point corresponding to Layer 2, most neural representations will be of the level of complexity corresponding to three features, but a few will be at the level of two features and some at the level of four features. Hence, the function of those brain regions emerges as continuous under the present account.

Feature Ambiguity at Different Levels of Stimulus Complexity

The extended PMFC model predicts that if the feature ambiguous discrimination experiment of Bussey et al. (2002) were replicated using simple stimuli, such as lines of varying orientation, monkeys with posterior VVS lesions would be more impaired than monkeys with PRh lesions. This prediction follows from a key assumption of the model: Resolution of feature ambiguity is performed at all stages of the VVS, not just at the anterior end. Whether a region of VVS is critical for a discrimination depends on whether the feature ambiguity present in the discrimination is best resolved in the region of brain damaged by the lesion.

Rosenthal and Behrmann (2006) presented some data from human subjects that speak to the prediction for monkeys described in the previous paragraph. They tested a patient, J.W.—who has an extensive lesion of V2 in both hemispheres and no evidence of damage at higher level areas of the ventral stream—on a visual discrimination learning task with simple stimuli. J.W. and age-matched controls were required to learn to discriminate between

three classes of visual stimulus, with feedback. All stimuli were composed of a pair of white stripes on a gray background, and the only difference between the three categories was the width of the white stripes: in category A, both stripes were narrow; in category B, both stripes were of intermediate width; and in category C, both stripes were wide. Thus, the stimuli across the three categories shared many perceptual properties, making their category membership highly ambiguous. J.W. was impaired in the acquisition of this discrimination, requiring many sessions of learning to acquire good discrimination performance, while controls acquired the discrimination in the first training session. Therefore, in human subjects, damage to early stations in the VVS is sufficient to cause impairments in the classification of simple visual stimuli, just as observed in nonhuman primates by Iwai and Mishkin (1968).

Representational Requirements

A prediction of our framework for understanding lesion effects is that the specificity of representations required for solving a task is determined by the representational requirements of the task (Tyler et al., 2004). According to our view, representational requirements are affected by at least two factors: the properties of the stimulus material and the task instructions. The first factor—the properties of the stimulus material—influences representational requirements by determining the degree of detail in which stimuli must be represented to resolve the task. For example, if the stimuli are complex but possess no ambiguous features, even posterior stages will be able to discriminate them using representations of simple features. The anterior stage that represents these stimuli conjunctively only becomes critical for the discrimination if the stimuli possess ambiguous features. The second factor—the task instructions—determines the representational requirements of the task in a similar way. Perhaps the way in which a stimulus is represented is not consistent across different tasks because the representation constructed need only be as detailed as the task solution demands (Palmeri & Tarr, 2008). For example, subjects might be required to discriminate between stimuli in one task and to categorize the same stimuli on the basis of a simple visual property in another task. Discriminating between complex, feature-ambiguous stimuli requires conjunctive representations of anterior-level complexity, but a categorization task using the same stimuli, in which categories are defined by one visual feature, would require only simple (posterior) representations of the features of those complex stimuli. This analysis provides a very different account of categorization to that provided by the traditional, modular view (Knowlton & Squire, 1993).

Functional Dissociations

The present simulations demonstrate a double dissociation, which is often interpreted as evidence for two mod-

ules. Yet we claim that our account of object processing is a continuous, hierarchical one. The present study shows how our claim can be reconciled with the data, by demonstrating that a double dissociation does not necessarily imply modularity at the cognitive level, because the simulation results emerge from a mechanism that eschews cognitive modules. Our model is, of course, a simplification of the brain: We have instantiated successive regions of VVS in only three layers. To better conceptualize how the function of this system could be truly continuous, one can imagine a model with ten layers or one hundred. With one hundred layers, we could take pairs of layers chosen at random, and the closer those two layers happen to lie in the continuum, the more similar would be their contribution to discriminating stimuli of a given complexity. In the limit of a very large number of layers (as in the brain), the change in function across the layers would be revealed as continuous. The double dissociation revealed in the simulations arises through the comparison of two tasks that use stimuli lying at opposite ends of the spectrum of stimulus complexity (or, at least, the spectrum represented in our model). There is no difference in processing mechanism at the two points in the model at which lesions produce opposite behavioral effects; the only difference is the complexity of stimulus representations residing at those points. Similarly, Plaut (1995) has presented a connectionist account of a double dissociation between concrete and abstract word reading in dyslexia. The double dissociation produced by damaging the network is a consequence of a functional specialization that arises in the model, without an assumption of modularity. As Shallice (1988) remarked, “If modules exist, then double dissociations can reveal them. However, finding double dissociations is no guarantee that modules exist.”

One prediction of this framework is that if discrimination performance were tested on a range of stimuli whose complexity increased incrementally, the effect of a lesion in anterior VVS would also increase incrementally, and the effect of a lesion in posterior VVS would decrease incrementally. A related prediction is that object processing in the healthy brain should differentially engage different regions of the VVS, depending on the level of detailed information that must be extracted from an object. A neuroimaging study by Tyler et al. (2004) has tested this hypothesis. Healthy subjects were shown pictures of everyday objects and asked to name them either at a “basic” level (e.g., as a donkey or hammer) or at a “domain” level (e.g., as a living thing or manmade object). When the task required detailed information to differentiate between objects (naming at a basic level), anteromedial regions of the left temporal lobe were recruited, and when the task did not demand such detailed information (naming at a domain level), activation was limited to bilateral posterior regions of IT. Tyler et al. also tested two patients with bilateral anterior temporal lobe damage on a picture-naming study comparable to that used in healthy subjects. The patients were impaired on naming at a basic level but very accurate at domain level naming.

An Alternative to Cognitive Modularity

The account that we present represents a completely different paradigm from the prevailing view in the field, in which the brain is divided into discontinuous modules, each defined according to a particular psychological function. The two accounts differ in at least two important respects.

First, our model is a single system in which the computational processes at each stage in the hierarchy are the same. We were able to reproduce the double dissociation of Iwai and Mishkin (1968) without any differences in cognitive processing across brain regions. Therefore, our results challenge the interpretation of Iwai and Mishkin's observations in terms of cognitive functional modules. The commonly encountered assumption that double dissociations such as Iwai and Mishkin's force an interpretation in terms of psychological modules cannot be correct. The success of our simulations suggests that the labels "perception" and "memory" may be the wrong way to describe cognition in VVS because these labels imply a host of functional and neural-algorithmical differences between the different regions of the pathway. Rather, differences in the cognitive role of different brain regions may be due to differences in the representations that those brain regions contain.

Second, the modular view posits modules that are involved in a single psychological function: A region involved in, say, memory performs that function to the exclusion of any role in perception. In this view, the brain adheres to the rule, "one region, one function." In contrast, our account claims that different regions contain different stimulus representations but that each region can, in principle, be involved in a range of cognitive functions. It may be that some representations (the simpler ones, in posterior regions) tend to be more useful for traditional "perceptual" tasks, whereas other representations (the complex ones, in anterior regions) are often more useful for "memory" tasks. When representational requirements and task labels are unconfounded by careful manipulation of stimulus material, it has been revealed that anterior regions previously thought to mediate only memory are also involved in perception (Lee, Scabill, & Graham, 2008; Bartko, Winters, Cowell, Saksida, & Bussey, 2007). Similarly, López-Aranda et al. (2009) recently demonstrated that cells in Layer 6 of the visual cortical area V2 are involved in object recognition memory, contradicting the notion that regions in posterior VVS are involved in only visual perception.

In line with the model's suggestion that visual perception and visual memory are distributed across common brain regions, several authors have rejected the modular approach to higher order perception and memory. Gaffan (1996, 2002) has argued strongly against using the concept of a module for memory or "memory system" in the brain. Bussey (2004) also questions the utility of the modular, "multiple memory systems" framework. Palmeri and Tarr (2008) join these authors in outlining the weaknesses of a multiple memory systems approach. They note that multiple memory systems theories typically do not answer

specific questions about underlying cognitive mechanisms such as how memories are encoded, represented, and processed. Further, Palmeri and Tarr point out that this weakness of the modular approach to memory is paralleled in modular accounts of visual perception that assign independent systems to particular kind of objects. Finally, Saksida (2009) has argued that the VVS and MTL should not be segregated according to their putative roles in perception and memory, rather the entire pathway is important for both of these cognitive functions.

In addition, there is a movement in the human cognitive literature advocating an account of cognition in terms of distributed cortical function. Some have proposed a process-driven account of cognition (e.g., Kolers & Roediger, 1984) or a levels-of-processing framework (Craik & Lockhart, 1972), in which memory is conceived of as a by-product of perception. More recently, Goldstone and Barsalou (1998) have suggested that higher order conceptual processes are grounded in perception. Fuster (2003) referred to the alternative, unitary account as a "network model of cognition"; Foster and Jelicic (1999) described it as a "processing" view.

A Continuous Hierarchical Account of Object Processing

Thus, the case against a modular organization of perception and memory is gaining momentum. The model we present in this article is one example of the general approach, which ties in the levels-of-processing view of perception and memory advocated by cognitive psychologists (e.g., Craik & Lockhart, 1972) with neurobiological properties of the brain.

Several strands of research are now converging to provide support for a continuous, hierarchical account of object processing. In addition to the anatomical and neurophysiological evidence, there is a growing body of theoretical work in favor of the new view. For example, Ullman et al. (2002) performed a computational analysis of visual images that measured the amount of information delivered by particular features about the class membership of the images. The features they sampled were subregions of the full images at a range of sizes and resolutions, and the features that emerged as most informative for classification were those of intermediate level complexity. The proposed reason for the superiority of intermediate complexity features over both smaller, simpler components and larger, more complex fragments is that intermediate features strike the optimal balance between specificity and frequency of occurrence. That is, for classification at the basic level, small, simple features occur so frequently in visual objects that they appear in many classes of object and cannot be used for discriminating between classes, but larger, complex features are so specific to one instance of a visual object that they cannot be used to generalize to different instances of the same class. The authors speculated that the features their information-theoretic model showed

to be optimal might map onto the features preferred by neurons in IT. In addition, Ullman et al. (2002) suggested that in tasks requiring the identification of specific objects, the optimal features would instead be global object views, that is, more complex features. Indeed, Zhang and Cottrell (2005) tested this hypothesis by extending the method of Ullman et al. (2002) to look for the features that are most useful in a subordinate level classification task, namely, face identification. They found that large areas of faces constituted the most informative fragments for face identification, confirming that the optimal representations for a task that demands detailed discriminations between similar stimuli are those with a high level of complexity.

Conclusions

We have presented an extension of the PMFC model to investigate the effect of lesions in VVS. Our single-system account can explain data from lesion studies including those of the canonical study of Iwai and Mishkin (1968), in which lesions in posterior versus anterior regions of VVS differentially affected performance on two visual discrimination learning tasks. We ask whether observed dissociations between memory and perception often arise because putative memory tasks typically require complex representations found in more anterior areas of VVS, whereas tasks presumed to tap perception require simpler representations located in posterior regions? Perhaps, in general, double dissociations are best interpreted not as an indication of functional modularity but as a demonstration of differences in the level of processing carried out by different points in a representational hierarchy. The present results could have far-reaching implications for categorization, perceptual learning, and recognition memory: Patterns of impairment in these functions following brain damage may reflect the demand that each task places on stimulus representations rather than dissociable cognitive modules.

APPENDIX

Network Overview

The network consists of an input layer, three stimulus representation layers (Layer 1, Layer 2, and Layer 3), and a single outcome node whose activation represents an event, for example, reward. Weights on the links between the input layer and Layers 1 to 3 are fixed. Upon presentation of an input stimulus, units in Layers 1, 2, and 3 are activated according to the similarity of the input pattern to the conjunction of features the unit represents. All units in Layers 1, 2, and 3 are linked to the outcome node with weights that are adjustable by an associative learning rule. Weights are adjusted every time the network makes a “choice” response between a pair of stimuli; the size of the adjustment is proportional to the activation of the unit and the discrepancy between

the response and the outcome. Each layer computes a response to a stimulus, which is the sum of the units’ activations multiplied by their associative weight strengths.

Initialization

Weights on the links between the input layer and Layers 1, 2, and 3 are set on initialization of the network. Each Layer 1 unit is connected to two input units, each Layer 2 unit is connected to three input units, and each Layer 3 unit is connected to four input units. Weight values are chosen so that the total weight value converging on any unit in Layers 1, 2, and 3 is 1. Thus, weights connecting an input unit to a Layer 1 unit are set to a value of 0.5, weights connecting an input unit to a Layer 2 unit are set to 0.33, and weights connecting an input unit to a Layer 3 unit are set to 0.25. It is assumed that a subset of all possible combinations of visual features are already “known” to the network. Layer 1 contains all possible combinations of two visual features into a conjunction, giving 496 units, Layer 2 also contains 496 units, that is, a randomly selected 10% subset of all possible three-feature combinations, and Layer 3 contains 496 units, each representing the conjunction of four visual features, which corresponds to 1.38% of all possible four-feature conjunctions. In addition, extra units in Layers 2 and 3 can be recruited as required on presentation of stimuli that contain conjunctions not already existent in the network.

Calculating the Activation of a Unit in Layers 1, 2, and 3

A stimulus S is represented on the input layer as a vector of activations, each element taking a value between 0 and 1. An activation value of 1 for an element represents the presence of the “visual feature” corresponding to that element in the vector, and an activation value of 0 represents the absence of that feature. Activations that are intermediate between 0 and 1 represent features that are only weakly present.

Because the calculation of activation is the same for units in Layers 1, 2, and 3, we refer to the layers collectively as Layer N . Every Layer N unit, indexed by j , is connected to certain units in the input layer indexed by i , via a link with weight w_{ji} . For every unit j in Layer N , there exists a set of weighted links \mathbf{w}_j (a vector of weight values) between each unit i in the input layer that is connected to that Layer N unit. The stimulus vector is mapped onto Layer N units according to Equation A1.

$$a_j = \sum_i a_i w_{ji}, \quad (\text{A1})$$

where a_j is the activation of the Layer N unit, a_i is the activation of the input layer unit, and i refers to the set of all input units to which the Layer N unit is connected.

When a stimulus is presented to the network, the unit or units in Layer N with the highest activation value are chosen as the “winning unit” or “winning units.” A lateral inhibition function is then applied to the activations. The function depends on the number of winners in the layer and the activation level of that winner according to the following equation:

$$\hat{a}_j = \exp(-Q(1 - a_j)), \quad (\text{A2})$$

where

$$Q = q \left(\frac{n_N}{\text{winners}_N} \right) a_j^* \quad (\text{A3})$$

In which q is a constant, n_N is the number of units in Layer N , winners_N is the number of units in Layer N with the maximum activation level, and a_j^* is the maximum activation level in Layer N . Equations A2 and A3 ensure that the lateral inhibition function is the steepest when there are few winners and the activation level of those winners is high, which corresponds to the likely outcome of lateral competition between collateral units.

The effect of this dampening function is that any unit in any layer of the network strongly prefers an exact match between the input pattern and the precise conjunction of features that the unit represents; that is, “the whole is more than the sum of the parts.” If there is more than one winning unit in Layer N , for example, when the input stimulus is not of the level of complexity best represented by that layer, the lateral inhibition is much weaker, according to Equations A2 and A3. This reflects the lack of response selectivity found in IT neurons for stimuli that are not of the preferred level of stimulus complexity for neurons in that region of IT (Kobatake & Tanaka, 1994).

The activation value, a_j , of each unit in Layer N produced by Equation A2 is then normalized according to Equation A4.

$$b_j = \frac{\hat{a}_j}{\sqrt{\sum_{j=1}^n \hat{a}_j^2}}, \quad (\text{A4})$$

where b_j is the normalized activation of Layer N unit j , and n is the number of units in Layer N .

Associative Learning

After the pattern of activation due to a stimulus has been determined in all layers, the weights on the links between units in Layer N and the outcome node are adjusted to an extent dependent on the activation of the sending unit (described subsequently). Subsequent presentations of the stimulus will activate the outcome node and lead to performance of a conditioned response or CR.

The tasks simulated in the present article are from a simultaneous visual discrimination learning paradigm in which an animal would be required to choose one of two simultaneously presented stimuli. In these simulations, the stimuli are presented to the network one at a time, so a “choice” is simulated by first presenting stimulus A and calculating a CR value, then presenting stimulus B and calculating the corresponding CR elicited, then comparing the two CR values in a probabilistic fashion (described below) to determine the network’s response.

The total CR output of the network for a given pattern is the sum of the independent CR contributions from each of Layers 1, 2, and 3. The CR value for Layer N , CR_N , is the sum over all units of the product of the activation of each unit j and its weight to the outcome node w_{oj} (Equation A5).

$$\text{CR}_N = \sum_{j=1}^n w_{oj} * b_j, \quad (\text{A5})$$

where b_j is the normalized activation value of Layer N unit j , w_{oj} is the weight from unit j to the outcome node, and n is the number of units in Layer N . The total CR for a pattern, CR_{tot} , is the sum of the three CR_N values from each of Layers 1, 2, and 3.

The associative weights for each Layer N are updated according to the delta rule (Rescorla & Wagner, 1972) in proportion to the activation of the corresponding Layer N unit (Equation A6).

$$\Delta w_{oj} = \alpha b_j \left(\frac{\lambda}{3} - \frac{\text{CR}_{\text{tot}}}{3} \right), \quad (\text{A6})$$

in which α is a constant learning rate, and λ is the value of the reinforcer. In simulations in which a layer is lesioned, so that only two layers remain, both λ and CR_{tot} are divided by 2 rather than 3.

Behavioral Output

The output of the model is the CR to each of the presented stimuli. The behavioral response—the selection of one stimulus from the pair—is a stochastic choice that depends on the magnitudes of $\text{CR}(\text{A})$ and $\text{CR}(\text{B})$, elicited by paired patterns A and B, respectively. To choose a response, a random number between 0 and 1 is generated. If learning is not yet at asymptote and the random number is greater than the sum of $\text{CR}(\text{A})$ and $\text{CR}(\text{B})$, a choice is selected randomly. If $\text{CR}(\text{A})$ is greater than the random number, action A is chosen. If $\text{CR}(\text{A})$ is less than the random number, action B is chosen. Thus, as $\text{CR}(\text{A})$ increases, so does the likelihood that $\text{CR}(\text{A})$ is greater than the random number and that a stimulus A is chosen. In addition, if the sum of $\text{CR}(\text{A})$ and $\text{CR}(\text{B})$ is greater than 1, the random number is multiplied by the sum of the two CRs before comparison with $\text{CR}(\text{A})$ is made, to scale up

the comparison number appropriately. This avoids a bias to choose CR(A) in highly trained networks where both CR(A) and CR(B) are high.

Parameters

Input stimuli consisted of a vector of 32 elements—each element corresponding to a visual feature—in which a number of features were “present” (taking activation level 1) and all other possible features in the input layer were “absent” (activation 0). Simple stimuli contained two active input elements, complex stimuli contained four active input elements (except in Simulation 3 with “morphed” stimuli; see Methods), and intermediate stimuli contained three active input elements. Identical network parameters were used for all simulations presented in this article: $\lambda = 1.0$, $\alpha = .0324$, $q = .02$.

Reprint requests should be sent to Rosemary A. Cowell, Psychology Department, University of California, San Diego, 9500 Gilman Drive #0109, La Jolla, CA 92093-0109, or via e-mail: rosie.cowell@gmail.com.

REFERENCES

- Baker, C. I., Behrmann, M., & Olson, C. R. (2002). Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. *Nature Neuroscience*, 5, 1210–1216.
- Barese, M. D., Gaffan, D., & Graham, K. S. (2007). The human medial-temporal lobe processes on-line representations of complex objects. *Neuropsychologia*, 45, 2963–2974.
- Bartko, S. J., Winters, B. D., Cowell, R. A., Saksida, L. M., & Bussey, T. J. (2007). Perirhinal cortex resolves feature ambiguity in configural object recognition and perceptual oddity tasks. *Learning and Memory*, 14, 821–832.
- Baxter, M. G. (2009). Involvement of medial-temporal lobe structures in memory and perception. *Neuron*, 61, 667–677.
- Blake, L., Jarvis, C. D., & Mishkin, M. (1977). Pattern discrimination thresholds after partial inferior temporal of lateral striate lesions in monkeys. *Brain Research*, 120, 209–220.
- 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. *Learning and Memory*, 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. *Learning and Memory*, 7, 375–382.
- Bussey, T. J. (2004). Multiple memory systems: Fact or fiction? *Quarterly Journal of Experimental Psychology*, 57B, 89–94.
- Bussey, T. J., & Saksida, L. M. (2002). The organization of visual object representations: A connectionist model of effects of lesions in perirhinal cortex. *European Journal of Neuroscience*, 15, 355–364.
- Bussey, T. J., & Saksida, L. M. (2007). Memory, perception, and the ventral visual-perirhinal-hippocampal stream: Thinking outside of the boxes. *Hippocampus*, 17, 898–908.
- Bussey, T. J., Saksida, L. M., & Murray, E. A. (2002). Perirhinal cortex resolves feature ambiguity in complex visual discriminations. *European Journal of Neuroscience*, 15, 365–374.
- Bussey, T. J., Saksida, L. M., & Murray, E. A. (2003). Impairments in visual discrimination after perirhinal cortex lesions: Testing “declarative” versus “perceptual-mnemonic” views of perirhinal cortex function. *European Journal of Neuroscience*, 17, 649–660.
- Butter, C. M. (1972). Detection of masked patterns in monkeys with inferotemporal, striate or dorsolateral frontal lesions. *Neuropsychologia*, 10, 241–243.
- Cowell, R. A., Bussey, T. J., & Saksida, L. M. (2006). Why does brain damage impair memory? A connectionist model of object recognition memory in perirhinal cortex. *Journal of Neuroscience*, 26, 12186–12197.
- Cowey, A., & Gross, C. G. (1970). Effects of foveal prestriate and inferotemporal lesions on visual discrimination by rhesus monkeys. *Experimental Brain Research*, 11, 128–144.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 671–684.
- Dean, P. (1974). Choice reaction times for pattern discriminations in monkeys with inferotemporal lesions. *Neuropsychologia*, 12, 465–476.
- Desimone, R., Albright, T. D., Gross, C. G., & Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *Journal of Neuroscience*, 4, 2051–2062.
- Desimone, R., & Schein, S. J. (1987). Visual properties of neurons in area V4 of the macaque: Sensitivity to stimulus form. *Journal of Neurophysiology*, 57, 835–868.
- Desimone, R., & Ungerleider, L. G. (1989). Neural mechanisms of visual processing in monkeys. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (Vol. 2, pp. 267–299). New York: Elsevier Science.
- Devlin, J. T., & Price, C. J. (2007). Perirhinal contributions to human visual perception. *Current Biology*, 17, 1484–1488.
- Eacott, M. J., Gaffan, D., & Murray, E. A. (1994). Preserved recognition memory for small sets, and impaired stimulus identification for large sets, following rhinal cortex ablations in monkeys. *European Journal of Neuroscience*, 6, 1466–1478.
- Eysel, U. T., Worgotter, F., & Pape, H. C. (1987). Local cortical lesions abolish lateral inhibition at direction selective cells in cat visual cortex. *Experimental Brain Research*, 68, 606–612.
- Foster, J. K., & Jelicic, M. (1999). *Memory: Systems, process, or function?* Oxford: Oxford University Press.
- Fukushima, K. (1980). Neocognitron: A self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36, 193–202.
- Fuster, J. M. (2003). *Cortex and mind: Unifying cognition*. New York: Oxford University Press.
- Gaffan, D. (1996). Associative and perceptual learning and the concept of memory systems. *Cognitive Brain Research*, 5, 69–80.
- Gaffan, D. (2002). Against memory systems. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, 357, 1111–1121.
- Gaffan, D., Harrison, S., & Gaffan, E. A. (1986). Visual identification following inferotemporal ablation in the monkey. *Quarterly Journal of Experimental Psychology: B*, 38, 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. *Behavioral Neuroscience*, 106, 30–38.
- Gilbert, P. E., & Kesner, R. P. (2003). Recognition memory for complex visual discriminations is influenced by stimulus interference in rodents with perirhinal cortex damage. *Learning and Memory*, 10, 525–530.

- Gluck, M. A., & Bower, G. H. (1988). Evaluating an adaptive network model of human learning. *Journal of Memory and Language*, 27, 166–195.
- Goldstone, R. L., & Barsalou, L. W. (1998). Reuniting perception and conception. *Cognition*, 65, 231–262.
- Gross, C. G., Cowey, A., & Manning, F. J. (1971). Further analysis of visual discrimination deficits following foveal prestriate and inferotemporal lesions in rhesus monkeys. *Journal of Comparative & Physiological Psychology*, 76, 1–7.
- Hampton, R. R. (2005). Monkey perirhinal cortex is critical for visual memory, but not for visual perception: Re-examination of the behavioural evidence from monkeys. *Quarterly Journal of Experimental Psychology*, 58B, 283–299.
- Hartley, T., Bird, C. M., Chan, D., Cipolotti, L., Husain, M., Vargha-Khadem, F., et al. (2007). The hippocampus is required for short-term topographical memory in humans. *Hippocampus*, 17, 34–48.
- Hubel, D., & Wiesel, T. (1962). Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *Journal of Physiology*, 160, 106–154.
- Hubel, D., & Wiesel, T. (1965). Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *Journal of Neurophysiology*, 28, 229–289.
- Iversen, S. D., & Humphrey, N. K. (1971). Ventral temporal lobe lesions and visual oddity performance. *Brain Research*, 30, 253–263.
- Iwai, E., & Mishkin, M. (1968). Two visual foci in the temporal lobe of monkeys. In N. Yoshii & N. Buchwald (Eds.), *Neurophysiological basis of learning and behavior* (pp. 1–11). Japan: Osaka University Press.
- Kikuchi, R., & Iwai, E. (1980). The locus of the posterior subdivision of the inferotemporal visual learning area in the monkey. *Brain Research*, 198, 347–360.
- Knowlton, B. J., & Squire, L. R. (1993). The learning of categories: Parallel brain systems for item memory and category knowledge. *Science*, 262, 1747–1749.
- Kolers, P. A., & Roediger, H. L. (1984). Procedures of mind. *Journal of Verbal Learning and Verbal Behavior*, 23, 425–429.
- Lee, A., Bandelow, S., Schwarzbauer, C., Henson, R. N., & Graham, K. S. (2006). Perirhinal cortex activity during visual object discrimination: An event-related fMRI study. *Neuroimage*, 33, 362–373.
- Lee, A., Scahill, V., & Graham, K. S. (2008). Activating the medial temporal lobe during oddity judgment for faces and scenes. *Cerebral Cortex*, 18, 683–696.
- Lee, A. C., Buckley, M. J., Gaffan, D., Emery, T., Hodges, J. R., & Graham, K. S. (2006). Differentiating the roles of the hippocampus and perirhinal cortex in processes beyond long-term declarative memory: A double dissociation in dementia. *Journal of Neuroscience*, 26, 5198–5203.
- Lennie, P. (1998). Single units and visual cortical organisation. *Perception*, 27, 885–1010.
- Levy, D. A., Shrager, Y., & Squire, L. R. (2005). Intact visual discrimination of complex and feature-ambiguous stimuli in the absence of perirhinal cortex. *Learning and Memory*, 12, 61–66.
- López-Andrade, M. F., López-Téllez, J. F., Navarro-Lobato, I., Masmudi-Martín, M., Gutiérrez, A., & Khan, Z. U. (2009). Role of layer 6 of V2 visual cortex in object-recognition memory. *Science*, 325, 87–89.
- Manning, F. J. (1971a). Punishment for errors and visual-discrimination learning by monkeys with inferotemporal cortex lesions. *Journal of Comparative and Physiological Psychology*, 75, 146–152.
- Manning, F. J. (1971b). The selective attention “deficit” of monkeys with ablations of foveal prestriate cortex. *Psychonomic Science*, 25, 291–292.
- 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. *Journal of Neuroscience*, 13, 5418–5432.
- Mishkin, M. (1982). A memory system in the monkey. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, 298, 83–95.
- Moss, H. E., Rodd, J. M., Stamatakis, E. A., Bright, P., & Tyler, L. K. (2005). Anteromedial temporal cortex supports fine-grained differentiation among objects. *Cerebral Cortex*, 15, 616–627.
- Murray, E. A., & Bussey, T. J. (1999). Perceptual-mnemonic functions of perirhinal cortex. *Trends in Cognitive Sciences*, 3, 142–151.
- Nakamura, H., Gattass, R., Desimone, R., & Ungerleider, L. G. (1993). The modular organization of projections from areas V1 and V2 to areas V4 and TEO in macaques. *Journal of Neuroscience*, 13, 3681–3691.
- Norman, G., & Eacott, M. J. (2004). Impaired object recognition with increasing levels of feature ambiguity in rats with perirhinal cortex lesions. *Behavioural Brain Research*, 148, 79–91.
- Palmeri, T. J., & Gauthier, I. (2004). Visual object understanding. *Nature Reviews Neuroscience*, 5, 291–304.
- Palmeri, T. J., & Tarr, M. (2008). Object recognition and long-term visual memory for objects. In S. Luck & A. Hollingsworth (Eds.), *Visual memory*. Oxford: Oxford University Press.
- Perrett, D., & Oram, M. (1993). Neurophysiology of shape processing. *Image and Vision Computing*, 11, 317–333.
- Plaut, D. C. (1995). Double dissociation without modularity: Evidence from connectionist neuropsychology. *Journal of Clinical and Experimental Neuropsychology*, 17, 291–321.
- Preston, A. R., & Gabrieli, J. D. (2008). Dissociation between explicit memory and configural memory in the human medial temporal lobe. *Cerebral Cortex*, 18, 2192–2207.
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton Century Crofts.
- Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2, 1019–1025.
- Riesenhuber, M., & Poggio, T. (2000). Models of Object Recognition. *Nature Neuroscience*, 3, 1199–1204.
- Rosenthal, O., & Behrmann, M. (2006). Acquiring long-term representations of visual classes following extensive extrastriate damage. *Neuropsychologia*, 44, 799–815.
- Sakai, S., & Miyashita, Y. (1993). Memory and imagery in the temporal lobe. *Current Opinion in Neurobiology*, 3, 166–170.
- Saksida, L. M. (2009). Neuroscience. Remembering outside the box. *Science*, 325, 40–41.
- Saleem, K., Tanaka, K., & Rockland, K. S. (1992). PHA-L study of connections from TEO and V4 to TE in the monkey visual cortex. *Society for Neuroscience Abstracts*, 18294.
- Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery and Psychiatry*, 20, 11–21.
- Shallice, T. (1988). *From neuropsychology to mental structure*. Cambridge: Cambridge University Press.
- Shrager, Y., Gold, J. J., Hopkins, R. O., & Squire, L. R. (2006). Intact visual perception in memory-impaired patients with medial temporal lobe lesions. *Journal of Neuroscience*, 26, 2235–2240.
- Sillito, A. M. (1979). Inhibitory mechanisms influencing complex cell orientation selectivity and their modification at high resting discharge levels. *Journal of Physiology*, 289, 33–53.

- Sillito, A. M., Kemp, J. A., Milson, J. A., & Berardi, N. (1980). A re-evaluation of the mechanisms underlying simple cell orientation selectivity. *Brain Research*, 194, 517–520.
- Squire, L. R., Stark, C. E., & Clark, R. E. (2004). The medial temporal lobe. *Annual Review of Neuroscience*, 27, 279–306.
- Squire, L. R., & Zola-Morgan, S. (1983). The neurology of memory: The case for correspondence between the findings for human and nonhuman primate. *The physiological basis of memory* (pp. 199–267). Academic Press.
- Squire, L. R., & Zola-Morgan, S. M. (1991). The medial temporal lobe memory system. *Science*, 253, 1380–1386.
- Stark, C. E., & Squire, L. R. (2000). Intact visual perceptual discrimination in humans in the absence of perirhinal cortex. *Learning and Memory*, 7, 273–278.
- Suzuki, W., & Amaral, D. G. (1994). Perirhinal and parahippocampal cortices of the macaque monkey: Cortical afferents. *Journal of Comparative Neurology*, 350, 497–533.
- Suzuki, W. A. (2009). Perception and the medial temporal lobe: Evaluating the current evidence. *Neuron*, 61, 657–666.
- Suzuki, W. A., & Baxter, M. G. (2009). Memory, perception, and the medial temporal lobe: A synthesis of opinions. *Neuron*, 61, 678–679.
- 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. *Journal of Neuroscience*, 13, 2430–2451.
- Tanaka, K., Saito, H., Fukada, Y., & Moriya, M. (1991). Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *Journal of Neurophysiology*, 66, 170–189.
- Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. *Science*, 247, 301–306.
- Tyler, L. K., Stamatakis, E. A., Bright, P., Acres, K., Abdallah, S., Rodd, J. M., et al. (2004). Processing objects at different levels of specificity. *Journal of Cognitive Neuroscience*, 16, 351–362.
- Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience*, 5, 682–687.
- van Strien, N. M., Scholte, H. S., & Witter, M. P. (2008). Activation of the human medial temporal lobes by stereoscopic depth cues. *Neuroimage*, 40, 1815–1823.
- Wallis, G., & Rolls, E. T. (1997). Invariant face and object recognition in the visual system. *Progress in Neurobiology*, 51, 167–194.
- Warrington, E. K., & Weiskrantz, L. (1968). New method of testing long-term retention with special reference to amnesic patients. *Nature*, 217, 972–974.
- Wilson, C. J. (1995). The contribution of cortical neurons to the firing pattern of striatal spiny neurons. In J. C. Houk, J. L. Davis, & D. G. Beiser (Eds.), *Models of information processing in the basal ganglia* (pp. 29–50). Cambridge, MA: MIT Press.
- Wilson, M., & Kaufman, H. M. (1969). Effect of inferotemporal lesions upon processing of visual information in monkeys. *Journal of Comparative and Physiological Psychology*, 69, 44–48.
- Wilson, M., Kaufman, H. M., Ziemer, R. E., & Lieb, J. P. (1972). Visual identification and memory in monkeys with circumscribed inferotemporal lesions. *Journal of Comparative and Physiological Psychology*, 78, 173–183.
- Zhang, L., & Cottrell, G. (2005). *Holistic processing develops because it is good*. Proceedings of the 27th Annual Cognitive Science Conference, Stresa, Italy. Mahwah: Lawrence Erlbaum.