

# A Feedback Model of Perceptual Learning and Categorisation

**M. W. Spratling**

Division of Engineering, King's College London. UK.  
and Centre for Brain and Cognitive Development, Birkbeck College, London. UK.

**M. H. Johnson**

Centre for Brain and Cognitive Development, Birkbeck College, London. UK.

## Abstract

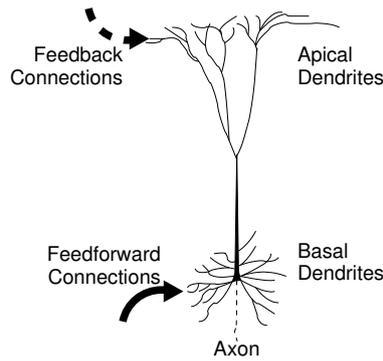
Top-down, feedback, influences are known to have significant effects on visual information processing. Such influences are also likely to affect perceptual learning. This article employs a computational model of the cortical region interactions underlying visual perception to investigate possible influences of top-down information on learning. The results suggest that feedback could bias the way in which perceptual stimuli are categorised and could also facilitate the learning of sub-ordinate level representations suitable for object identification and perceptual expertise.

**Keywords:** Perception; Learning; Neural Networks; Representation.

## 1 Introduction

Visual perception is accomplished by a distributed network of interconnected and interacting neural populations, or cortical regions. The activities of the neurons within these regions underly perceptual processes. Furthermore, changes in the response properties of such neurons underly changes in perceptual abilities (Karni, 1996; Walsh et al., 1998; Gauthier et al., 2000; Sigman and Gilbert, 2000). The interconnectivity between different cortical regions makes it possible to identify specific information processing pathways, along which neural populations are arranged in a hierarchy (Felleman and Van Essen, 1991; Crick and Koch, 1998; Mountcastle, 1998). During perception, information propagates through the visual processing hierarchy from primary sensory areas to higher cortical regions (via feedforward connections), and in the reverse direction (via feedback connections). The resulting neural activity, within each cortical region, is thus influenced by both feedforward and feedback information (Lamme et al., 1998; Lee et al., 1998; Bullier and Nowak, 1995; Lamme and Roelfsema, 2000). Hence, both feedforward and feedback connections cause short-term changes in cortical information processing by directly influencing neural activity. In addition, since learning is activity-dependent, each type of connection is involved in modifying the response properties of neurons, in the long-term, through learning. In this article we explore the role of cortical feedback connections in perceptual learning and in subsequent perceptual processing using a simple model of two interacting cortical regions at adjacent stages along an information processing pathway. Our results suggest that feedback may be involved in learning representations suitable for performing both categorisation and identification.

Neurons become active in response to particular patterns of pre-synaptic stimulation. These stimulus selectivities, for neurons in all cortical regions, are dependent on prior experience (Mountcastle, 1998; Logothetis, 1998; Wallis and Bühlhoff, 1999; Sigala and Logothetis, 2001) and are subject to change throughout life (Sagi and Tanne, 1994; Desimone, 1996; Gilbert, 1996; Karni and Bertini, 1997; Kobatake et al., 1998; Goldstone, 1998; Logothetis, 1998). Individual neurons learn to become selective for statistical regularities in the input environment which they experience (Földiák, 1990; Barlow, 1990). This 'environment' is the pattern of pre-synaptic activity to which that neuron is exposed. The environment is thus different for neurons in different cortical regions. Neurons in one cortical region will learn to respond to the patterns of activity generated by the neurons in other cortical regions from which they receive their inputs. The uniformity of cortical anatomy and physiology (Ebdon, 1992; Mountcastle, 1998) suggests that the same processes of representation and adaptation underly both low-level, perceptual, and high-level, cognitive, systems and the learning within each (Goldstone and Barsalou, 1998). Hence, many different phenomena in the domain of perceptual and conceptual learning are likely to result from the same mechanisms of change in neural response properties. Consistent with this view, the model presented in this paper demonstrates that a single learning mechanism can account for several different forms of visual adaptation, such as, imprinting, perceptual category learning, and perceptual expertise. Furthermore, the model demonstrates that common mechanisms, involving cortical feedback pathways, can account for a range of visual processes, such as, object parsing, categorical perception and dimension sensitisation. In addition to providing a unified account of a



**Figure 1:** A pyramidal cell. Pyramidal cells are the predominant class of excitatory neuron in the cerebral cortex. The detailed dendritic morphology of pyramidal cells is highly variable, however, one relatively consistent feature is the segregation into two distinct dendritic projections: the apical dendrite which extends towards more superficial cortical layers, and the basal dendrites which extend laterally within the same layer as the cell body. For pyramidal cells in the superficial layers of the cortex, feedforward connections (from regions lower in the cortical hierarchy), relayed by spiny-stellate cells in layer IV, target the basal dendrites, while feedback connections (from regions higher in the cortical hierarchy) target the apical dendrites. Cell morphology adapted from (Paré et al., 1998).

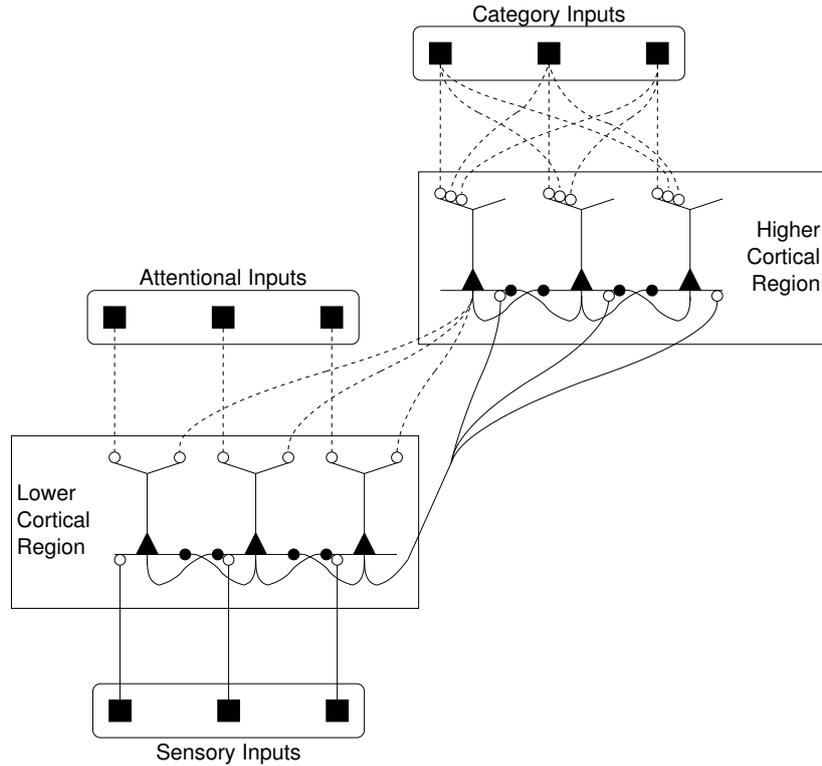
number of perceptual processes that are currently considered distinct, the proposed model is consistent with neuroscience. If our objective in studying psychology rather than artificial intelligence is to explain human behaviour, then there is a need to focus on those models that are most compatible with the neural mechanisms which underly psychological phenomena. Hence, we propose that a proper consideration of brain mechanisms is likely to result in more adequate models of cognition. The same model has been applied previously to simulating single-cell data in the domain of spatial and object-based attention (Spratling and Johnson, 2004a). The range of processes that can be simulated by this model supports the view that the same underlying neural mechanisms give rise to a number of distinct psychological phenomena.

## 2 Methods

Cortical feedforward and feedback connections preferentially target distinct regions of pyramidal cell dendrites (see figure 1). For example, pyramidal cells in cortical layers II and III receive feedforward information at the basal dendrites and feedback information at the apical dendrites. Physiological evidence suggests that this anatomically observed segregation of input sources, between the apical and basal dendrites, may have functional significance (Spratling, 2002). The apical dendrite appears to act as a functionally distinct dendritic compartment capable of integrating the stimulation it receives relatively independently of the basal dendrite (Häusser and Mel, 2003; Larkum et al., 1999; Yuste et al., 1994). Furthermore, activation of the apical dendrite causes weaker depolarisation of the cell than basal dendrite stimulation (Rockland, 1998; Budd, 1998), which is consistent with the suggestion that feedback signals act to modulate responses that are primarily driven by feedforward inputs (Crick and Koch, 1998; Hupé et al., 1998; Friston and Büchel, 2000).

Functionally distinct apical and basal dendrites enable information arising from different sources to have disparate affects on neural activity and (thereby) learning (Spratling, 2002). Based on these observations we have developed a neural network model in which neurons have two dendritic compartments capable of independently integrating the feedforward and feedback signals that they receive (Spratling and Johnson, 2004a). In this model, the response of a neuron is driven by the activity generated in its basal dendrite and modulated by the activity of the apical dendrite. The response is also affected by lateral inhibitory connections targeting the basal dendrites, via which neurons in the same region compete to represent distinct stimuli (Spratling and Johnson, 2001, 2002). This model thus combines two mechanisms, competition and gain modulation, which have both been separately suggested as common computational principles employed throughout the cortex (Keysers and Perrett, 2002; O'Reilly, 1998; Phillips and Singer, 1997; Salinas and Thier, 2000).

This proposed neural network architecture can be used to model the interactions between cortical regions, as illustrated (for two interacting regions) in figure 2. In the model, visual stimuli compete to be represented by cortical activity. The outcome of this competition is influenced not only by bottom-up, sensory-driven, processes, but also by top-down information, which can originate from a range of different sources, to modulate neural

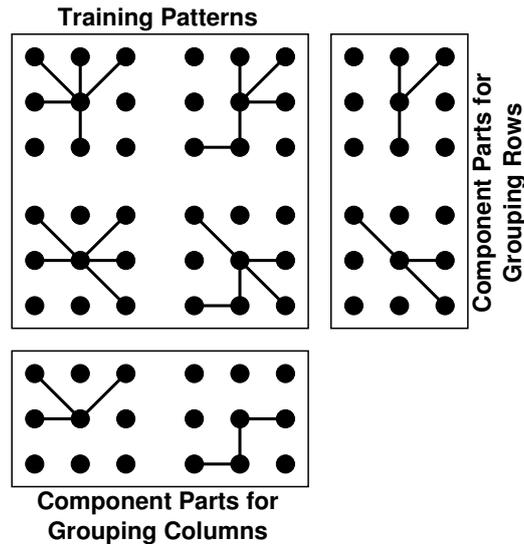


**Figure 2:** A schematic of the model showing two interacting cortical regions. Each region contains one layer of model pyramidal cells (somata shown as filled triangles). These model neurons each have two, independent, dendritic compartments: the basal dendrites, which receive feedforward connections (shown as solid lines), and the apical dendrites, which receive feedback connections (shown as dashed lines). Nodes in different regions are reciprocally connected by feedforward and feedback projections (only the connections originating from one pair of nodes are shown) and are targeted by connections originating from other sources, such as sensory inputs (shown as filled squares). Nodes within the same region compete via lateral inhibitory connections which target the basal dendrites (only selected connections between neighbouring nodes are shown). Excitatory and inhibitory synapses are shown as open and filled circles respectively.

activity and bias competition. Hence, in the model, neural activity at later stages of perceptual processing can influence the on-going processing occurring at earlier stages. Full implementation details are provided in the Appendix.

Figure 2 illustrates the structure of the model used in all the simulations reported in this article. The lower and upper regions contained 29, and six, neurons respectively. These figures for the number of nodes were determined by structure of the particular tasks the network was applied to. Sensory input to the lower region consisted of a 29-element vector, each element of which indicated the strength of activation of an image component (a line segment or a dot). The 29 neurons in the lower region each received input from one element of the input vector. Neurons in the lower region thus had predefined preferences, such that each node was selective for a unique, independent, image feature. It was thus assumed, for the purposes of the experiments described here, that the lower region has previously learnt appropriate low-level sensory representations. The model was used to explore the effects of subsequent learning in the upper region. The nodes in the upper region received feedforward connections from all the nodes in the lower region. These nodes thus had larger receptive fields and could learn to become selective to patterns of activity in the outputs of all the nodes in the lower region, and hence, to larger features within the input images. The synaptic strengths of the reciprocal, feedforward and feedback, connections between the two regions were learnt (as described in the Appendix). The resulting selectivities of the nodes in the upper region were thus dependent on prior experience.

Separate sources of feedback targeted the apical dendrites of nodes in each region. For feedback to the lower region, each node received top-down input from a specific source. Such connections would be necessary to be able to direct attention to a particular feature or to a particular spatial location (Spratling and Johnson, 2004a). In contrast, feedback inputs to the upper region were in the form of an array of input signals fully connected to all the



**Figure 3:** The task used in [Pevtzow and Goldstone \(1994\)](#) to assess the effects of prior categorisation experience on object segmentation. When the two objects in each row were placed in the same category, the parts shown on the right were diagnostic for those categories. When the two objects in each column were placed in the same category, the parts shown at the bottom were diagnostic for those categories.

nodes. The strength of the synaptic weights from these inputs to the apical dendrites of nodes in the upper-region were learnt at the same time as the synaptic weights at the basal dendrites were learnt. Distinct, non-overlapping, input patterns denoting the category of the perceptual stimulus were used as top-down input to the upper region in certain experiments.

## 3 Results

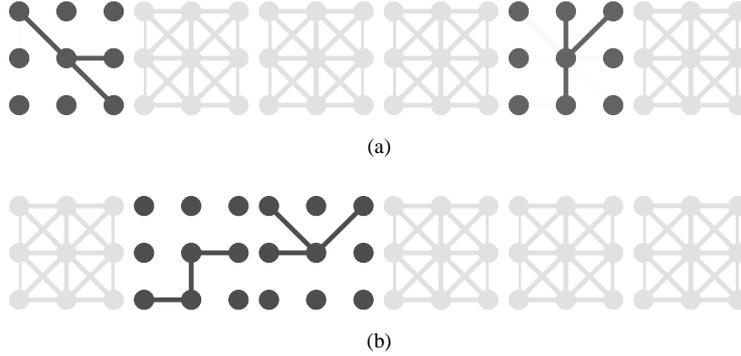
### 3.1 Imprinting and Unsupervised Perceptual Learning

Previous experiments with this model, have demonstrated that it provides a powerful, and efficient, mechanism for the unsupervised learning of perceptual stimuli. The algorithm finds statistical regularities in the sensory input and is able to learn the underlying, independent, components of images ([Spratling and Johnson, 2001, 2002, 2003](#)). Furthermore, the same learning rule can be applied, simultaneously, at each stage in a hierarchy of regions so that higher-level neurons become tuned to more abstract features of the training environment ([Spratling and Johnson, 2004b](#)). The following sections explore the influence of top-down information on the learning process.

### 3.2 Category Learning and Object Parsing

[Pevtzow and Goldstone \(1994\)](#) explored the effects of category learning on a subsequent perceptual segmentation task. During the initial category training phase of the experiment, subjects were shown distortions of four objects. These objects were made up of line segments, or bars, joining dots on a three-by-three grid (see figure 3) and were distorted for each presentation by the addition of a further, randomly chosen, line segment. One set of participants learnt that the pairs of objects in each row of figure 3 were members of the same category, while a second set of subjects were trained to place each pair of objects in the columns of figure 3 into the same category. Following category training, subjects were tested on their ability to determine if a novel object contained a specific part. Participants were faster to correctly identify a part as present when that part was diagnostic for a previously learnt perceptual category ([Pevtzow and Goldstone, 1994; Goldstone et al., 2000](#)).

This task was simulated using the model described in the Methods. Distorted versions of the four training patterns shown in figure 3 were presented as sensory input to the basal dendrites of the lower region. The training patterns were distorted by adding one extra bar, chosen at random, at each iteration. In addition, inputs indicating the category of the current stimulus were provided to the feedback connections targeting the apical dendrites of the upper region. Figure 4 shows typical examples of the stimulus selectivities learnt by neurons in the upper region following training. Nodes in the upper region learnt to represent the common features (or prototypes) corresponding to whichever categories the network was biased towards learning. Hence, when the category information

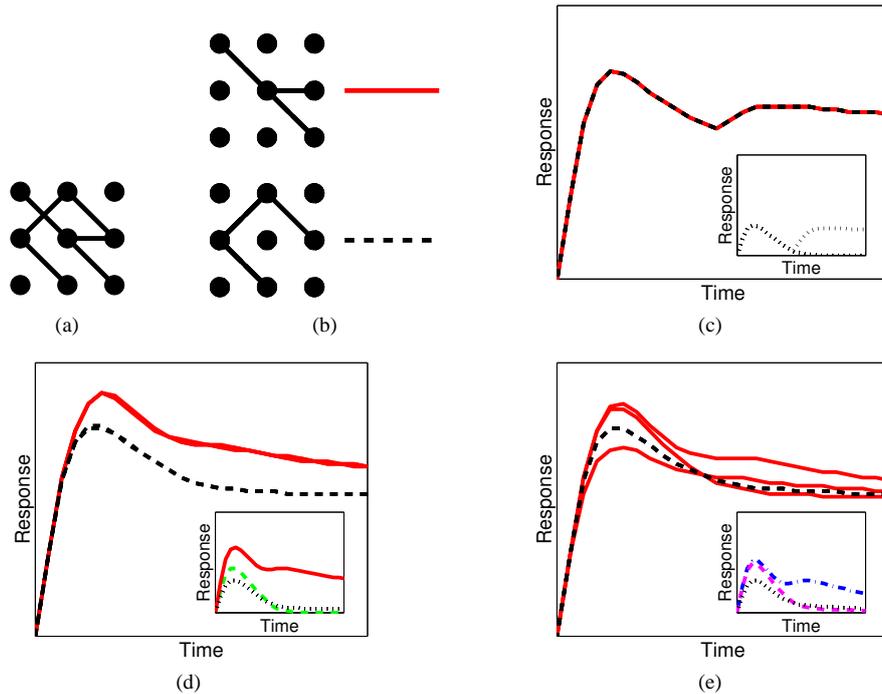


**Figure 4:** The representations learnt by neurons in the upper region, for the task shown in figure 3, (a) when categorisation information biases grouping rows, and, (b) when categorisation information biases grouping columns. The strength of synaptic weights received by each node from each input feature is proportional to the darkness of that feature in the image.

indicated that patterns within the same row were in the same category, one node in the upper region learnt to represent the common features of the two patterns in the top row, while another node learnt the common features of the two patterns in the bottom row. This occurred for 17 trials out of 20. A typical example of the selectivities learnt by nodes in the upper region is shown in figure 4(a). In contrast, when the category information indicated that patterns within the same column formed distinct categories, one node in the upper region learnt to represent the common features of the two patterns in the left column, while another node learnt the common features of the two patterns in the right column. This occurred for 19 trials out of 20. A typical example of the selectivities learnt by nodes in the upper region is shown in figure 4(b).

The response of the network was tested using a novel stimulus, similar to that used by [Pevtzow and Goldstone \(1994\)](#) in their whole/part judgement task. No category information was supplied during this test. The stimulus (figure 5(a)) contained six bars, three of which were members of one of the prototypes learnt by the network when grouping rows. When this stimulus was presented to a network that had been previously trained to classify patterns within rows, the node in the upper region that represented this prototype became strongly active. This node thus provided feedback that enhanced the responses of the nodes in the lower region representing the individual bars within this prototype (figure 5(d)). No such consistent enhancement to the activity of nodes representing one or other part of the object occurred when the network was trained to group each column of patterns into the same category (figure 5(e)).

Similar enhanced neural responses have been observed in single-cell electro-physiological data recorded in adult monkeys performing figure/ground segmentation tasks. In these experiments feedback enhances the activities of neurons representing spatially distributed image features on the figure ([Zipser et al., 1996](#); [Lamme, 1995](#); [Hupé et al., 1998](#)) and, hence, provides these elements with a competitive advantage over neurons representing image features on the background ([Reynolds and Desimone, 1999](#)). We have previously employed a similar model to the one described here to simulate these effects in a figure/ground segmentation task ([Spratling and Johnson, 2004a](#)). In the current application, the model suggests that participants were faster to correctly identify a part that was diagnostic for a previously learnt perceptual category ([Pevtzow and Goldstone, 1994](#); [Goldstone et al., 2000](#)) due to segmentation being facilitated by feedback from a previously learnt category prototype. The influence of high-level shape information on the grouping of low-level image features has also been observed in other behavioural experiments. For example, a familiar or identifiable region of an ambiguous image is more likely to be perceived as the figure ([Peterson and Gibson, 1993](#)). Furthermore, familiarisation with an object can bias subsequent perceptual grouping and facilitate processing of disparate image segments as a unitary whole ([Zemel et al., 2002](#)). Similarly, prior experience can enable the segmentation of abutting stimuli with ambiguous part membership even when low-level cues (such as common motion or spatial separation) are unavailable ([Bravo and Farid, 2003](#)), even for infants as young as 4 months ([Needham, 2001](#)). More generally, a range of contextual cueing tasks ([Chun, 2002](#)) demonstrate that high-level knowledge produces top-down biases that influence low-level perceptual processing. The model suggests that the same mechanisms underly all these phenomena. Neurons with large receptive fields, that are sensitive to high-level information in the image, provide feedback modulation to neurons with smaller receptive fields in earlier regions. Such top-down signals will be dependent upon the response properties of the neurons which generate them. Hence, the processing of perceptual data, in lower cortical regions, is biased by the selectivities learnt, from prior experience, by neurons in higher areas ([Lee et al.,](#)



**Figure 5:** Responses to a stimulus, before and after training on a perceptual categorisation task. A novel stimulus, shown in (a) was presented to the network. Responses are shown for two sets of nodes in the lower region, that were selective for the subsets of line segments indicated in (b). The insets in each figure show the responses of the nodes in the upper region: dotted lines show responses for nodes that are unselective to any patterns, while coloured solid and dashed lines show the responses of upper-region nodes that have learnt to represent prototypes. (c) Before training, the upper region nodes generate a weak response to the stimulus, and feedback is unselective so that the responses of all the recorded nodes in the lower region are equal. (d) After training to group rows, when nodes in the upper region represent the stimuli shown in figure 4(a), the novel stimulus strongly activates one upper region node. Feedback activation from this node results in enhanced responses for one subset of lower region nodes. (e) After training to group columns, when nodes in upper region represent the stimuli shown in figure 4(b), the novel pattern partially activates both prototypes in the upper region, and there is no consistent enhancement to the response of either subset of lower region nodes. Both response and time are measured in arbitrary units, but the same scale has been used in each plot.

2002). High-level object and category knowledge (resulting from prior experience) can thus enhance the activities of lower-level neurons representing certain image features to affect on-going perceptual information processing.

### 3.3 Categorical Perception and Dimension Sensitisation

Categorical perception is another phenomenon that can be induced by prior perceptual experience. It describes a warping of perceptual space, such that there is a perceived increase in the similarity between items within the same category, and a perceived increase in dissimilarity between items in distinct categories (Harnad, 1987, 2003; Goldstone, 1998; Goldstone et al., 2001). The model we have presented above can also simulate these categorical perception effects. In the previous experiment, activity in the upper region resulted in selective enhancement of neural responses in the lower region. One consequence of this top-down modulation of neural responses is that similarities between the current image and previously learnt object representations or category prototypes will be exaggerated. Hence, if a novel object is sufficiently similar to a previously learnt object to cause activation of the stored object representation, this will result in feedback enhancement to those neurons representing features of the stimulus that match the learnt object. Similarly, if a novel object is sufficiently similar to a previously learnt category prototype to cause activation of this stored representation, this will result in feedback enhancement to those neurons representing features of the novel object that match the learnt prototype (as in figure 5(d)). Such an enhancement in the neural responses to features that match the prototype is likely to result in an increased perceptual similarity between the stimulus and the learnt category. When the stimulus is similar to more than one

stored representation, competition between these representations will determine which stimulus features receive top-down support. Hence, similar stimuli which fall within different categories will receive distinct patterns of feedback modulation that will lead to an exaggeration of the differences between the stimuli.

This is illustrated in figure 6. A network was used that had been trained in the [Pevtzow and Goldstone \(1994\)](#) task (see figure 3) to classify patterns within rows, and which had thus learnt representations in the upper region for the prototypes of each row, like those shown in figure 4(a). Figure 6 shows the response of this network to a range of stimuli, which have varying similarities to the two prototype that had been learnt by the network. Responses are shown for the nodes in the lower region that are selective to the three bars that form each of the previously learnt prototypes (as indicated in figure 6(a)). The insets of each figure show the responses of the nodes in the upper region that are selective to each of these two prototypes. It can be seen that in each case one node in the upper region wins the competition and suppresses the response of all other upper-region nodes. The winning node provides feedback that amplifies the responses of neurons encoding image features consistent with that category, and hence, enhances the perceptual similarity between the stimulus and the winning category as well as exaggerating the differences between stimuli in different categories. The outcome of the competition within the upper region is sensitive to small differences in the stimulus. Hence, near the boundary between the two categories, a small change in the stimulus can give rise to a large difference in the resulting neural response. The response to a stimulus that is equally similar to both stored prototypes is shown in the central plot of figure 6(b). In this case, noise within the neural activations results in one upper-region node winning the competition. Previous models of categorical perception have also relied on stimuli from different categories causing the network to fall into distinct activation states ([Damper and Harnad, 2000](#)). For our model the different activation states result from competition between the nodes in the upper-region, which affects the activation states of nodes throughout the network via feedback.

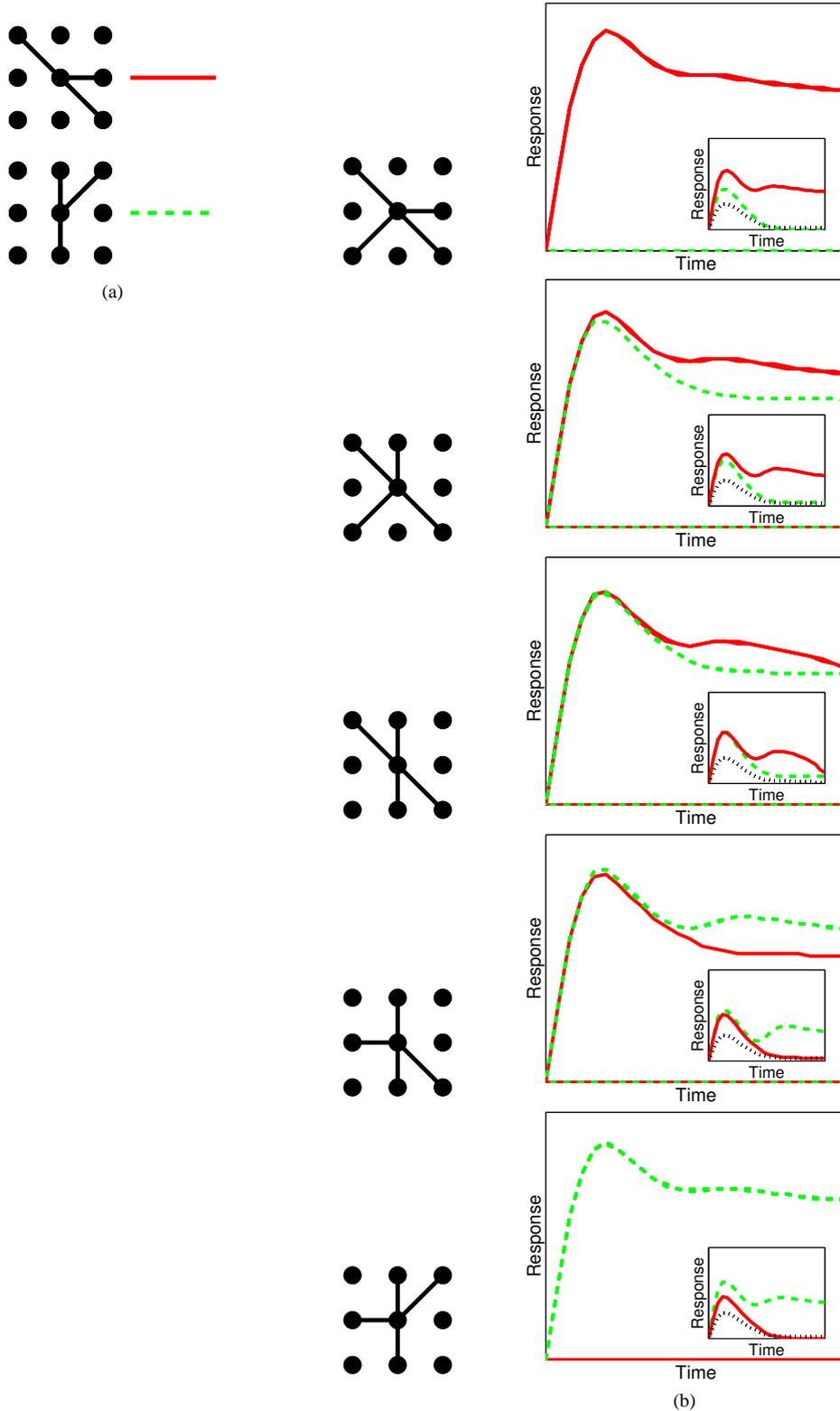
To quantify the effects of category learning on perceptual similarity and dissimilarity a second simulation was performed. [Goldstone et al. \(2001, 2000\)](#) describe how categorisation training can increase a subject's sensitivity to a specific perceptual dimension. In these experiments a four-by-four array of training images was used. These images were created either by changing parameters along two perceptual dimensions (*i.e.*, the brightness and size of images of squares) or by morphing between four stimuli (*i.e.*, faces). Subjects were trained on a categorisation task that divided the stimulus array either vertically or horizontally into two equal groups. Hence, for successful categorisation one dimension was relevant and the other irrelevant. Subsequent testing revealed that subjects were most sensitive to perceiving stimulus differences in the relevant dimension across the boundary between the two categories, and were least sensitive to perceiving changes along the dimension that was irrelevant for categorisation.

The model was applied to a similar task, in which training data consisted of the four-by-four array of images shown in figure 7(a). This array was created by morphing between the four stimuli employed previously (as shown in figure 3). Images were presented in random order to the basal dendrites of the lower region, together with inputs targeting the apical dendrites of the upper region that indicated the category of the current stimulus. Networks were either trained on a categorisation task that divided the stimulus array horizontally or vertically. This is similar to the previous task in which patterns within rows or columns were placed in the same category. As in the previous task, nodes in the upper region often learnt prototypes similar to those shown in figure 4. Before and after training, the responses generated in the lower region to each image in the array were recorded. These patterns of neural activity were compared to determine how distinctly different images were represented by the network (see section A.1). The distinctiveness of the neural response, to adjacent images from the training array, averaged over ten trials with each of the two classification tasks is shown in figures 7(b) and 7(c). It can be seen that, as with the behavioural data ([Goldstone et al., 2001, 2000](#)), the largest changes in response occur across the boundary between the two learnt categories, and that the smallest differences in response occur along the dimension that was irrelevant for categorisation.

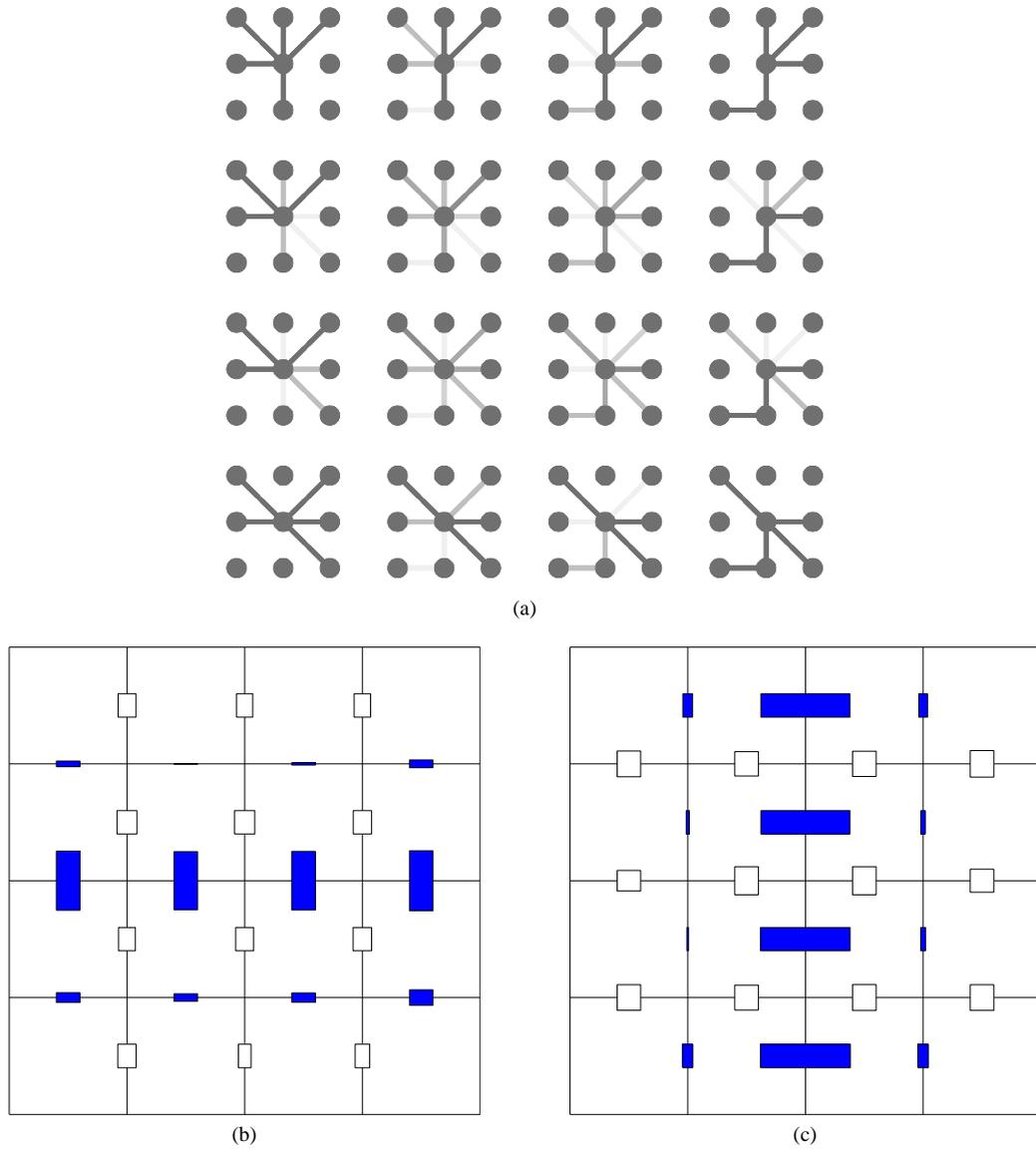
Similar effects on the sensitivity to specific stimulus dimensions, following category training, have been observed in monkeys ([Sigala and Logothetis, 2001](#)). After training on a categorisation task for which only certain stimulus features were relevant, neurons in inferior temporal cortex were found to generate responses that were most distinct for stimulus changes along the relevant dimensions rather than the irrelevant dimensions ([Sigala and Logothetis, 2001](#)). One possible explanation for these effects, in agreement with our model, is that top-down information from other cortical regions serves to shape the sensitivity of IT neurons to the diagnostic stimulus dimensions ([Hasegawa and Miyashita, 2002](#)).

### 3.4 The Effects of Feedback on Categorisation and Differentiation

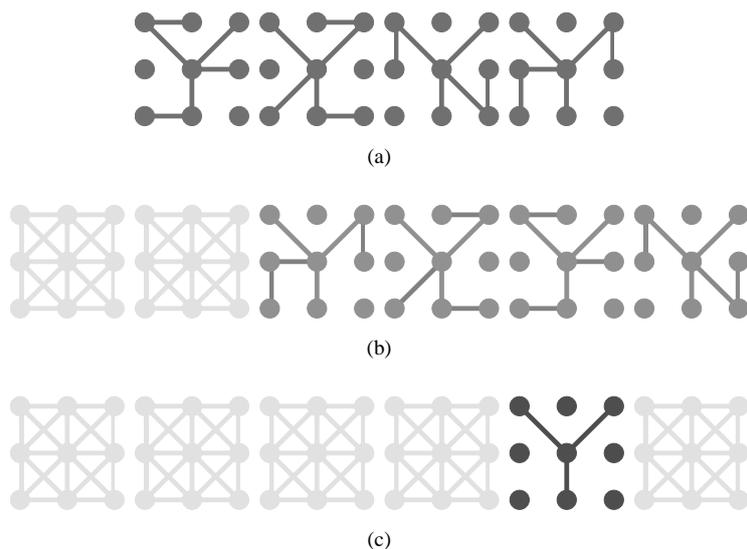
The previous experiments have shown that category information, input to the apical dendrites of the upper region, can influence the learning of prototype representations at the basal dendrites. If there is insufficient overlap



**Figure 6:** The response of a network to patterns that vary in their similarity to two learnt prototypes. Responses are shown, in (b), for two sets of nodes in the lower region that represent the line segments indicated in (a). Note that in the top and bottom graphs all the nodes representing one or other of these sets of line segments have zero activation at all times. Insets, show the responses of nodes in the upper region that are selective to the two prototypes.



**Figure 7:** (a) A four-by-four array of training images generated by morphing between the four images shown in figure 3. The plots in (b) and (c) show how distinct the neural responses were to adjacent images within the array, when (b) the network had been trained to classify images in the upper and lower halves of the array into separate categories, and (c) when the network had been trained to classify images in the left and right halves of the array into separate categories. The length of each rectangle is proportional to the magnitude of the  $\Delta\epsilon$  values (determined for the lower region using equation 9), such that a value of  $\Delta\epsilon = \pm 1$  would be shown as a rectangle with length equal to the side of one of the small squares in the grid. Positive  $\Delta\epsilon$  values are shown by filled rectangles, and negative values are shown using empty rectangles. Positive values indicate that training has resulted in improved discrimination between adjacent stimuli, while negative values indicate increased similarity in the representations of adjacent stimuli.

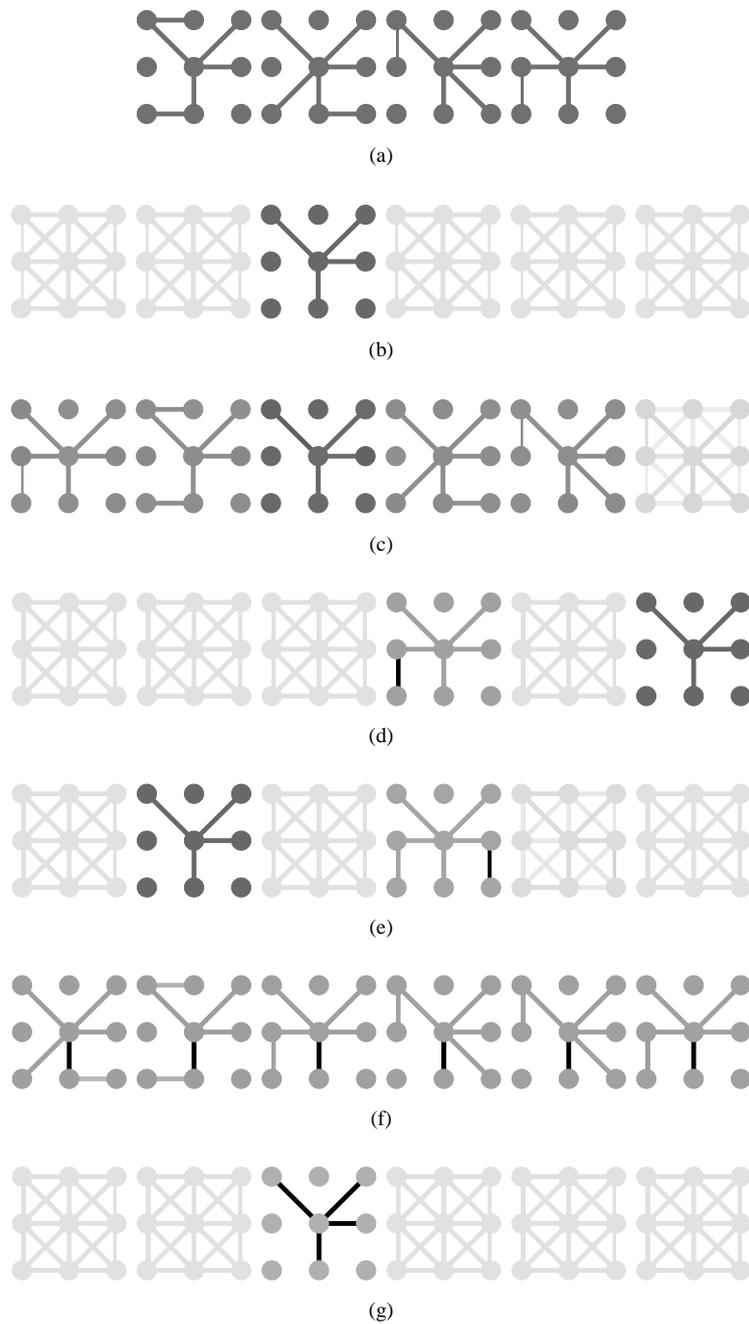


**Figure 8:** (a) Training patterns that have less overlap with each other than the previous example. The representations learnt by neurons in the upper region, when trained on this data, (b) with no top-down bias, and (c) with a top-down signal to the upper region that is the same for each exemplar. The strength of synaptic weights received by each node from each input feature is proportional to the darkness of that feature in the image.

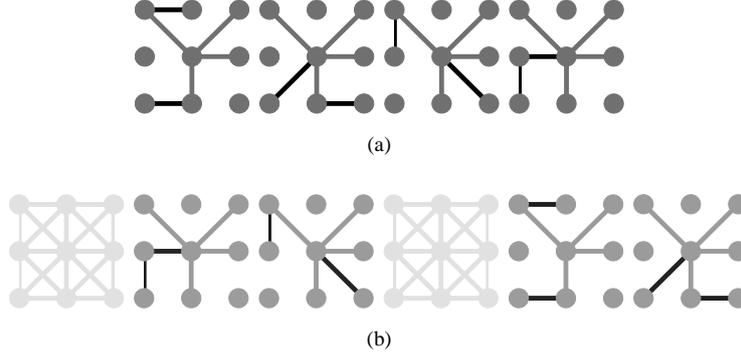
between the different training patterns then, in the absence of such category information, the network will learn to represent the individual exemplars. For example, distorted versions of the four training patterns shown in figure 8(a) were presented as sensory input to the lower network. The training patterns were distorted by adding one extra bar, chosen at random, at each iteration. A typical example of the stimulus selectivities learnt by nodes in the upper region is shown in figure 8(b). All the individual exemplars were learnt in 100% of 20 trials. To demonstrate the influence of apical inputs on learning, an unvarying input pattern was supplied to the feedback connections targeting the upper region. This input indicates that all the exemplars were members of the same perceptual category and it biases the network to use the same node to represent all the input images. Hence, a neuron in the upper region learns the prototype of the four training images (as shown in figure 8(c)). The prototype was learnt in 100% of 20 trials.

In contrast, if there is a significant overlap between the different training patterns then the network will learn to represent that overlap in the absence of any category information. For example, distorted versions of the four training patterns shown in figure 9(a) were presented as sensory input to the lower network. The training patterns were distorted by adding one extra bar, chosen at random, at each iteration. A typical example of the stimulus selectivities learnt by nodes in the upper region is shown in figure 9(b). The prototype was learnt in 90% of 20 trials. In this case, feedback information could be used in an attempt to bias learning of individual exemplars. For each exemplar a distinct pattern could be supplied to the feedback connections targeting the upper region. However, in only 20% of trials did such feedback to the upper region succeed in generating separate representations of each exemplar.

Feedback directed to the apical dendrites of nodes in the lower region can bias the network to learn representations of individual exemplars. In contrast to the feedback supplied to the upper region, each neuron received input from one element of a 29-element input vector, so that feedback to the lower region was directed towards individual nodes (see Methods). When feedback was directed to one lower region node, chosen at random on each iteration, nodes in the upper region learnt to represent the individual exemplars in 95% of trials. A typical example of the stimulus selectivities learnt by nodes in the upper region is shown in figure 9(c). It can be seen that in this example the network has learnt the prototype in addition to the individual exemplars (this occurred in 80% of trials). This result suggests that in order to differentiate perceptually similar stimuli it is necessary to selectively amplify responses to the individual features of those stimuli. In the cortex, selective enhancement of specific neural responses can occur as a result of spatial or object-based attention (Olson, 2001; Reynolds et al., 2000; Kastner and Ungerleider, 2000; Luck et al., 1997; McAdams and Maunsell, 2000). Attention operates via the cortical feedback projections (Treue, 2001; Schroeder et al., 2001; Mehta et al., 2000; Desimone and Duncan, 1995; Olson et al., 2001), hence, feedback to the lower region in the model is assumed to result from attention (Spratling



**Figure 9:** (a) Training patterns that have more overlap with each other than the previous example. The representations learnt by neurons in the upper region, when trained on this data, (b) with no top-down bias, (c) when feedback is directed to an image feature chosen at random at each cycle, (d) when feedback is directed to the vertical bar at the lower-left corner of the input image (a feature that is unique to one exemplar), (e) when feedback is directed to the vertical bar at the lower-right corner of the input image (a feature that is not part of any exemplar), (f) when feedback is directed to the vertical bar at the centre of the lower-half of the input image (a feature that is shared by all the exemplars), (g) when feedback is directed to a all the bars in the prototype. The strength of synaptic weights received by each node from each input feature is proportional to the darkness of that feature in the image.



**Figure 10:** (a) Training patterns like those in figure 9(a), but in which the differences between exemplars are highlighted by being presented at a stronger contrast. (b) The representations learnt by neurons in the upper region, when trained on this data.

and Johnson, 2004a). The following experiments explored the effects of providing top-down enhancement (or directing attention) to different image features.

When feedback was directed to one lower region node, representing a feature that is unique to one exemplar, nodes in the upper region learnt to represent that particular exemplar and the prototype in 70% of trials. A typical example of the stimulus selectivities learnt by nodes in the upper region is shown in figure 9(d). Alternatively, when feedback was directed to one lower region node, representing a feature that is not part of any exemplar, nodes in the upper region learnt to represent the prototype in 90% of trials. In many of these trials, representations of exemplars together with the attended feature were also learnt. These patterns, containing the extra feature, were presented to the network during training, since patterns were distorted by adding one extra bar, chosen at random, at each iteration. A typical example of the stimulus selectivities learnt by nodes in the upper region is shown in figure 9(e). These two results suggest that attention to a specific feature, will lead to patterns containing that feature being more easily learnt.

When feedback was directed to one lower region node, representing a feature that is shared by all the exemplars (*i.e.*, one of the features of the prototype), nodes in the upper region learnt to represent all four exemplars in 100% of trials (on 20% of these trials the prototype was also learnt). A typical example of the stimulus selectivities learnt by nodes in the upper region is shown in figure 9(f). This effect results from an idiosyncrasy of the algorithm. Without feedback a node in the upper region will win the competition to represent the first exemplar presented to the network. Due to the similarity of other exemplars, the same node will also win subsequent competitions, and hence will refine its afferent weights to represent the overlap between all the exemplars. In the presence of feedback, one input to the upper region is amplified and the node that wins the initial competition will learn a stronger synaptic weight to this feature. Due to the normalisation of the synaptic weights that occurs during competition between nodes (see equation 4 in the Appendix), this increased weight makes the node less able to prevent other nodes from being activated by other inputs. Other nodes in the upper region are thus able to successfully compete to represent other input patterns. A more expected result occurred when feedback was directed to all the lower region nodes representing line segments shared by all the exemplars (*i.e.*, all the bars contained in the prototype). In this condition, nodes in the upper region learnt to represent the prototype in 100% of trials. A typical example of the stimulus selectivities learnt by nodes in the upper region is shown in figure 9(g).

Feedback directed to the apical dendrites of nodes in the lower region selectively enhances the responses of those nodes. The strength of the response of a neuron in the lower region can also be affected by changes in the strength of the stimulation presented at the basal dendrites. Hence, results similar to those presented above can also be produced by manipulating the relative contrasts of specific image features. In the experiments presented above, feedback caused an apical activation of 0.75, and hence the response of the attended node (from equation 5, see Appendix) was  $y_{jk}^t = 1.75 \times y_{jk,basal}^t$ . A change in stimulus contrast that increased the activation of the basal dendrite by a factor of 1.75 would thus cause an equal increase in the node's activity in the absence of any feedback. The effects of changing the stimulus strength is illustrated in figure 10. The network was trained on stimuli identical to those used in the experiments immediately above, but with the unique features of each exemplar presented at 33% higher contrast (figure 10(a)). Nodes in the upper region learnt to represent all the individual exemplars in 19 out of 20 trials. A typical example of the stimulus selectivities learnt by nodes in the upper region is shown in figure 10(b). This compares to the prototype being learnt in 100% of trials when the stimuli were presented at a uniform contrast (see above).

Perceptual categories become more differentiated with age (for developing infants, [Quinn and Johnson, 2000](#)) and with expertise (for learning adults, [Goldstone, 1998](#)), enabling finer discriminations to be made ([Goldstone, 1998](#)). Experience thus appears to lead to the formation of more specific representations. The results above thus suggest that learning perceptual expertise within a domain will be facilitated if the distinctive or unique features of different exemplars are highlighted either by drawing attention to them or by presenting them at a higher contrast.

Figure 11 explores the response of the network when it has learnt a number of similar patterns. The network was trained on the patterns shown in figure 9(a), with feedback directed to randomly chosen nodes in the lower region. This resulted in the prototype and each individual exemplar being represented by nodes in the upper region (similar to the previous result shown in figure 9(c)). For each test pattern presented to the network, nodes in the upper region generate an initial transient activation, during which the competition between the nodes is resolved. This initial activity is followed by the sustained response of a single node.

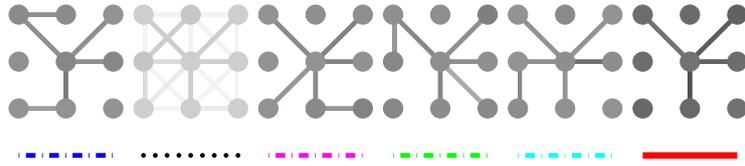
When a familiar exemplar is presented to the network (see the first four response recordings in figure 11(b)), the nodes representing the prototype and other exemplars from the same perceptual category are initially active in addition to the node representing the presented pattern. However, each exemplar causes only sustained activation of the node which has learnt weights selective to that specific pattern. The initial response of the network, thus carries information mostly about the perceptual category of the presented pattern, while later activity carries information about the stimulus identity. A similar change from global to fine information encoding has been observed for neurons in the inferior temporal cortex ([Sugase et al., 1999](#)). For both the cortical neurons and those of the model, the peak in global information transmission occurs near the peaks in the neural responses, while the peak in fine information transmission occurs once neural responses have decayed to the spontaneous or sustained activity level. This sequence of information encoding at the neural level could be related to the observation that classifications are faster and more accurate than subordinate, identity, judgements ([Perrett et al., 1998](#)).

The fifth and sixth response recordings in figure 11(b) show that a novel pattern, that is part of the same perceptual category, causes sustained activation of the node which has learnt the prototype, as does presentation of the prototype itself. The seventh response recording in figure 11(b) illustrates the response of the network to a familiar exemplar, in which the contrast of the distinguishing features has been increased relative to all the other features. In each of the last three recordings, the activity of the node that generates the sustained response is more distinct from the activity of the other nodes during the initial, transient, response phase. This would suggest that the final response to these patterns would be discernible earlier than for familiar exemplars.

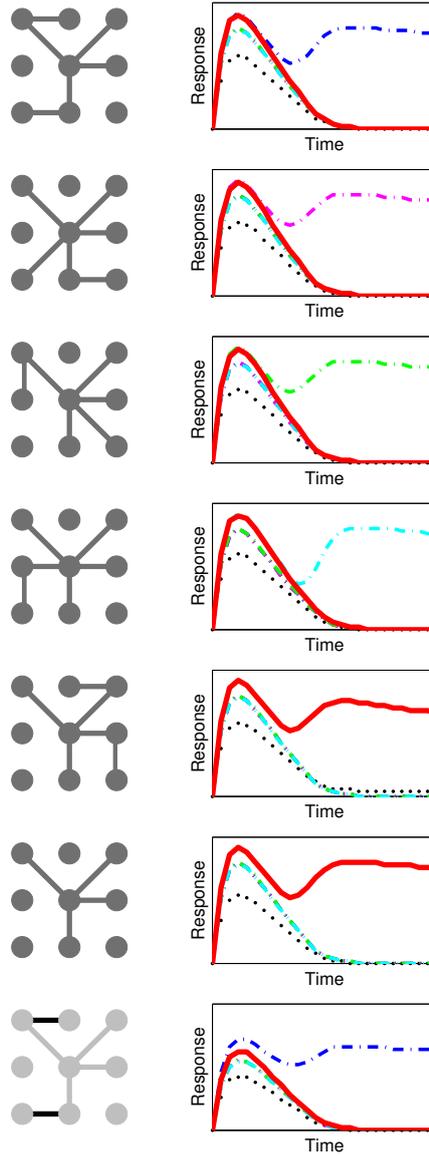
## 4 Discussion

In a perceptual processing hierarchy, like that in the visual system, neurons in higher cortical regions learn to respond to patterns of activity across the populations of neurons in the more peripheral cortical areas from which they receive their inputs. Hence, the representations learnt by neurons in lower regions can be employed by neurons in subsequent regions to build representations of more abstract or specialised features of the environment. These neurons act as ‘virtual sensors’ ([Thornton, 1996, 1997](#)) or ‘feature detectors’ ([Barlow, 1990, 1995](#)) for certain important, recurring, or behaviourally relevant stimuli.

Which stimuli become represented by individual neurons, at some level in the processing hierarchy, will depend on experience. In the model advanced in this paper, this learning process is influenced by both exogenous factors (such as the perceptual similarity between different stimuli), and endogenous factors (such as the existence of feedback signals). The model allows feedback information to bias neural activity and so affect learning. The effects of two different forms of feedback were explored. Feedback to the region in which the learning occurs was shown to bias the way in which perceptual stimuli were categorised. In contrast, targeted feedback to an earlier region could facilitate the learning of sub-ordinate level representations. Feedback can thus act both to enhance the similarity and to highlight the distinctions between stimuli. Such feedback information is assumed to originate in other cortical regions which have not been explicitly modelled. Feedback information about perceptual category membership (supplied to the upper region of the model) could come from cortical regions processing information from different sensory modalities or from regions processing linguistic or semantic information. A number of previous neural network models have found significant advantages when learning is influenced by information derived from different processing streams ([Becker and Hinton, 1992](#); [de Sa, 1994](#); [Phillips et al., 1995](#); [Balkenius, 1995](#); [Becker, 1996](#); [Kay and Phillips, 1997](#); [Phillips and Singer, 1997](#); [de Sa and Ballard, 1998](#); [Der and Smyth, 1998](#); [Körding and König, 2001](#)). Feedback information that selectively modulates the processing of lower-level image features (supplied to the lower region of the model) could come from cortical regions controlling spatial or object-based attention. Empirical evidence that is consistent with this proposal comes from experiments demonstrating that attention is required to induce learning under certain task conditions ([Karni and Bertini, 1997](#); [Sigman and Gilbert, 2000](#); [Watanabe et al., 2002](#)). It would therefore be expected that



(a)



(b)

**Figure 11:** The responses of nodes in the upper region after learning to represent the prototype and each exemplar of the training patterns shown in figure 9(a). (a) The representations learnt by each node in the upper region and the key to how the response of each node is plotted. (b) The activity generated in response to each of the test patterns shown.

the parietal and frontal system responsible for attentional control (Kanwisher and Wojciulik, 2000; Kastner and Ungerleider, 2000; Corbetta and Shulman, 2002) would be active during certain perceptual learning tasks, but would not need to be active once learning has been achieved.

Neurons respond to stimuli based on the selectivity of their synaptic weights. Hence, prior experience (recorded in the synaptic weights) affects how current perceptual information is processed (via those synaptic weights). Learning thus influences subsequent perceptual processing. The effect of prior experiences are not only evident during the initial feedforward activation of neurons: feedback enables neural activity at later stages of perceptual processing to influence the on-going processing occurring at earlier stages. In the model, such feedback has been shown to enable prior experience to influence both object segmentation and categorical perception. However, the formation of a specialised representation in the upper region of the model can enable familiar stimuli to be represented rapidly regardless of feedback; compatible with observations of rapid visual perception in humans and monkeys (Keysers et al., 2001; Fabre-Thorpe et al., 2001). The formation of specialised feature detectors is likely to enable more reliable, as well as more rapid, recognition of stimuli which might otherwise be difficult to identify based on the distributed activity of a number of neurons in more peripheral cortical areas. Rather than relying on dynamic binding (Roskies, 1999; Hummel, 2001) or top-down, attentional, processes (Reynolds and Desimone, 1999) such stimuli can be identified via feedforward information processing. In this way, the recognition process for familiar stimuli can become automatic in much the same way that the control of behaviours, that initially require deliberation, can become reactive after extended practise (Posner et al., 1997; Toates, 1998; Arbib and Liaw, 1995; Karni, 1996). In humans, extended experience with visual stimuli, leading to perceptual expertise, does result in faster and less deliberative processing (Palmeri and Gauthier, 2004; Johansen and Palmeri, 2002). Furthermore, processing becomes more holistic or configural with expertise (Palmeri and Gauthier, 2004; Hancock et al., 2000; Baker et al., 2002), for example, category decisions initially based on the properties of individual stimulus dimensions, become based on multiple stimulus dimensions after training (Johansen and Palmeri, 2002). Empirical evidence thus supports the proposition that expert performance results from the formation of more specific and specialised representations.

Categorisation requires generalisation across similar items, while identification requires discrimination among similar items. In the model, feedback can be used to guide the learning of category prototypes and to induce the learning of specific object representations. The same model can thus be used to simulate both perceptual categorisation and object recognition and suggests that the same representational and learning mechanisms underly both these perceptual processes. The proposed model thus combines aspects of both prototype and exemplar based theories of perceptual categorisation (Palmeri and Gauthier, 2004). Furthermore, prior to the learning of more specialised representations (either prototypes or exemplars), the representations of individual features at a lower-level in the hierarchy could be used to make rule-like category decisions based on the properties of individual stimulus dimensions (Johansen and Palmeri, 2002). The proposed model thus combines more abstract and specific object representations in common with several other theories (Palmeri and Gauthier, 2004; Johansen and Palmeri, 2002). However, the proposed model is not designed to provide an account of conceptual learning. Many conceptual categories are not based on perceptual similarity and can not be learnt with the current model. Nodes in the model learn to represent recurring *conjunctions* of inputs and are thus incapable of forming category representations for stimuli that do not overlap in the input space (*i.e.*, for sets of stimuli that do not have a prototype). In order to assign such perceptually dis-similar stimuli to the same category it would be necessary for the model to also be capable of learning *disjunctions*. Similarly, a more complete model of object recognition would also require disjunctive (as well as conjunctive) learning in order to form representations that are invariant to viewpoint (Fukushima, 1980, 1988; Riesenhuber and Poggio, 2000, 1999b; Spratling, 2005)

Traditional theories of perception have assumed a purely serial, feedforward, sequence of processing stages (Marr, 1982; Hubel and Wiesel, 1977). Such “feedforward” theories thus presume that low-level information is processed prior to high-level information (Driver et al., 2001). In contrast, current theories stress the importance of cortical region interactions, in both the feedforward and feedback directions (Lamme et al., 1998; Lamme and Roelfsema, 2000; Friston and Price, 2001; Bressler, 2002; Hochstein and Ahissar, 2002). Consistent with this view is the observation that not only are concepts influenced by perceptual representations (feedforward) but also percepts are influenced by concepts (feedback) (Schyns et al., 1998; Goldstone et al., 2000, 2001). This article has explored the effects of cortical feedback connections on both neural activity, and learning. We have demonstrated that feedback enables prior learning to influence subsequent perceptual processing, and that feedback can influence perceptual learning determining what stimuli become represented and how stimuli are categorised.

## Appendix: Implementation Details

In order to learn reciprocal connection weights between the different neural regions, it was necessary to use different activation functions and learning rules for the feedforward and feedback connections. A node in the upper region learns to become selective for a combination of co-active lower-level nodes. Hence, the basal dendrites need to learn to represent a *conjunction* of pre-synaptic inputs. In contrast, each node in the lower region needs to increase the strength of the feedback connection from the single upper-level node that is responding to the pattern of which this node forms a part. Since each lower-level node may be a part of several distinct input patterns, the apical dendrites need to form connections with multiple, non-coactive, upper region nodes. Hence, the apical dendrites need to learn to represent a *disjunction* of pre-synaptic inputs. In order to respond to a conjunction of inputs, a standard weighted sum of pre-synaptic activation values can be used. Such a function will cause the activation of the dendrite to be a maximum when all the pre-synaptic features, to which it responds, are simultaneously active. In contrast, to respond to a disjunction of inputs, a function can be used which causes the output to depend on the maximum input activity (Riesenhuber and Poggio, 1999a,b; Spratling, 2005).

### A.1 Activation

For each node the activations of the apical and basal dendrites were calculated as:

$$y_{jk,apical}^t = \max_{i=1}^l \{Z_{ijk}^t\} \quad (1)$$

$$y_{jk,basal}^t = \sum_{i=1}^m (w_{ijk} X_{ijk}^t) \quad (2)$$

Where  $y_{jk,apical}^t$  is the activation of the apical dendrite of node  $j$  in region  $k$  at time  $t$ ,  $y_{jk,basal}^t$  is the activation of the basal dendrite of the same node,  $l$  is the total number of synapses on the apical dendrite,  $m$  is the total number of synapses on the basal dendrite,  $w_{ijk}$  is the synaptic weight from input  $i$  to the basal dendrite of node  $j$  in region  $k$ ,  $Z_{ijk}^t$  is the normalised weighted activation received at input  $i$  to the apical dendrite of node  $j$  in region  $k$  and  $X_{ijk}^t$  is the activation received by the basal dendrite of node  $j$  in region  $k$  from input  $i$  after pre-integration lateral inhibition<sup>1</sup>:

$$Z_{ijk}^t = z_{ijk}^t \left( \frac{v_{ijk}}{\max_{q=1}^l \{v_{qjk}\}} \frac{v_{ijk}}{\max_{q=1}^n \{v_{iqk}\}} \right) \quad (3)$$

$$X_{ijk}^t = x_{ijk}^t \left( 1 - \alpha^t \max_{\substack{p=1 \\ (p \neq j)}}^n \left\{ \frac{w_{ipk}}{\max_{q=1}^m \{w_{qp k}\}} \frac{y_{pk}^{t-1}}{\max_{q=1}^n \{y_{qk}^{t-1}\}} \right\} \right)^+ \quad (4)$$

Where  $z_{ijk}^t$  is the input activity received at the apical dendrite of node  $j$  in region  $k$ ,  $v_{ijk}$  is the synaptic weight from input  $i$  to the apical dendrite of node  $j$  in region  $k$ ,  $x_{ijk}^t$  is the input activity received at the basal dendrite of node  $j$  in region  $k$ ,  $\alpha^t$  is a scale factor controlling the strength of lateral inhibition,  $y_{pk}^{t-1}$  is the activation of node  $p$  in region  $k$  at time  $t - 1$  (defined below in equation 8), and  $(z)^+$  is the positive half-rectified value of  $z$  (i.e.,  $(z)^+ = z$  if  $z$  is positive,  $(z)^+ = 0$  otherwise). In the reported simulations, the value of  $\alpha^t$  was gradually increased at each time-step, from an initial value of zero to a maximum value of five in steps of 0.1.

The value of  $Z_{ijk}^t$  depends upon the strength of pre-synaptic activity and the strength of the weight received from that input ( $i$ ). This value is adjusted using both post- and pre-synaptic weight normalisation. Such normalisation causes the value of  $Z_{ijk}^t$  to be reduced if the node receives a stronger connection from another input, or if another node receives a stronger connection from input  $i$ . There is thus implicit competition which, through activity-dependent learning (see section A.2), causes each node to become selective for a disjunctive set of pre-synaptic inputs. Specifically, if more than one upper-level node is active, the feedback connection from only one of those nodes will be strengthened for each lower-level node.

The value of  $X_{ijk}^t$  depends upon the strength of pre-synaptic activity and the strength of the lateral inhibition directed towards that particular input. The strength of lateral inhibition is normalised such that (1) the lateral weight is divided by the maximum lateral weight originating from the inhibiting node and (2) the inhibiting node's activity is divided by the maximum activity of all the nodes in the network. This normalisation ensures (1) that lateral inhibition is effective at the start of training when weights have small, undifferentiated, values and (2)

<sup>1</sup>For full details of the implementation of pre-integration lateral inhibition, and for a justification of this scheme on both biological and computational grounds, see Spratling and Johnson (2001, 2002, 2004b).

that lateral inhibition remains effective during the competition process (*i.e.*, for all values of  $t$ ) when nodes inhibit each other and hence have reduced activity values. The value of  $X_{ijk}^t$  will be strongly inhibited if another node ( $p$ ) is strongly activated by the overall stimulus (*i.e.*, if  $y_{pk}^{t-1}$  has a high value relative to all other node activations) and that other node receives a strong synaptic weight from input  $i$  (*i.e.*, if  $w_{ipk}$  has a high value relative to all the other weights received by node  $p$ ). Hence, this form of lateral inhibition provides competition by enabling each node to ‘block’ its preferred inputs from activating the basal dendrites of other nodes. There is thus strong competition which causes nodes to become selective for distinct conjunctive sets of inputs, but which does not prevent multiple nodes from responding simultaneously to the presentation of multiple, distinct, stimuli.

The values of the apical inputs ( $z_{ijk}^t$ ) are the activations of nodes in the higher cortical region at the previous time step (*i.e.*,  $y_{j'k+1}^{t-1}$ ), or are top-down signals that arise from external inputs to the network. The values of the basal inputs ( $x_{ijk}^t$ ) are the activations of nodes in the lower cortical region at the previous time step (*i.e.*,  $y_{j'k-1}^{t-1}$ ) or are external, sensory, inputs supplied to the network. Note that the activations of every node (in both regions) are determined in an identical way. Furthermore, external sources of input are treated in the same way as inputs arising from neural activity generated within the modelled circuits. Hence, external sources of feedback (or sensory input) provide another activation value,  $z$  (or  $x$ ), that contributes to the apical (or basal) activation via a synaptic weight,  $v$  (or  $w$ ).

The activation of the apical dendrite multiplicatively modulates the activation of the basal dendrite in order to determine the output activation of each node:

$$y_{jk}^t = y_{jk,basal}^t (1 + y_{jk,apical}^t) \quad (5)$$

Where  $y_{jk}^t$  is the activation of node  $j$  in region  $k$  at time  $t$ . This formulation enables bottom-up, sensory-driven, stimulation to drive the response of the node even in the absence of top-down activity. In contrast, feedback activation can not drive the node’s activity in the absence of feedforward activation, but it can amplify any response to feedforward stimulation. This is consistent with the widely held view that feedback signals act to modulate responses that are primarily driven by feedforward inputs (Crick and Koch, 1998; Hupé et al., 1998; Friston and Büchel, 2000). Furthermore, this formulation also provides a simple, first-order, approximation to the form of gain modulation, caused by stimulation of the apical dendrite, which is observed in cortical pyramidal cells (Larkum et al., 2004).

The presence of reciprocal excitatory connections can lead to positive feedback effects resulting in run-away activation values. To prevent this, the activity of each node is attenuated in proportion to the cumulative strength of its previous activity ( $C_{jk}^{t-1}$ ), and this attenuated activation is clipped to be in the range [0,1]:

$$y_{jk}^t = \left[ \frac{y_{jk}^t}{1 + C_{jk}^{t-1}} \right]_0^1 \quad (6)$$

Where the cumulative activity,  $C_{jk}^t$ , of the node is calculated as:

$$C_{jk}^t = \tau_c y_{jk}^t + (1 - \tau_c) C_{jk}^{t-1} \quad (7)$$

Where  $\tau_c = 0.1$  is a time constant. This attenuated activity has a time-varying profile which resembles that of biological neurons; having an initial burst of activity followed by a sustained response at a lower firing rate. The activity of each node was also modified by a small amount of noise, such that:

$$y_{jk}^t = y_{jk}^t (1 + \rho) \quad (8)$$

The noise values,  $\rho$ , were logarithmically distributed positive real numbers in the range [0,0.01]. Since the magnitude of the noise is small it has very little effect on neural activity except when multiple nodes have virtually identical synaptic weights. When this occurs, the noise causes one of these nodes to win the competition to be active in response to the current stimulus. Finally, to make the plotted response profiles more realistic the change in activity of each node was smoothed by taking into account its previous activity:

$$y_{jk}^t = \tau_y y_{jk}^t + (1 - \tau_y) y_{jk}^{t-1}$$

Where  $\tau_y = 0.4$  is a time constant.

The nodes in a neural network generate activity in response to the input they receive. This response can be considered to be a representation of the input stimulus. How distinctly two different stimuli are represented can be measured by comparing the patterns of activity generated in response to these different stimuli. This comparison was made by calculating the mean of the absolute difference in the activity values generated by the two stimuli:

$$\epsilon_{ab} = \frac{\sum_{j=1}^n \| {}^a y_{jk}^t - {}^b y_{jk}^t \|}{n}$$

where  ${}^a y_{jk}^t$  and  ${}^b y_{jk}^t$  are the activity values for a node in response to stimuli  $a$  and  $b$  respectively. For these calculations the activity at a value of  $t$  when  $\alpha^t = 2$  were used. This measure provides an indication of how sensitive the network is at detecting the difference between two stimuli. To determine to what extent prior learning has affected the sensitivity of the network, such a comparison between the neural response patterns needs to be performed before and after training:

$$\Delta\epsilon_{ab} = \frac{\epsilon_{ab}^T - \epsilon_{ab}^U}{\epsilon_{ab}^T + \epsilon_{ab}^U} \quad (9)$$

where  $\epsilon_{ab}^U$  and  $\epsilon_{ab}^T$  are the sensitivity values before and after training respectively. This equation was used to generate the results shown in figure 7.

## A.2 Learning

All the undefined synaptic weights targeting each dendrite were initially given equal values. Synaptic weights were modified using the final activation values found after iteration of the equations given above (*i.e.*, for values of  $t$  when  $\alpha^t = 5$ ). The  $t$  superscript is thus dropped from subsequent equations. For the apical dendrites, the following learning rule was employed:

$$\Delta v_{ijk} = \pm \frac{\gamma z_{ijk}}{\sum_{p=1}^n y_{pk}} (y_{jk} - \bar{y}_k)^+ \quad (10)$$

Where  $\bar{y}_k$  is the mean of the output activations (*i.e.*,  $\bar{y}_k = \frac{1}{n} \sum_{j=1}^n y_{jk}$ ), and  $\gamma$  is a parameter controlling the learning rate ( $\gamma = \frac{1}{4}$  was used in all the simulations presented in this text). Learning only occurs for nodes that were more active than average at the previous iteration, and at synapses with currently active inputs. This learning rule has a positive value at synapses where  $Z_{ijk} = y_{jk}$  and a negative value otherwise. Hence, only the weight of the most active input was increased, and weights to all other active inputs were decreased. This learning rule thus encourages each node to learn weights selective for a set of non-coactive inputs. This is achieved since when a node is more active than average it increases its synaptic weights to a single active input and decreases its weights to all other active inputs. Hence, only sets of upper-level nodes which are seldom coactive will generate strong feedback weights to the same lower-level node. Following learning, synaptic weights were clipped at zero such that  $w_{ijk} = (w_{ijk})^+$  and were normalised such that  $\sum_{j=1}^n w_{ijk} = 1$ .

For basal dendrites, the following learning rule was employed:

$$\Delta w_{ijk} = \frac{\beta (x_{ijk} - \bar{x}_{jk})}{\sum_{p=1}^m x_{pjk}} (y_{jk} - \bar{y}_k)^+ \quad (11)$$

Where  $\bar{x}_{jk}$  is the mean of the input activations (*i.e.*,  $\bar{x}_{jk} = \frac{1}{m} \sum_{i=1}^m x_{ijk}$ ), and  $\beta$  is a parameter controlling the learning rate ( $\beta = \frac{1}{2}$  was used in all the simulations presented in this text). Following learning, synaptic weights were clipped at zero such that  $w_{ijk} = (w_{ijk})^+$  and were normalised such that  $\sum_{i=1}^m w_{ijk}^+ = 1$ . Synaptic weights that reached a value of zero became unmodifiable in future learning (*i.e.*, if  $w_{ijk} = 0$  then this weight was fixed at zero). This learning rule encourages each node to learn weights selective for a set of coactive inputs. This is achieved since when a node is more active than average it increases its synaptic weights to active inputs and decreases its weights to inactive inputs. Hence, only sets of inputs which are consistently coactive will generate strong afferent weights. In addition, the learning rule is designed to ensure that nodes can represent stimuli which share input features in common (*i.e.*, to allow the network to represent overlapping patterns). This is achieved by rectifying the post-synaptic term of the rule so that no weight changes occur when the node is less active than average. If learning was not restricted in this way, whenever a pattern was presented, all nodes which represented patterns with overlapping features would reduce their weights to these features.

The learning rule for the feedforward connections enables neurons to become selective to re-occurring patterns of activity within the training data they are exposed to. The learning rule for the feedback connections results in corresponding feedforward and feedback weights having similar strengths. These learning rules are not entirely biologically plausible, since both rules employ one non-local variable: the sum of all node activations. This value is used to calculate the value of the denominator in equation 10 and to calculate the value of mean output activation ( $\bar{y}_k$ ) used as a threshold in both rules.

In all the experiments described in section 3, the network was trained for 200 iterations, with different, randomly chosen, inputs being presented to the network at each cycle. The one exception, is that the results shown in figure 7 were produced using 500 training cycles. Results were not particularly sensitive to the chosen parameter values, hence, all the simulations reported in section 3 were performed using the same parameter values.

### A.3 Comparison with Other Models

In common with many other neural network models, the algorithm presented here employs a mechanism of competition between the nodes in a region. In most competitive learning algorithms, nodes attempt to ‘block’ other nodes from generating an *output* in response to the current stimulus. A node’s success in this competition is dependent on the total strength of the stimulation it receives and nodes which compete unsuccessfully have their output activity suppressed. This form of competition can be implemented either by using a selection process which chooses the ‘winning’ node(s) (e.g., Rumelhart and Zipser, 1985; Kohonen, 1997; Grossberg, 1987; O’Reilly and Munakata, 2000) or by using lateral inhibitory connections between nodes (e.g., Földiák, 1990; Marshall, 1995; von der Malsburg, 1973; Oja, 1989). Such algorithms require an *a priori* choice to be made as to whether the training data will be represented using a local or distributed code (Spratling and Johnson, 2002, 2004b).

In contrast, the mechanism for generating competition in the architecture used here causes each node to attempt to ‘block’ its preferred *inputs* from activating other nodes. Hence, if a node is strongly activated by the stimulus, and it has a strong synaptic weight to a certain feature of that stimulus, then it will inhibit other nodes from responding to that feature. On the other hand, if an active node receives a weak synapse from a certain input, then it will weakly inhibit other nodes from responding to that specific input. Thus, if individual nodes have learnt weights that are selective to certain stimuli then when multiple stimuli are simultaneously presented to the network each of the nodes representing one of these stimuli can be simultaneously active. A network using this form of lateral inhibition can thus respond appropriately to any combination of input patterns. Specifically, it is capable of learning, and generating, both local and distributed codes (Spratling and Johnson, 2002, 2004b).

At first glance, the proposed architecture appears to have some superficial similarity with neural networks that implement adaptive resonance theory (ART) (Grossberg, 1986). However, in the present model feedback modulates neural activity, while in ART networks the reciprocal feedforward and feedback connections are designed to mediate a matching process in order to locate the stored representation that is most similar to the input pattern. In the algorithm proposed here, this matching process is achieved by using an improved model of competition within each region. This mechanism of competition enables a population of nodes to deal naturally with stimuli which share sub-features in common (*i.e.*, patterns which overlap in the input space). Following competition only the best matching node(s) remain active even when there is considerable overlap between stored representations (Spratling and Johnson, 2002, 2003). Hence, in this model the stored representations that match the current input are found via competition within a single region, rather than via an iterative procedure of competition, comparison and resetting involving multiple neural populations as used in ART.

The proposed mechanism of competition ensures the accurate activation of stored representations by matching input patterns. However, when novel or poorly represented patterns are presented, the competition mechanism results in the activation of a previously unallocated node. The maximum value of the parameter  $\alpha$ , *i.e.*, the maximum strength of lateral inhibition, determines the degree of mismatch required between a new pattern and all the stored patterns before a previously unallocated node wins the competition. Since this parameter value was kept constant in all the experiments reported here, whether the network learnt to represent patterns as individual exemplars or as prototypes depended purely on the properties of the inputs to the network (*i.e.*, the structure of the training patterns and any top-down information supplied to the network) rather than changes to parameter values. Since learning is activity-dependent, previously learnt representations remain stable (as they are only activated by matching input patterns), while new patterns can be represented by new nodes. The stability/plasticity dilemma is thus solved by an improved mechanism for competition (within a single layer of nodes) rather than the use of a separate novelty detection mechanism as employed in ART.

In addition to the intra-regional competition, described above, the present model also incorporates mechanisms for inter-regional cooperation: excitatory feedforward and feedback connections between distinct neural populations. Reciprocal connectivity between separate neural modules is a common feature of many algorithms. However in this particular model, nodes have two functionally distinct dendritic compartments in which separate sets of feedforward and feedback inputs are integrated independently. Feedforward and feedback stimulation have distinct effects on neural activity and thereby learning. Specifically, the response of the node is driven by bottom-up activation and modulated by top-down activity.

In contrast, other models that incorporate reciprocal connections between neural populations have proposed that feedforward and feedback activations have the same, additive, effect on node activity (e.g., McClelland and Rumelhart, 1981; O’Reilly and Farah, 1999; Humphreys and Forde, 2001). Such models thus make no functional distinction between feedforward and feedback connections. Other algorithms have proposed distinct functional roles for feedforward and feedback connections, but have only allowed feedback to affect learning rather than node activation (e.g., Rumelhart et al., 1986; Körding and König, 2000). Neural architectures that do assign distinct roles to feedforward and feedback connections and which also allow feedback (as well as feedforward) connections to influence activity have mostly suggested that top-down information has a suppressive rather than a modulatory

effect (e.g., Mumford, 1994; Barlow, 1994; Rao, 1999; Rao and Ballard, 1997, 1999; Koerner et al., 1997). Hence, few previous models have proposed that feedback (or contextual) information should have a modulatory affect on neural activation (e.g., Grossberg and Raizada, 2000; Salinas and Abbott, 1996; Salinas and Thier, 2000; Salinas and Sejnowski, 2001; Phillips and Singer, 1997; Phillips et al., 1995). Each of these previous models have been implemented using conventional, one-dendrite, neurons and none have employed the same form of modulation as proposed here. The proposed model is thus distinct in using nodes with two functionally distinct dendritic compartments to allow feedback to modulate neural activity.

A number of other artificial neural network algorithms have employed neurons with functionally distinct clusters of synapses within which nonlinear operations are performed (e.g., Rumelhart et al., 1986; Mel and Koch, 1990; Durbin and Rumelhart, 1989; Spratling and Hayes, 2000). However, these models make no distinction between feedforward and feedback connections. The model presented in this paper thus exploits a different way in which computationally distinct dendritic compartments can contribute to the computational properties of neurons: by enabling different sets of synaptic inputs to have distinct functional roles.

To conclude, the computational model employed in this paper is distinct from previous algorithms in using a novel method for intra-regional competition and a novel method of inter-regional interaction.

## Acknowledgements

This work was funded by MRC Grant number G9715587 and EPSRC Grant number GR/S81339/01.

## References

- Arbib, M. A. and Liaw, J.-S. (1995). Sensorimotor transformations in the worlds of frogs and robots. *Artificial Intelligence*, 72:53–80.
- Baker, C. I., Behrmann, M., and Olson, C. R. (2002). Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. *Nature Neuroscience*, 5(11):1210–6.
- Balkenius, C. (1995). Multi-modal sensing for robot control. In Niklasson, L. F. and Bodén, M. B., editors, *Current trends in connectionism*, pages 203–16. Lawrence Erlbaum, Hillsdale, NJ.
- Barlow, H. B. (1990). Conditions for versatile learning, Helmholtz’s unconscious inference, and the task of perception. *Vision Research*, 30:1561–71.
- Barlow, H. B. (1994). What is the computational goal of the neocortex? In Koch, C. and Davis, J. L., editors, *Large-Scale Neuronal Theories of the Brain*, chapter 1. MIT Press, Cambridge, MA.
- Barlow, H. B. (1995). The neuron doctrine in perception. In Gazzaniga, M. S., editor, *The Cognitive Neurosciences*, chapter 26. MIT Press, Cambridge, MA.
- Becker, S. (1996). Mutual information maximization: models of cortical self-organization. *Network: Computation in Neural Systems*, 7:7–31.
- Becker, S. and Hinton, G. E. (1992). A self-organizing neural network that discovers surfaces in random-dot stereograms. *Nature*, 355(6356):161–3.
- Bravo, M. J. and Farid, H. (2003). Object segmentation by top-down processes. *Visual Cognition*, 10(4):471–91.
- Bressler, S. L. (2002). Understanding cognition through large-scale cortical networks. *Current Directions in Psychological Science*, 11(2):58–61.
- Budd, J. M. L. (1998). Extrastriate feedback to primary visual cortex in primates: a quantitative analysis of connectivity. *Proceedings of the Royal Society of London. Series B*, 265(1400):1037–44.
- Bullier, J. and Nowak, L. G. (1995). Parallel versus serial processing: new vistas on the distributed organization of the visual system. *Current Opinion in Neurobiology*, 5(4):497–503.
- Chun, M. M. (2002). Contextual cueing of visual attention. *Trends in Cognitive Sciences*, 4(5):170–8.
- Corbetta, M. and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, 3(3):201–15.
- Crick, F. and Koch, C. (1998). Constraints on cortical and thalamic projections: the no-strong-loops hypothesis. *Nature*, 391:245–50.
- Damper, R. I. and Harnad, S. R. (2000). Neural network models of categorical perception. *Perception and Psychophysics*, 62(4):843–67.
- de Sa, V. R. (1994). Learning classification with unlabeled data. In Cowan, J. D., Tesauro, G., and Alspector, J., editors, *Advances in Neural Information Processing Systems 6*, pages 112–9, San Francisco, CA. Morgan Kaufmann.
- de Sa, V. R. and Ballard, D. (1998). Category learning through multi-modality sensing. *Neural Computation*, 10(5):1097–117.

- Der, R. and Smyth, D. (1998). Local online learning of coherent information. *Neural Networks*, 11(5):909–25.
- Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proceedings of the National Academy of Sciences USA*, 93:13494–9.
- Desimone, R. and Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18:193–222.
- Driver, J., Davis, G., Russell, C., Turatto, M., and Freeman, E. (2001). Segmentation, attention and phenomenal visual objects. *Cognition*, 80(1-2):61–95.
- Durbin, R. and Rumelhart, D. E. (1989). Product units: a computationally powerful and biologically plausible extension to backpropagation networks. *Neural Computation*, 1:133–42.
- Ebdon, M. (1992). The uniformity of cerebral neocortex and its implications for cognitive science. Technical Report CSRP-228, School of Cognitive and Computing Sciences, University of Sussex.
- Fabre-Thorpe, M., Delorme, A., Marlot, C., and Thorpe, S. (2001). A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. *Journal of Cognitive Neuroscience*, 13:171–80.
- Felleman, D. J. and Van Essen, D. C. (1991). Distributed hierarchical processing in primate cerebral cortex. *Cerebral Cortex*, 1:1–47.
- Földiák, P. (1990). Forming sparse representations by local anti-Hebbian learning. *Biological Cybernetics*, 64:165–70.
- Friston, K. J. and Büchel, C. (2000). Attentional modulations of effective connectivity from V2 to V5/MT in humans. *Proceedings of the National Academy of Sciences USA*, 97(13):7591–6.
- Friston, K. J. and Price, C. J. (2001). Dynamic representations and generative models of brain function. *Brain Research Bulletin*, 54(3):275–85.
- Fukushima, K. (1980). Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4):193–202.
- Fukushima, K. (1988). Neocognitron: a hierarchical neural network capable of visual pattern recognition. *Neural Networks*, 1(2):119–30.
- Gauthier, I., Skudlarski, P., Gore, J. C., and Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, 3(2):191–7.
- Gilbert, C. D. (1996). Plasticity in visual perception and physiology. *Current Opinion in Neurobiology*, 6(2):269–74.
- Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology*, 49:585–612.
- Goldstone, R. L. and Barsalou, L. (1998). Reuniting perception and conception. *Cognition*, 65:231–62.
- Goldstone, R. L., Lippa, Y., and Shiffrin, R. M. (2001). Altering object representations through category learning. *Cognition*, 78:27–43.
- Goldstone, R. L., Steyvers, M., Spencer-Smith, J., and Kersten, A. (2000). Interactions between perceptual and conceptual learning. In Diettrich, E. and Markman, A. B., editors, *Cognitive Dynamics: Conceptual Change in Humans and Machines*, pages 191–228. Lawrence Erlbaum Associates, Mahwah, NJ.
- Grossberg, S. (1986). The adaptive self-organisation of serial order in behaviour: speech, language, and motor control. In Schwab, E. C. and Nusbaum, H. C., editors, *Pattern Recognition by Humans and Machines, volume 1: Speech Recognition*, London, UK. Academic Press.
- Grossberg, S. (1987). Competitive learning: from interactive activation to adaptive resonance. *Cognitive Science*, 11:23–63.
- Grossberg, S. and Raizada, R. (2000). Contrast-sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex. *Vision Research*, 40(10-12):1413–32.
- Hancock, P. J. B., Bruce, V., and Burton, A. M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences*, 4(9):330–7.
- Harnad, S. (1987). Psychophysical and cognitive aspects of categorical perception: a critical overview. In Harnad, S., editor, *Categorical Perception; The Groundwork of Cognition*, chapter 1. Cambridge University Press, Cambridge, UK.
- Harnad, S. (2003). Categorical perception. In *Encyclopedia of Cognitive Science*. Macmillan: Nature Publishing Group.
- Hasegawa, I. and Miyashita, Y. (2002). Categorizing the world: expert neurons look into key features. *Nature Neuroscience*, 5(2):90–1.
- Häusser, M. and Mel, B. (2003). Dendrites: bug or feature. *Current Opinion in Neurobiology*, 13(3):372–83.
- Hochstein, S. and Ahissar, M. (2002). View from the top: hierarchies and reverse hierarchies in the visual system. *Neuron*, 36(5):791–804.
- Hubel, D. H. and Wiesel, T. N. (1977). Functional architecture of macaque monkey visual cortex. *Proceedings of the Royal Society of London. Series B*, 198:1–59.
- Hummel, J. E. (2001). Complementary solution to the binding problems in vision: implications for shape percep-

- tion and object recognition. *Visual Cognition*, 8:489–517.
- Humphreys, G. W. and Forde, E. M. E. (2001). Hierarchies, similarity, and interactivity in object recognition: ‘category-specific’ neuropsychological deficits. *Behavioral and Brain Sciences*, 24(3):453–76.
- Hupé, J. M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P., and Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature*, 394(6695):784–7.
- Johansen, M. K. and Palmeri, T. J. (2002). Are there representational shifts during category learning? *Cognitive Psychology*, 45:482–553.
- Kanwisher, N. and Wojciulik, E. (2000). Visual attention: insights from brain imaging. *Nature Reviews Neuroscience*, 1(2):91–100.
- Karni, A. (1996). The acquisition of perceptual and motor skills: a memory system in the adult human cortex. *Cognitive Brain Research*, 5:39–48.
- Karni, A. and Bertini, G. (1997). Learning perceptual skills: behavioral probes into adult cortical plasticity. *Current Opinion in Neurobiology*, 7(4):530–5.
- Kastner, S. and Ungerleider, L. G. (2000). Mechanisms of visual attention in the human cortex. *Annual Review of Neuroscience*, 23:315–41.
- Kay, J. and Phillips, W. A. (1997). Activation functions, computational goals and learning rules for local processors with contextual guidance. *Neural Computation*, 9(4):895–910.
- Keyser, C. and Perrett, D. I. (2002). Visual masking and RSVP reveal neural competition. *Trends in Cognitive Sciences*, 6(3):120–5.
- Keyser, C., Xiao, D. K., Földiák, P., and Perrett, D. I. (2001). The speed of sight. *Journal of Cognitive Neuroscience*, 13(1):90–101.
- Kobatake, E., Wang, G., and Tanaka, K. (1998). Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys. *Journal of Neurophysiology*, 80(1):324–30.
- Koerner, E., Tsujino, H., and Masutani, T. (1997). A cortical-type modular neural network for hypothetical reasoning. *Neural Networks*, 10(5):791–814.
- Kohonen, T. (1997). *Self-Organizing Maps*. Springer-Verlag, Berlin.
- Körding, K. P. and König, P. (2000). Learning with two sites of synaptic integration. *Network: Computation in Neural Systems*, 11(1):25–39.
- Körding, K. P. and König, P. (2001). A spike based learning rule for generation of invariant representations. *Journal of Physiology (Paris)*, 94(5–6):539–48.
- Lamme, V. A. F. (1995). The neurophysiology of figure-ground segregation in primary visual cortex. *Journal of Neuroscience*, 15(2):1605–15.
- Lamme, V. A. F. and Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, 23(11):571–9.
- Lamme, V. A. F., Supér, H., and Spekreijse, H. (1998). Feedforward, horizontal, and feedback processing in the visual cortex. *Current Opinion in Neurobiology*, 8(4):529–35.
- Larkum, M. E., Senn, W., and Lüscher, H.-R. (2004). Top-down dendritic input increases the gain of layer 5 pyramidal neurons. *Cerebral Cortex*, 14(10):1059–70.
- Larkum, M. E., Zhu, J. J., and Sakmann, B. (1999). A new cellular mechanism for coupling inputs arriving at different cortical layers. *Nature*, 398(6725):338–41.
- Lee, T. S., Mumford, D., Romero, R., and Lamme, V. A. F. (1998). The role of primary visual cortex in higher level vision. *Vision Research*, 38(15–16):2429–54.
- Lee, T. S., Yang, C. F., Romero, R. D., and Mumford, D. (2002). Neural activity in early visual cortex reflects behavioral experience and higher-order perceptual saliency. *Nature Neuroscience*, 5(6):589–97.
- Logothetis, N. (1998). Object vision and visual awareness. *Current Opinion in Neurobiology*, 8(4):536–44.
- Luck, S. J., Chelazzi, L., Hillyard, S. A., and Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *Journal of Neurophysiology*, 77:24–42.
- Marr, D. (1982). *Vision*. Freeman, San Francisco, CA.
- Marshall, J. A. (1995). Adaptive perceptual pattern recognition by self-organizing neural networks: context, uncertainty, multiplicity, and scale. *Neural Networks*, 8(3):335–62.
- McAdams, C. J. and Maunsell, J. H. R. (2000). Attention to both space and feature modulates neuronal responses in macaque area V4. *Journal of Neurophysiology*, 83(3):1751–5.
- McClelland, J. L. and Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception, part I: an account of basic findings. *Psychological Review*, 88:375–407.
- Mehta, A. D., Ulbert, I., and Schroeder, C. E. (2000). Intermodal selective attention in monkeys. II: physiological mechanisms of modulation. *Cerebral Cortex*, 10:359–70.
- Mel, B. W. and Koch, C. (1990). Sigma-pi learning: on radial basis functions and cortical associative learning. In Touretzky, D. S., editor, *Advances in Neural Information Processing Systems 2*, pages 474–81, San Francisco,

- CA. Morgan Kaufmann.
- Mountcastle, V. B. (1998). *Perceptual Neuroscience: The Cerebral Cortex*. Harvard University Press, Cambridge, MA.
- Mumford, D. (1994). Neuronal architectures for pattern-theoretic problems. In Koch, C. and Davis, J. L., editors, *Large-Scale Neuronal Theories of the Brain*, pages 125–52. MIT Press, Cambridge, MA.
- Needham, A. (2001). Object recognition and object segregation in 4.5-month-old infants. *Journal of Experimental Child Psychology*, 78(1):3–22.
- Oja, E. (1989). Neural networks, principle components, and subspaces. *International Journal of Neural Systems*, 1:61–8.
- Olson, C. R. (2001). Object-based vision and attention in primates. *Current Opinion in Neurobiology*, 11(2):171–9.
- Olson, I. R., Chun, M. M., and Allison, T. (2001). The contextual guidance of attention: human intracranial event-related potential evidence for feedback modulation in anatomically early temporally late stages of visual processing. *Brain*, 124(7):1417–25.
- O’Reilly, R. C. (1998). Six principles for biologically based computational models of cortical cognition. *Trends in Cognitive Sciences*, 2(11):455–62.
- O’Reilly, R. C. and Farah, M. J. (1999). Simulation and explanation in neuropsychology and beyond. *Cognitive Neuropsychology*, 16:49–72.
- O’Reilly, R. C. and Munakata, Y. (2000). *Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain*. MIT Press, Cambridge, MA.
- Palmeri, T. J. and Gauthier, I. (2004). Visual object understanding. *Nature Reviews Neuroscience*, 5(4):291–303.
- Paré, D., Lang, E. J., and Destexhe, A. (1998). Inhibitory control of somatodendritic interactions underlying action potentials in neocortical pyramidal neurons in vivo: an intracellular and computational study. *Neuroscience*, 84(2):377–402.
- Perrett, D. I., Oram, M. W., and Ashbridge, E. (1998). Evidence accumulation in cell populations responsive to faces: an account of generalisation of recognition without mental transformations. *Cognition*, 67:111–45.
- Peterson, M. A. and Gibson, B. S. (1993). Shape recognition inputs to figure-ground organisation in three-dimensional displays. *Cognitive Psychology*, 25:383–429.
- Pevtzow, R. and Goldstone, R. L. (1994). Categorization and the parsing of objects. In *Proceedings of the 16th Annual Conference of the Cognitive Science Society*, pages 717–22, Hillsdale, NJ. Lawrence Erlbaum Associates.
- Phillips, W. A., Kay, J., and Smyth, D. (1995). The discovery of structure by multi-stream networks of local processors with contextual guidance. *Network: Computation in Neural Systems*, 6(2):225–46.
- Phillips, W. A. and Singer, W. (1997). In search of common foundations for cortical computation. *Behavioural and Brain Sciences*, 20(4):657–722.
- Posner, M. I., DiGirolamo, G. J., and Fernandez-Duque, D. (1997). Brain mechanisms of cognitive skills. *Consciousness and Cognition*, 6:267–90.
- Quinn, P. C. and Johnson, M. H. (2000). Global-before-basic object categorization in connectionist networks and 2-month-old infants. *Infancy*, 1(1):32–46.
- Rao, R. P. N. (1999). An optimal estimation approach to visual perception and learning. *Vision Research*, 39(11):1963–89.
- Rao, R. P. N. and Ballard, D. H. (1997). Dynamical model of visual recognition predicts neural response properties in the visual cortex. *Neural Computation*, 9(4):721–63.
- Rao, R. P. N. and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1):79–87.
- Reynolds, J. H. and Desimone, R. (1999). The role of neural mechanisms of attention in solving the binding problem. *Neuron*, 24(1):19–29.
- Reynolds, J. H., Pasternak, T., and Desimone, R. (2000). Attention increases sensitivity of V4 neurons. *Neuron*, 26:703–14.
- Riesenhuber, M. and Poggio, T. (1999a). Are cortical models really bound by the ”binding problem”? *Neuron*, 24(1):87–93.
- Riesenhuber, M. and Poggio, T. (1999b). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2(11):1019–25.
- Riesenhuber, M. and Poggio, T. (2000). Models of object recognition. *Nature Neuroscience*, 3(supplement):1199–1204.
- Rockland, K. S. (1998). Complex microstructures of sensory cortical connections. *Current Opinion in Neurobiology*, 8:545–51.
- Roskies, A. L. (1999). The binding problem. *Neuron*, 24(1):7–9.

- Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning internal representations by error propagation. In Rumelhart, D. E., McClelland, J. L., and The PDP Research Group, editors, *Parallel Distributed Processing: Explorations in the Microstructures of Cognition. Volume 1: Foundations*, pages 318–62. MIT Press, Cambridge, MA.
- Rumelhart, D. E. and Zipser, D. (1985). Feature discovery by competitive learning. *Cognitive Science*, 9:75–112.
- Sagi, D. and Tanne, D. (1994). Perceptual learning: learning to see. *Current Opinion in Neurobiology*, 4:195–9.
- Salinas, E. and Abbott, L. F. (1996). A model of multiplicative neural responses in parietal cortex. *Proceedings of the National Academy of Sciences USA*, 93:11956–61.
- Salinas, E. and Sejnowski, T. J. (2001). Gain modulation in the central nervous system: where behavior, neurophysiology and computation meet. *The Neuroscientist*, 7(5):430–40.
- Salinas, E. and Thier, P. (2000). Gain modulation: a major computational principle of the central nervous system. *Neuron*, 27:15–21.
- Schroeder, C. E., Mehta, A. D., and Foxe, J. J. (2001). Determinants of attentional control in cortical processing: evidence from human and monkey electrophysiologic investigations. *Frontiers in Bioscience*, 6:d672–84.
- Schyns, P. G., Goldstone, R. L., and Thibaut, J.-P. (1998). The development of features in object concepts. *Behavioural and Brain Sciences*, 21(1):1–54.
- Sigala, N. and Logothetis, N. K. (2001). Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature*, 415:318–20.
- Sigman, M. and Gilbert, C. D. (2000). Learning to find a shape. *Nature Neuroscience*, 3(3):264–9.
- Spratling, M. W. (2002). Cortical region interactions and the functional role of apical dendrites. *Behavioral and Cognitive Neuroscience Reviews*, 1(3):219–28.
- Spratling, M. W. (2005). Learning viewpoint invariant perceptual representations from cluttered images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(5):753–61.
- Spratling, M. W. and Hayes, G. M. (2000). Learning synaptic clusters for non-linear dendritic processing. *Neural Processing Letters*, 11(1):17–27.
- Spratling, M. W. and Johnson, M. H. (2001). Dendritic inhibition enhances neural coding properties. *Cerebral Cortex*, 11(12):1144–9.
- Spratling, M. W. and Johnson, M. H. (2002). Pre-integration lateral inhibition enhances unsupervised learning. *Neural Computation*, 14(9):2157–79.
- Spratling, M. W. and Johnson, M. H. (2003). Exploring the functional significance of dendritic inhibition in cortical pyramidal cells. *Neurocomputing*, 52-54:389–95.
- Spratling, M. W. and Johnson, M. H. (2004a). A feedback model of visual attention. *Journal of Cognitive Neuroscience*, 16(2):219–37.
- Spratling, M. W. and Johnson, M. H. (2004b). Neural coding strategies and mechanisms of competition. *Cognitive Systems Research*, 5(2):93–117.
- Sugase, Y., Yamane, S., Ueno, S., and Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex. *Nature*, 400:869–73.
- Thornton, C. (1996). Re-presenting representation. In Peterson, D. M., editor, *Forms of Representation: An Interdisciplinary Theme for Cognitive Science*, pages 152–62. Intellect Books, Exeter, UK.
- Thornton, C. (1997). Truth-from-trash learning and the mobot footballer. Technical Report CSRP 504, School of Cognitive and Computing Sciences, University of Sussex.
- Toates, F. (1998). The interaction of cognitive and stimulus-response processes in the control of behaviour. *Neuroscience and Biobehavioural Reviews*, 22(1):59–83.
- Treue, S. (2001). Neural correlates of attention in primate visual cortex. *Trends in Neurosciences*, 24(5):295–300.
- von der Malsburg, C. (1973). Self-organisation of orientation sensitive cells in the striate cortex. *Kybernetik*, 14:85–100.
- Wallis, G. and Bühlhoff, H. (1999). Learning to recognize objects. *Trends in Cognitive Sciences*, 3(1):22–31.
- Walsh, V., Ashbridge, E., and Cowey, A. (1998). Cortical plasticity in perceptual learning demonstrated by transcranial magnetic stimulation. *Neuropsychologia*, 36(4):363–7.
- Watanabe, T., Sr., J. E. N., Koyama, S., Mukai, I., Liederman, J., and Sasaki, Y. (2002). Greater plasticity in lower-level than higher-level visual motion processing in a passive perceptual learning task. *Nature Neuroscience*, 5(10):1003–9.
- Yuste, R., Gutnick, M. J., Saar, D., Delaney, K. R., and Tank, D. W. (1994).  $Ca^{2+}$  accumulations in dendrites of neocortical pyramidal neurons: an apical band and evidence for two functional compartments. *Neuron*, 13(1):23–43.
- Zemel, R. S., Behrmann, M., Mozer, M. C., and Bevalier, D. (2002). Experience-dependent perceptual grouping and object-based attention. *Journal of Experimental Psychology: Human Perception and Performance*, 28(1):202–17.

Zipser, K., Lamme, V. A. F., and Schiller, P. H. (1996). Contextual modulation in primary visual cortex. *Journal of Neuroscience*, 16(22):7376–89.