Modeling Learned Categorical Perception in Human Vision

Matthew C. Casey[a,*], Paul T. Sowden[b]

[a] *Department of Computing, University of Surrey, UK*
[b] *Department of Psychology, University of Surrey, UK*

Abstract

A long standing debate in cognitive neuroscience has been the extent to which perceptual processing is influenced by prior knowledge and experience with a task. A converging body of evidence now supports the view that task does influence perceptual processing, leaving us with the challenge of understanding the locus of, and mechanisms underpinning, these influences. An exemplar of this influence is learned categorical perception (CP), in which there is superior perceptual discrimination of stimuli that are placed in different categories. Psychophysical experiments on humans have attempted to determine whether early cortical stages of visual analysis change as a result of learning a categorization task. However, while some results indicate that changes in visual analysis occur, the extent to which earlier stages of processing are changed is still unclear. To explore this issue, we develop a biologically motivated neural model of hierarchical vision processes consisting of a number of interconnected modules representing key stages of visual analysis, with each module learning to exhibit desired local properties through competition. With this system level model, we evaluate whether a CP effect can be generated with task influence to only the later stages of visual analysis. Our model demonstrates that task learning in just the later stages is sufficient for the model to exhibit the CP effect, demonstrating the existence of a mechanism that requires only a high-level of task influence. However, the effect generalizes more widely than is found with human participants, suggesting that changes to earlier stages of analysis may also be involved in the human CP effect, even if these are not fundamental to the development of CP. The model prompts a hybrid account of task-based influences on perception that involves both modifications to the use of the outputs from early perceptual analysis along with the possibility of changes to the nature of that early analysis itself.

Keywords: Categorical perception, Modular neural networks, Hebbian learning, Task influence

1. Introduction

Understanding and modeling human vision is a highly challenging area of computational intelligence given the complexity and scale of the processing involved. One key question regarding vision is to what extent does prior knowledge and experience with a given task influence perceptual processing? Evidence suggests that such task influences happen as early as the primary visual cortex for visual discrimination (Li, Piëch, & Gilbert, 2004; Sowden & Schyns, 2006), and earlier for other activities including multi-sensory integration in the superior colliculus (Stein, 2005), cortical moderation of fear responses in the amygdala (Shi & Davis, 2001), and decision making in visual tasks with the basal ganglia (Bogacz & Gurney, 2007). What is interesting about this feedback is that it is between functional areas, is adaptive, and often changes as a function of task. It is therefore clear that feedback is a key component of visual analysis, but what is its computational role (Olshausen & Field, 2005) and at what stage does it occur in order to produce observed behavior?

Work on visual categorization has been one fruitful area for exploring the links between task and visual perception. For instance, psychophysical work suggests that the spatial frequency analysis carried out in the initial cortical stages of visual processing varies as a function of the categorization task that an individual is trying to perform (Sowden & Schyns, 2006). Categorization is a fundamental mechanism for dealing with the infinite variation in stimulation from our environment. By quickly identifying the category to which a visual stimulus belongs, we can rapidly access information about the likely properties of that stimulus and how we might interact with it. Without such abstracted knowledge we would have to deal with each visual event as entirely new, losing the benefit of previous learning about other similar stimuli.

In order to make rapid and consistent categorizations it is helpful if members of a category share many features in common that have low overlap with members of other categories. One approach is to attempt to 'carve nature at its joints' (Harnad, 1987) finding natural discontinuities in the dimensions of visual stimulation along which stimuli vary. However, in many cases there will be no ready joints

*Corresponding address: Department of Computing, University of Surrey, Guildford, Surrey, GU2 7XH, UK. Tel.: +44 (0)1483 689635; fax: +44 (0)1483 686051.

Email addresses: m.casey@surrey.ac.uk (Matthew C. Casey), p.sowden@surrey.ac.uk (Paul T. Sowden)

at which to separate visual stimuli into distinct and functional categories. It has been argued that in such circumstances, an individual learns to place stimuli which vary along such a dimension(s) into separate categories, with the relevant dimension(s) of variation becoming warped. This warping is such that stimuli that are placed into the same category come to appear more similar, while stimuli that are placed into different categories become less similar. This warping process marks an important difference between semantic categorization processes and the categorization of visual stimuli. Whereas semantic categories are defined by the abstract relations between objects (e.g. 'desk' and 'computer' are both members of the category of equipment used for academic work) visual categories are defined by the perceptual relationships between stimuli and only these categories acquire changes in the perception of physical stimulus properties through experience, resulting in the phenomenon of *categorical perception* (CP).

Behaviorally, learned CP effects are shown as more accurate and rapid discrimination of stimuli that are placed in different categories compared to discrimination of equally different (on some physical metric) stimuli that are placed in the same category. Put simply *between* category discrimination is superior to *within* category discrimination. CP therefore provides a key example of how task can influence perceptual analysis. By exploring how and where neural processing changes in order to facilitate the sharpening of the distinction between categories, we can improve our understanding of the feedback that occurs between neural circuits.

So where is the locus of change in visual CP? Evidence points to a wide range of brain areas involved in visual categorization and therefore potentially CP. It is known that cells in the pre-frontal cortex form strong categorical representations (Freedman, Riesenhuber, Poggio, & Miller, 2001, 2003); that patterns of activity in the inferotemporal cortex code for different object categories (Freedman, Riesenhuber, Poggio, & Miller, 2006; Kiani, Esteky, Mirpour, & Tanaka, 2007; Op de Beeck, Deutsch, Vanduffel, Kanwisher, & DiCarlo, 2008); and that cells there become tuned along diagnostic category dimensions (Sigala & Logothetis, 2002). Further, changes to early visual analysis in the occipital lobe could serve to amplify/attenuate differences in basic visual properties that serve to distinguish members of different categories. Any or all of these sites could be involved in the development of CP effects, from pre-cortical areas, through the occipital lobe, to ventral visual processing and beyond.

Notman, Sowden, & Özgen (2005) attempted to determine the possible neural locus of the CP effect. They trained human observers to categorize Gabor patch stimuli that varied in spatial phase and measured their ability to discriminate within and between category differences before and after training. They found that a CP effect developed as a result of training and that it did not generalize to exactly the same stimuli rotated to a different visual orientation. Because it is known that cells in the primary visual cortex are highly selective for stimulus orientation, Notman et al. deduced that the pattern of specificity of CP to stimulus orientation was consistent with changes to the processing conducted in these initial cortical visual processing stages. These may be implemented as dynamic task specific changes to the strengths of connections in these areas driven through feedback connections from later categorical processing stages. However, another possibility is that later stages in the visual processing hierarchy learn to make better use of the information being fed forwards from these initial processing stages (Mollon & Danilova, 1996; Petrov, Dosher, & Lu, 2005).

In this paper, our hypothesis is that only changes to the later stages of processing are required to induce CP; that is, feedback to initial cortical processing is not required. In order to evaluate this hypothesis, we need to develop a sufficiently plausible model of vision that demonstrates key properties needed for categorization, namely feature selection through localized receptive fields, and categorization through a global combination of features. There have been many attempts at building models of such aspects of visual analysis as well as those that model the combination of different visual functions (for example, von der Malsburg, 1973; Grossberg, 1976; Kohonen, 1982; Linsker, 1988; Jacobs, Jordan, & Barto, 1991; Miikkulainen, Bednar, Choe, & Sirosh, 2005; Serre, Oliva, & Poggio, 2007). Some common attributes of these models include hierarchical modules performing successive visual analysis, and self-organization through competition.

An exemplar of hierarchical vision is the modular model of early visual cortical processing developed by Itti, Koch, & Braun (1999), which exhibited equivalent psychophysically observed results. Their model consisted of layers representing orientation columns in V1, connected together to represent the interactions between feature selectors, which then fed a decision process. While the later stages are loosely based on mechanisms thought to occur in the brain, the model parameters are estimated by minimizing the error to the training data, rather than a more concrete understanding of the processing. However, this model does show that a simple modular architecture can demonstrate known psychophysical phenomena through the interaction of the different layers. Furthermore, it has been extended to explore task-driven attention, albeit using representations for long term and working memory (Navalpakkam & Itti, 2005). Although these models successfully demonstrate the properties we are looking for, we are motivated to find a more biologically plausible approach to modeling the modular nature of vision that uses neuronal principles to learn, and which can encode the influence of task. One such model is reported by Deco & Zihl (2001), which is a hierarchical model of visual attention which uses a recurrent network for feedback to guide attention. Other similar models have also been proposed (Spratling & Johnson, 2004; Hamker, 2007). Each of these builds on biological principles to provide a system level model of attention that compares well with behavioral data. Spratling & Johnson

(2006) go one step further and explore feedback in perceptual learning. By using a biologically plausible model of feedforward and feedback 'cortical regions', they demonstrated how task influence could induce a simple CP effect. However, while showing that feedback can induce a CP effect, they did not explore the interaction between different stages of visual analysis and task influence.

How do we model visual analysis so that we can test for an induced CP effect? One possible modular architecture which is both biologically motivated and encodes task training has been developed by Armony et al. (1995). While their approach is much simpler than the attentional models that have been developed, their uniform architecture of layers of identical, interconnected neurons uses Hebbian adaptation (Hebb, 1949) with lateral inhibition through competition (Rumelhart & Zipser, 1986) to train neurons to become sensitive to different, overlapping patterns. Similar competitive architectures built on biological principles have proven effective at modeling vision (Linsker, 1988; Miikkulainen et al., 2005), but have not been used for task training. Armony et al. (1995) used the competitive principle to model sub-cortical and cortical auditory pathways leading to the amygdala, the response from which could be adapted through the use of a conditioning signal. The idea of conditioning being used for category learning has been applied by Gluck & Bower (1988). In Armony et al.'s model, the conditioning signal is applied as an additional input to selected modules, modifying the response from their neurons as well as subsequently connected modules. While each module is formed by a single layer of neurons only, each learns to exhibit desired local properties (feature selection) as well as the whole model exhibiting the required global properties (conditioned responses).

Of particular interest to us is that the architecture defined by Armony et al. has been used to make biological processing predictions, which have subsequently been tested (Armony, Servan-Schreiber, Cohen, & LeDoux, 1997a; Armony, Servan-Schreiber, Romanski, Cohen, & LeDoux, 1997b), thereby validating the model and closing the circle between computational modeling and neuroscience. For us, this model provides a way in which different stages of vision can be modeled (modules) with task influence (categorization) being learned through association (conditioning). Both the local properties of the modules (feature detectors) and the global properties (categorization) can then be analyzed and compared with observed human data, building upon the successful use of such approaches before to model aspects of visual, cortical processing (Linsker, 1988; Miikkulainen et al., 2005). This approach for the first time therefore allows us to build a model of the different stages of visual analysis so that we can explore whether CP emerges as a result of changes to later stages of visual processing only, comparing this with relevant behavioral experiments.

In this paper we present an adaptive model of hierarchical vision processes (section 2) that is trained to perform a categorization task. This model is based upon the work by Armony et al. (1995), and hence we model hierarchical vision at a system level, and include task training to generate a CP effect. This work is novel because we are 1) evaluating whether an abstract, system level model of visual processing can sufficiently model early visual analysis such that it exhibits both local (feature detection) and global (categorization) behavior, 2) whether a CP effect can be induced through task influence to the later stages of processing, and 3) comparing our results systematically with behavioral experiments. This will allow us to provide computational evidence for the extent to which task influence is required through later or earlier stages of visual analysis in humans for categorical perception to arise. To achieve this, we execute the experiments conducted by Notman et al. (2005) on the model. These experiments tested whether early visual analysis changed as a result of learning a categorization. Our results show that the model is capable of reproducing the desired visual analysis behavior (section 3). We then go on to show that task influence is only needed in the highest level of analysis that we model (ventral visual processing), rather than at early stages of analysis as hypothesized in humans. However, while our model shows an existence proof for CP without early perceptual changes, we note that this effect does not fully replicate the specificity of the human data to stimulus orientation, suggesting that some other change at early stages of analysis (for example, see Furmanski et al., 2004; Schoups et al., 2001; Yang & Maunsell, 2004) may be needed to fine-tune the CP effect.

2. Modeling learned categorical perception

The model we present in this paper explores whether task influence to later stages in the visual processing hierarchy is sufficient to drive CP effects. As such, the model must provide sufficient capability to explicitly encode different hierarchical vision processes and allow for task influence to be input into these at relevant stages. Furthermore, the model needs to clearly represent these different processing stages without being overly complex to allow for interpretation and understanding of the results. Key elements are therefore the input representation, the modularity of the processing, and finally how it can adapt when repeatedly performing the human equivalent task. All these must come together to allow the model to plausibly (as much as can be achieved with a system level model) represent how human vision behaves when performing the selected task.

To achieve this, we base our model on the modular architecture developed by Armony et al. (1995). While there are successful neural models of category learning (Gluck & Bower, 1988; Kruschke, 1992; Jäkel, Schölkopf, & Wichmann, 2008), one question that arises from selecting a neural model is whether it will be sufficient to allow a CP effect to be generated? Damper & Harnad (2000) conducted a systematic review and comparative experiments

to determine if neural models are valid in such studies. They concluded that if categorization is in some way built into a sensory stimulus, "any general learning system operating on broadly neural principles ought to exhibit the essentials of CP" (p.862). This is backed up by evidence for a number of key architectures with *synthetic CP*, as Damper & Harnad term it, occurring to some degree irrespective of the way in which the model is defined or the parameters that are used. Instead, for synthetic CP to emerge, specialized processing is not required because of the potential for categorization in the input. So if such a general, neural learning system is sufficient, a competitive model that consists of hierarchical visual processing modules should also be capable of exhibiting CP. However, no such model has yet been developed to explore CP in hierarchical vision.

For us, the use of a simple Hebbian learning model that can be studied at both a behavioral and physiological level is key because it allows us to observe whether the CP effect has been learned, while ensuring that the components of the model themselves are acting at least plausibly as individual populations of neurons (for example, visual receptive fields). A schematic of our model of learned categorical perception in human vision is shown in Fig. 1.

[Figure 1 about here.]

2.1. Input representation

In the first human experiment by Notman et al. (2005), eight images were constructed, each comprised of a pair of Gaussian windowed gratings (Gabor patches) with different spatial frequencies (f and $3f$) that were combined to form compound Gabors varying in the relative spatial phase of the two components (Fig. 2). During the experiment, the observers' ability to discriminate pairs of images presented side-by-side on a computer display at a fixed viewing distance was measured before and after they had learned to categorize the images displayed at a single orientation (45° from vertical) on the basis of the spatial phase variation. A given pair of images could either be identical or different in spatial phase.

[Figure 2 about here.]

While presenting the same images to the computational model would be desirable, even simple images such as these would require significant pre-processing in order to extract separate information channels for spatial frequency and phase, something achieved in humans at early stages of vision. Therefore, because the images are designed to vary the $3f$ phase only, for the model we chose to represent each image as a pattern of phase activity for $3f$ phases as $P = \{0°, 45°, ..., 315°\}$. For completeness, we also include an f phase input with a constant value of 0 to represent the combination of phase information. Each of the eight images was therefore represented as a 9-dimensional vector.

For a given phase, the appropriate input is formed as the center of a Gaussian pattern of activity with mean equal to the selected phase S_p and bandwidth $\lambda_p = 106°$ (to match approximate phase selectivity of early cortical neurons), decreasing in strength with increasing difference in phase and wrapping around so that a phase of 360° is equivalent to 0°. Here then, an input x corresponding to an activation for phase $p \in P$ and orientation $o \in O$ is given by:

$$x_{po} = e^{-\Lambda_p(p-S_p)^2 - \Lambda_o(o-S_o)^2} \quad (1)$$

$$\Lambda_p = \frac{-\ln 1/2}{(\lambda_p/2)^2} \quad (2)$$

$$\Lambda_o = \frac{-\ln 1/2}{(\lambda_o/2)^2} \quad (3)$$

where S_p and S_o are the stimulus phase and orientation, and λ_p and λ_o the phase and orientation bandwidth, respectively. Note that the values Λ_p and Λ_o are chosen so that the associated bandwidth is achieved with the Gaussian at half the height of the curve. Two example inputs are shown in Fig. 3.

[Figure 3 about here.]

In a second experiment, Notman et al. extended their initial approach to explore the effect of orientation on CP. They modified the stimuli described above by rotating each image to provide distinct orientations. Category training carried out with the subjects during the task still only used examples with an orientation of 45° relative to vertical. The additional orientations were designed to test what generalization of CP had occurred to increasingly different orientations of ±2°, ±5°, ±15° and ±30° relative to the 45° training orientation, to give the set of images with orientation $O = \{0°, 15°, 30°, 40°, 43°, 45°, 47°, 50°, 60°, 75°, 90°\}$ (extended here to 0° and 90° for completeness).

For our model, the addition of stimuli at different orientations is encoded by extending the Gaussian pattern of activity used for phase into a two dimensional representation for phase versus orientation. Fig. 4 shows a plot of the $3f$ phase versus orientation input data for orientations 0° to 90°, with an input centered on a $3f$ phase of $S_p = 135°$ at orientation $S_o = 45°$. An orientation bandwidth of $\lambda_o = 30°$ was chosen for the Gaussian, again to match human data (for example, Campbell & Kulikowski, 1966; De Valois & De Valois, 1988).

[Figure 4 about here.]

The overall input to the model consists of four elements: the left and right visual field inputs together with the associated left and right category signals (Fig. 1). Each visual field input is selected to represent a chosen grating image, with a zero input used if no grating image is being presented. If more than one orientation is being input to the model, then the associated set of 9 dimensional vectors are combined in orientation order for each visual field by

concatenation into a single vector. The category signal is a single binary value per orientation representing whether the associated visual field input belongs to category A ('0') or category B ('1'). The category signal is only used during the category training stage, and is otherwise always presented as a '0'.

2.2. Modeling visual processing

Our model consists of three pairs of modules representing key stages in human vision processing (Fig. 1): pre-cortical (PC) processing (such as retina and LGN), early visual cortical (EC) processing (such as V1, V2) and ventral visual (VV) processing (for example, PIT, AIT). Matching the human experiments, we present two images to the model, one in the left and one in the right visual field. Image (and hence category) discrimination is then achieved by comparing the responses to these separate streams of processing, representing the left and right hemispheres of the brain (left processing the right visual field and vice versa). Here, the left PC feeds its output to the left EC, which in turn feeds its output to the left VV. Similarly for the right hemisphere.

According to Gross and colleagues (Gross, Bender, & Rocha-Miranda, 1969; Gross, Rocha-Miranda, & Bender, 1972), unlike earlier visual areas, the receptive fields of cells in inferotemporal cortex are often somewhat bilateral (56%), extending across the midline. Other IT cells respond only to the contralateral visual field (34%) and some only to the ipsilateral field (20%). However, in the case of bilateral cells, the receptive field centres rarely extended more than 4° into the ipsilateral visual field (less than 2% of cells), and were predominantly located in the contralateral field (79%) and generally showed a stronger response to contralateral stimulation. Since the human data we are comparing with are based upon responses to Gabor patches 7.1° away from the foveated fixation marker, the only predominant responses are contralateral, since in the ipsilateral visual field, these stimuli would only have been optimal for a tiny minority of bilateral IT cells. To therefore keep the model simple, we model only contralaterally responsive units. This very simplified view of visual processing therefore culminates in the VV, where neurons have been shown to become tuned along diagnostic category dimensions (Sigala & Logothetis, 2002).

Following the architecture defined by Armony et al. (1995), each module consists of a single layer of rate coded neurons fully connected to the input. The activation for a neuron j for an d-dimensional input x is calculated as:

$$u_j = \sum_{i=1}^{d} x_i w_{ij}(t) \tag{4}$$

$$y_j = \begin{cases} f(u_j) & \text{if } j = \arg\max_i f(u_i) \\ f(u_j - \mu y_{win}) & \text{otherwise} \end{cases} \tag{5}$$

$$f(u) = \begin{cases} 1 & u \geq 1 \\ u & 0 < u < 1 \\ 0 & u \leq 0 \end{cases} \tag{6}$$

where $w_{ij}(t)$ is the weight for input i to neuron j, $1 \leq j \leq M$, at time step t, which are all initialized uniformly to small random values (when $t = 1$), y_{win} is the activation value of the winning neuron so that $y_{win} = \max_i f(u_i)$, and μ is the inhibition rate. The time step t varies such that $1 \leq t \leq N$, where N is a multiple (epochs) of the number of inputs being presented.

Armony et al.'s model uses competitive learning to form feature detectors, and is itself a development of the competitive algorithm developed by Rumelhart & Zipser (1986). Competition is achieved through lateral inhibition in a module. Having calculated the activation y for all of the neurons in the module (M), the neuron with the maximum activation is chosen as the winner, with the activation value then used to suppress the activation of all other neurons (equation 5). The inhibition rate μ determines the rate at which the winner suppresses the other neurons, such that if $\mu = 0$ no inhibition is applied, versus $\mu = 1$ where effectively only the winning neuron is active. Therefore, μ controls the number of neurons active for a given input, determining the spread of activity in the module. For the case of the left and right PC and EC modules, these output activations are then fed as inputs to the next module.

2.3. Learning: feature selectors and category training

Competitive learning in each module is achieved through a simple Hebbian learning rule applied to all weights that have input values above a defined threshold:

$$w'_{ij}(t) = \begin{cases} w_{ij}(t) + \eta x_i y_j & \text{if } x_i > \rho\bar{x} \\ w_{ij}(t) & \text{otherwise} \end{cases} \tag{7}$$

$$\bar{x} = \frac{1}{m}\sum_{i=1}^{m} x_i \tag{8}$$

$$w_{ij}(t+1) = \frac{w'_{ij}(t)}{\sum_{k=1}^{d} w'_{kj}} \tag{9}$$

where η is the learning rate. Armony et al. (1995) just use the mean $\bar{x}$ of all the inputs to the module as the threshold, with the effect that the weights that have above average input are increased in strength, whereas those with below average input are decreased through normalization with respect to all of the inputs to a neuron j (equation 9). We modify the threshold to include the factor ρ so that we can vary the threshold above which an increase in the weight is made to determine what effect this has on feature selection.

While this Hebbian learning rule provides the ability for the modules to become feature selectors, we have not yet dealt with how task influence during category training can be incorporated into the learning process. For Armony et al., this change in the association was formed as a pairing between the conditioned stimulus (a selected input) and an unconditioned stimulus (for example a foot shock in animal studies). The effect is to bias learning for a particular conditioned stimulus, changing the pattern of

activation by causing associated neurons to become the winners and hence their weights to be strengthened. For us, our task influence is category training, where, in the human studies, the subjects are given feedback on whether a particular image belongs to category A or B. Here then, we use Armony et al.'s conditioning as category training by associating a category signal with a category's four example inputs. This is simply achieved by including a category input to the VV (Fig. 1) which is then active during category training for all category B examples. At all other times, the category input remains '0' (including for category A examples). The category input to VV is similar to the category signal thought to be fed back from the pre-frontal cortex, where strong category representations are held in memory (Freedman, Riesenhuber, Poggio, & Miller, 2003).

To ensure that the category signal always has an influence on the competitive process during learning, the category input weights are fixed at a predefined value W_c and are not subject to learning, however they do still influence weight normalization (equation 9). For both non-category and category training, learning is repeated on the training data set until stability in the weights is achieved.

2.4. Measuring CP

Our description of the model so far has shown how we present a series of paired inputs, feed forward activation through the left and right hemisphere modules, and train each module to form receptive fields with a category bias (details in section 3). However, how do we measure the behavior of this model and compare it to the human results? The human experiments were conducted in such a way as to allow changes in the participants' discrimination between categories to be measured (see Notman et al., 2005, for full details). For our model we will conduct a similar comparison.

Discrimination performance was measured by the number of hit and false alarm responses to the same-different image task. Hits are measured separately for within and between category image pairs. A $Hit(W)$ was counted if the participant correctly identified two images as being different for each pair of images taken from within the same category. Similarly, a $Hit(B)$ was counted if the subject correctly identified images as being different for each pair of images taken from between categories. Lastly, a *False Alarm* was counted if the subject identified the images as being different when they were identical. The ability of the participants to discriminate was then calculated as an A' score (Pollack & Norman, 1964), which is a non-parametric measure of the area under the single-point *Receiver Operating Characteristic* (ROC) curve. This is calculated from the probabilities of *Hits* and *False Alarms* separately for the within and between responses:

$$A' = \begin{cases} \frac{1}{2} & \text{if } H < F \\ \frac{1}{2} + \frac{(H-F)(1+H-F)}{4H(1-F)} & \text{otherwise} \end{cases} \quad (10)$$

where H is the probability of a *Hit* (separately within or between) occurring, and F is the probability of a *False Alarm*. We therefore obtain an $A'(W)$ for within and an $A'(B)$ for between that estimate the discrimination ability of the subject.

For our model, we attempt to carry out the same experimental method by using the relevant testing and training phases. As a consequence, once a stable model of human vision is established, we conduct discrimination testing to record the baseline performance, we then perform category training, and finally we test the model again to note the change in discrimination.

To obtain a *Hit* or *False Alarm* count, we compare the responses of the left and right modules at each processing stage. As in the human experiments, these values are obtained over 10 trials of the same pairs of images presented randomly to either visual field. For a given image pair, the sum of the outputs of the left and right module neurons are calculated separately, normalized, then compared with each other to determine if they are producing a similar or different value:

$$Y' = \sum_{i=1}^{n} y_i \quad (11)$$

$$Y = \frac{Y' - \min_i(y_i)}{\max_i(y_i) - \min_i(y_i)} \quad (12)$$

where Y is the normalized sum of the outputs for either the left or right module for a given input. The responses are treated as different if

$$|Y_{left} - Y_{right}| > \delta \quad (13)$$

where δ is a pre-defined difference threshold.

3. Experiments and results

We now consider the results from the model for two sets of experiments mimicking those performed on humans. The first performs category training on a single orientation to determine if we can model the development of the CP effect. The second determines whether the CP effect generalizes over multiple orientations.

3.1. Training data

While the range of stimuli for discrimination testing and category training are defined for us by the psychophysical experiments, we need to select an appropriate set of input data to develop a stable model in the pre-training phase. Since we first have to develop a model that is capable of discriminating at the neuronal level in each module between stimuli of different phases, a natural choice of training data at this stage is the set of all possible inputs. We therefore use a pre-training data set that consists of all eight inputs representing images 1 to 8 as shown in Fig. 2. Since the model has two visual fields, we need to present an input to both. To ensure that the model can discriminate

the absence of an input to a visual field, we pair each image with the zero-vector representing this, such that each input consists of an input representing a particular phase to one visual field, and the zero-vector to the other visual field. Which visual field receives which input is randomly chosen for each presentation. One training epoch therefore corresponds to the presentation of all eight phase inputs to one of the visual fields. Example responses from a model pre-trained for 10000 epochs are shown in Fig. 5.

[Figure 5 about here.]

For category training, we use the same range of stimuli as used for the psychophysical experiments. Here we restrict ourselves to the case when we have a single orientation $S_o = 45°$. One run of category training consists of a single block of the double and then the single training task. The double training input therefore consists of stimuli representing all possible image pairs (12 from the same category and 16 from different categories). An image from each pair (at random) is presented to the left and right visual hemifields. For single training, each presentation consisted of each phase input presented to one of the visual fields, randomly selected, with the other visual field receiving a zero-vector input representing no image. A training epoch therefore consisted of 28 double followed by three lots of 8 single inputs selected in random order (a total of 52 inputs). When providing category feedback, the corresponding category value for the visual field's input is given to the VV module. When not giving category feedback, zero is input.

3.2. Learned categorical perception

Key parameters in the model (Table 1) were determined through a systematic evaluation of the model's discrimination performance, coupled with assumptions about the values of the fixed weight and the learning rate, which are as per Armony et al. (1995). Discrimination testing matches that used for the human experiments, with pairs comprising the same input presented to both visual fields (images 1 to 8 in Fig. 2), and each consecutive pair (pairs 9 to 16). The within and between category responses for the relevant pairs are recorded by determining the difference between the left and right VV modules (equation 13) using a value of $\delta = 0.2$.

To determine if the model can exhibit the CP effect, we trained 100 models, all with different initial random weights, using the parameters in Table 1. Pre-training was conducted using the data set described in section 3.1, which was then followed by category training.

[Table 1 about here.]

The mean $A'(W)$ and $A'(B)$ values derived from the PC, EC and VV over the 100 networks are shown in Fig. 6a), b) and c), compared to the human data over 16 observers in Fig. 6d). We can see that the models' VV modules exhibit a similar profile of behavior to the humans in that before category training there is little difference between the mean $A'(W) = 0.56$ and $A'(B) = 0.58$ values. After category training, the influence of the category signal shows that the mean $A'(W)$ has dropped to 0.52, whereas the $A'(B)$ has increased significantly to 0.86. This is the CP effect. To determine whether changes to PC and EC during category training had any effect on the A' scores in VV we replaced the weights of the PC, EC and the PC and EC with their pre-category training weights. This had no effect on A' scores.

[Figure 6 about here.]

In order to facilitate comparison with the human data we assessed the statistical significance of these observations using analysis of variance (see Table 2). In the following description any differences referred to were statistically significant ($p < 0.001$) as assessed using Bonferroni post-hoc analyses. Overall these analyses confirmed the statistical significance of the observations made above.

[Table 2 about here.]

$A'(B)$ scores in the VV improved as a result of category training and this was true even when the weights in the PC and EC modules or both together were replaced with their pre-category-training values. The improvement in between category discrimination is consistent with the often observed between category expansion effects seen in humans where objects that are placed in different categories become perceptually less similar (also known as *acquired distinctiveness*).

$A'(W)$ scores in the VV actually reduced as a result of category training and again this was true when the weights in PC, EC or both together were replaced with their pre-training values. This decline in within category discrimination is consistent with the sometimes observed within category compression effects seen in humans where objects that are placed in the same category become perceptually more similar to each other (also known as *acquired equivalence*).

In summary, what this detailed analysis shows is that the model is capable of exhibiting the categorical perception effect under matched conditions to that of humans. That is, when discriminating between images that encode phase, and which are divided into two distinct categories based on these phases, the model has enhanced between category discrimination, versus unchanged (or suppressed) within category discrimination, and this is exhibited in the VV module only. This is in contrast to the results of Notman et al., who provided results suggesting that the CP effect occurred as a result of changes to earlier stages of perceptual processing, possibly as early as V1. Our model learns CP by the application of a category signal to the last module in our abstract visual processing stream, namely VV. However, in the present model there were no explicit connections back from VV to the EC and this may limit the possibility of observing CP at these earlier stages. This

leaves us to ask what influence the PC and EC modules have on our results, since the human evidence suggests that this earlier processing could be important? The analysis shows that replacing the PC, the EC and finally the PC and EC modules with their pre-trained variants does not affect the strength of the CP effect observed in VV. However, no CP effect is observed in VV prior to category training. This suggests that the CP effect may arise because during category training the VV module learns to make use of the phase selective outputs from PC and EC, which are present after pre-training. This is consistent with the hypotheses that CP may be based upon changes to the use of information from early visual filters by later stages of analysis (Mollon & Danilova, 1996; Petrov, Dosher, & Lu, 2005). However, a key result of the human studies was the high degree of specificity of the learned CP effect to the orientation of the visual stimulus during training. Consequently, we next explored whether the present model showed this same specificity.

3.3. Orientation generalization of categorical perception

Having shown that we can reproduce the CP effect in the model when discriminating between inputs representing phases in a single orientation, we now turn our attention to the second psychophysical experiments run by Notman et al. to determine how far this effect generalizes to inputs representing other orientations. Their experiments show that the CP effect is specific within a 6.5° orientation bandwidth (at half amplitude), when category training on a single orientation. We therefore repeat these experiments with our model to determine the specificity of CP to stimulus orientation.

To test our model on a variety of orientations, we use the input representation extended for orientations 0° through 90° as described in section 2.1. Here, to keep the input small (18 dimensional vectors), we pair an orientation of 45° with each of the orientations $O = \{0°, 15°, 30°, 40°, 43°, 47°, 50°, 60°, 75°, 90°\}$, and then combine the results.

[Figure 7 about here.]

For each pair of orientations, we trained 100 models with the same parameters as before, all with different initial random weights. Pre-training was conducted using the same data set, but with both orientations represented from the pair. However, category training was carried out with just the single orientation $S_o = 45°$. The average difference, over the 100 models, in mean $A'(B)$ values before and after category training for the VV module is shown in Fig. 7a). Here, we show the effect of the difference threshold on the $A'(B)$ values. Using the value of $\delta = 0.2$, the resulting orientation responses show a dip surrounding the category training orientation $S_o = 45°$. This is different to the observed human results, where the peak difference surrounds the trained orientation. However, when we vary the difference to make it smaller, the peak rises in the middle, and with $\delta = 0.01$, it is at its maximum. This parameter affects the way in which we calculate the same/difference response given by the model (equation 13). By making this smaller, we are looking for a normalized value from the left and right VV that differs by less to indicate a same response, and hence we are increasing the number of different responses with this smaller tolerance. In contrast, Fig. 7b) shows the effect of the weight change threshold ρ on the orientation generalization when $\delta = 0.01$. Whereas values of ρ from 1.5 and below tend to give sharp boundaries between orientation responses, values of ρ of 2 and 2.5 give a response that is closer in profile to the human data (normally distributed), albeit wider. Here we select $\rho = 2$, since this also maintains a stable value for the $A'(B)$ for 45° for all pairs, while higher values degrade the response to 0.

Having varied the difference and weight change thresholds to obtain an orientation tuned profile, we can compare the obtained results with the selected parameters. Fig. 8a) shows the change in the $A'(W)$ and $A'(B)$ responses over the 100 models in the VV module, compared to the human observations in Fig. 8b) for 12 observers. First, we note that the model clearly shows orientation specificity, with graded responses surrounding the trained orientation in a similar way to that observed in humans. This holds for both the within (suppression) and between (enhancement) responses. For the model responses, perhaps due to its deterministic nature and larger number of trials, the curves shown are far smoother. Second, whereas the CP effect is specific within a 6.5° bandwidth with humans, in the model we obtain a bandwidth of approximately 35°. No further tuning of the parameters could obtain a narrower response without degrading the learned CP, so although specificity has been demonstrated, the model's behavior differs from the humans in this respect.

[Figure 8 about here.]

To further facilitate comparison of the specificity of CP across orientation with the human data we assessed the statistical significance of these observations using four-way analysis of variance (orientations (11) – 0°, 15°, 30°, 40°, 43°, 45°, 47°, 50°, 60°, 75° or 90°; model variants (2) – after pre-training, after category training; modules (3) – PC, EC, VV; category positions (2) – within, between), which, importantly for the present analysis, showed a significant four way interaction ($F(20, 1980) = 67.4$, $p < 0.0000001$, observed power 1.0, partial $\eta^2 = 0.40$). Post-hoc analysis using three-way ANOVAs for each module and Bonferroni tests showed that only in the VV module did the CP effect vary as a function of stimulus orientation. Prior to category training there was no significant difference between within category and between category discrimination at any orientation, whilst after training these differences were significant for several orientations (from 30° to 60°, inclusive) that were closer to the training orientation. A between category expansion effect is seen as a result of category training for orientations between 30° and 60°

and a more narrowly tuned within category compression effect is seen for orientations between 40° and 50°. It is interesting to note that a similar, albeit more narrowly tuned, compression effect can be seen in the human data (Fig. 8b).

3.4. Discussion

Previous models of CP have tended to focus on speech, such as vowel (Anderson, Silverstein, & Ritz, 1977) and phoneme categorization (McClelland & Elman, 1986), and speech production (Kröger, Birkholz, Kannampuzha, & Neuschaefer-Rube, 2007). However, Goldstone, Steyvers, & Larimer (1996) explored categorization of visual stimuli in the form of Bezier curves, Padgett & Cottrell (1998) facial expressions and Spratling & Johnson (2006) a simple square of nine, joined dots. Our model falls somewhere between these two approaches by using visual stimuli, but which are represented in an abstract form to present phase information only from the Gabor patterns.

In addition to the types of task that have been modeled, various different architectures have been used to simulate CP. In the simulations conducted by Anderson et al. (1977), CP was modeled using Hebbian learning for positive feedback (the brain-state-in-a-box model). McClelland & Elman (1986) focused on using excitatory and inhibitory connections between neurons, whereas Kröger et al. (2007) used self-organizing maps. The model developed by Spratling & Johnson (2006) is perhaps the most plausible so far presented in that it consists of interacting layers of 'cortical regions' based on pyramidal neurons. In contrast to these simulations, which have some degree of biological plausibility in that they are each based around competitive learning or similar, other less plausible architectures have been used, such as Padgett & Cottrell (1998), who used an ensemble of multi-layered perceptrons trained using backpropagation.

In this paper we have gone beyond these existing models of CP. We have developed a biologically motivated model of the key stages of visual analysis that is capable of exhibiting local properties in each layer (feature selection) and global properties of the whole system (category learning). While this is similar to other models of hierarchical vision and categorization (Deco & Zihl, 2001; Spratling & Johnson, 2004; Hamker, 2007), our approach is novel because we then went on to demonstrate how the model can exhibit a CP effect as a result of sequential stages of visual analysis. These stages convert the input data into a form which can be categorized by our model of VV processing, which is coupled with a category signal as feedback.

In addition to developing a biologically inspired model of visual CP, we have also followed an equivalent training and testing regimen to that of a set of psychophysics experiments. Despite using simple neural structures and an abstract image representation, the results have both demonstrated a CP effect and its orientation specificity which compares well with human data. Here, our simple neural model applied a category signal only to the VV module, representing ventral visual processing, despite evidence from human data suggesting that the CP effect occurs as a result of changes at earlier stages of processing (for example, V1). If we treat our model as sufficiently plausible, our results provide evidence for the CP effect to occur as a result of changes to ventral processing only, although with some variation in the observed values and, in particular, the orientation bandwidth of the effect. These two different aspects of our findings may reconcile two different views of perceptual learning and categorical perception effects. One view, consistent with our finding that CP can arise from modifications to VV alone, posits that learning effects arise primarily because later stages of analysis learn to make better use of the fine grained information coming from early visual analysis (cf. Mollon & Danilova, 1996; Petrov et al., 2005). However, a second view proposes that learning more difficult tasks, or equivalently fine-tuning performance to a high degree, recruits successively earlier stages of analysis (cf. Ahissar & Hochstein, 2004). This latter view is consistent with the failure of our model to achieve the same degree of orientation tuning that is observed for the human data implying that, in humans, modifications are also driven to earlier stages of visual analysis to achieve this fine tuning.

Evidence in support of the likely involvement of early cortical stages of visual analysis in tasks involving orientation specific learning has come from perceptual learning research, which has sought to identify the neural correlates of such learning. For instance, Furmanski et al. (2004) showed that practice improved observers' ability to detect low contrast gratings in an orientation specific way and that activity in V1 in response to the practiced grating orientation increased following training. Further, Schoups et al. (2001) showed that in monkeys, practising orientation identification narrowed the tuning of orientation selective V1 neurons for practised orientations. In addition, Yang & Maunsell (2004) reported that practicing orientation discrimination produced narrower orientation tuning curves for units in V4 responding to the practised orientation at the practised stimulus location.

However, we must recognize that our model has three main limitations that we must consider along with this evidence. First, we must acknowledge that the model is an abstraction of a complex neural hierarchy, with arbitrary labels delineating broadly different stages of processing. What we have shown is that these three successive stages of processing can exhibit behavior which can be mapped to human data through the tuning of parameters, and that this behavior can be modified by the application of a category signal. The model also abstracts temporal information through the use of a rate coding architecture. However, despite these limiting factors, what the results have shown is that modeling can be used to explore complex biological processes and, more importantly, can be used to make predictions about these processes, namely that the CP effect appears to be possible without changes to early perceptual processing.

Second, in order to test orientation specificity, we paired orientations to keep the size of the input vectors small. While this is a valid and often used strategy for psychophysics experiments, humans are capable of processing multiple orientations at once. Tests that included all orientations in the input to our model did demonstrate the CP effect, but they did not show a decreasing profile of orientation generalization, with all orientations gaining the same $A'(B)$ values instead. This suggests that the model cannot adequately cope with significantly increased input dimensions (perhaps the curse of dimensionality being felt).

Third, one striking result is that we could not find parameters that would sufficiently tune the orientation bandwidth to the desired 6.5°. While we could dismiss this as a result of a fundamental limit of the architecture, this could also suggest that the model does not sufficiently represent the underlying biology. Here, the role of feedback connections underlying the CP effect is thought to be important to enable changes to be made to perceptual processing (for example, although not explicitly for CP, Crist & Gilbert, 2001; Gilbert, Sigman, & Crist, 2001). In our model, connections between modules are feedforward only; feedback connections may be important in achieving a smaller orientation bandwidth by modulating activation within layers as a result of category learning. This remains an important goal for future work.

4. Conclusions

In this paper we have provided computational evidence for the way in which prior knowledge of a task can influence perceptual processing. We have done this by presenting an adaptive model of hierarchical vision processes trained on a visual categorization task. In defining the model and establishing our experimental method, we have attempted to provide an equivalence between the model and human vision. To achieve this we have used biological inspiration in our choice of architecture and algorithm, systematically parameterized the model, executed equivalent experimental steps to that of the human experiments, and finally, rigorously analyzed the results using statistical approaches applied in psychophysics.

Contrary to current debate on the influence of task on visual analysis, our results demonstrate that a visual CP effect can be established with only task influence to the latter stages of analysis that we have modeled: ventral visual processing. However, our model also shows that the developed CP effect is less specific to the stimulus orientation than found in humans. So, while we propose that a basic CP effect may arise in later stages of processing learning to make better use of the phase selective outputs from early stages, a more complex form of CP may still require changes to perceptual processing in these earlier stages.

We have drawn this conclusion based upon the evidence that the model provides. Here, we find that the strength of the CP effect observed in VV is not affected by replacing the after category-training connection weights in PC, EC or both with their pre-training values. The phase selective outputs, which are present following pre-training and before any category training takes place, appear to be sufficient for CP when coupled with the learning that takes place in VV following application of the category signal. This proposal is consistent with late selection models of CP and perceptual learning (Mollon & Danilova, 1996; Petrov, Dosher, & Lu, 2005). However, this basic CP effect shows far less specificity to the orientation of the visual stimulus than is the case for the human data implying that some additional changes to perceptual analysis, possibly localized to earlier stages of processing, are involved in the human CP effect in line with the conclusions of Notman et al. (2005) and the Reverse Hierarchy Theory of Ahissar & Hochstein (2004). Such modifications might include changes to the strengths of intra-cortical connections (i.e. within module) under the influence of feedback connections from later stages of analysis (Crist & Gilbert, 2001; Gilbert, Sigman, & Crist, 2001).

Our results prompt further human experiments to test the hybrid account of CP effects proposed above involving both early and late stage modifications to perceptual analysis. For example, Transcranial Magnetic Stimulation designs might be used to selectively disrupt the processing at different stages of visual analysis at the point in time when task feedback is provided, and the effects of this on CP observed. Our model might predict that disruptions to early visual analysis could increase the generalization of CP effects across stimulus dimensions such as orientation.

However, in drawing these conclusions, we recognize that the model suffers from a number of limitations. For example, more capable models of vision may be built by considering the role of inter-module feedback, temporal coding (pulse coded networks) and the use of more sophisticated techniques to decrease the levels of abstraction (better models of receptive fields and the implementation of earlier stages of visual processing). Work on this has already started with, for example, the inclusion of topographic properties within the modules (Pavlou & Casey, 2009). Despite these limitations, the model does have a sufficient level of plausibility and replication of behavior to provide an existence proof that the CP effect can develop with only simple feedback to later stages of visual analysis.

Acknowledgments

We would like to thank Athanasios Pavlou for early discussions on this work as to how a model of fear conditioning could apply to category training. We would also like to thank the anonymous reviewers for their helpful comments.

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, *8*(10), 457–464.

Anderson, J. A., Silverstein, J. W., & Ritz, S. A. (1977). Vowel pre-processing with a neurally based model. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'77)*, vol. 2, (pp. 265–269).

Armony, J. L., Servan-Schreiber, D., Cohen, J. D., & LeDoux, J. E. (1995). An anatomically constrained neural network model of fear conditioning. *Behavioral Neuroscience*, *109*(2), 246–257.

Armony, J. L., Servan-Schreiber, D., Cohen, J. D., & LeDoux, J. E. (1997a). Computational modeling of emotion: Explorations through the anatomy and physiology of fear conditioning. *Trends in Cognitive Sciences*, *1*(1), 28–34.

Armony, J. L., Servan-Schreiber, D., Romanski, L. M., Cohen, J. D., & LeDoux, J. E. (1997b). Stimulus generalization of fear responses: Effects of auditory cortex lesions in a computational model and in rats. *Cerebral Cortex*, *7*(2), 157–165.

Bogacz, R., & Gurney, K. (2007). The basal ganglia and cortex implement optimal decision making between alternative actions. *Neural Computation*, *19*(2), 442–477.

Campbell, F. W., & Kulikowski, J. J. (1966). Orientation selectivity of the human visual system. *Journal of Neurophysiology*, *197*, 437–445.

Crist, R. E., & Gilbert, C. D. (2001). Learning to see: Experience and attention in primary visual cortex. *Nature Neuroscience*, *4*, 519–525.

Damper, R. I., & Harnad, S. R. (2000). Neural network models of categorical perception. *Perception and Psychophysics*, *62*(4), 843–867.

De Valois, R. L., & De Valois, K. K. (1988). *Spatial Vision*. Oxford, UK: Oxford University Press.

Deco, G., & Zihl, J. (2001). A Neurodynamical Model of Visual Attention: Feedback Enhancement of Spatial Resolution in a Hierarchical System. *Journal of Computational Neuroscience*, *10*(3), 231–253.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science*, *291*, 312–316.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *The Journal of Neuroscience*, *23*, 5235–5246.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2006). Experience-Dependent Sharpening of Visual Shape Selectivity in Inferior Temporal Cortex. *Cerebral Cortex*, *16*(11), 1631–1644.

Furmanski, C. S., Schluppeck, D., & Engel, S. A. (2004). Learning strengthens the response of primary visual cortex to simple patterns. *Current Biology*, *14*(7), 573–578.

Gilbert, C. D., Sigman, M., & Crist, R. E. (2001). The neural basis of perceptual learning. *Neuron*, *31*, 681–697.

Gluck, M. A., & Bower, G. H. (1988). From Conditioning to Category Learning: An Adaptive Network Model. *Journal of Experimental Psychology: General*, *117*(3), 227–247.

Goldstone, R. L., Steyvers, M., & Larimer, K. (1996). Categorical perception of novel dimensions. In *Proceedings of the 18th Annual Conference of the Cognitive Science Society*, (pp. 243–248).

Gross, C. G., Bender, D. B., & Rocha-Miranda, C. E. (1969). Visual receptive fields of neurons in inferotemporal cortex of the monkey. *Science*, *166*(3910), 1303–1306.

Gross, C. G., Rocha-Miranda, C. E., & Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the macaque. *Journal of Neurophysiology*, *35*(1), 96–111.

Grossberg, S. (1976). Adaptive pattern classification and universal recoding, i: Parallel development and coding of neural feature detectors. *Biological Cybernetics*, *23*, 121–134.

Hamker, F. H. (2007). The Mechanisms of Feature Inheritance as Predicted by a Systems-level Model of Visual Attention and Decision making. *Advances in Cognitive Psychology*, *3*(1-2), 111–123.

Harnad, S. (1987). *Categorical Perception: the Groundwork of Cognition*. Cambridge, UK: Cambridge University Press.

Hebb, D. O. (1949). *The Organization of Behavior: A Neuropsychological Theory*. New York: John Wiley and Sons.

Itti, L., Koch, C., & Braun, J. (1999). A quantitative model relating visual neuronal activity to psychophysical thresholds. *Neurocomputing*, *26-27*, 743–748.

Jacobs, R. A., Jordan, M. I., & Barto, A. G. (1991). Task decomposition through competition in a modular connectionist architecture: The what and where vision tasks. *Cognitive Science*, *15*, 219–250.

Jäkel, F., Schölkopf, B., & Wichmann, F. A. (2008). Generalization and similarity in exemplar models of categorization: Insights from machine learning. *Psychonomic Bulletin & Review*, *15*(2), 256–271.

Kiani, R., Esteky, H., Mirpour, K., & Tanaka, K. (2007). Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *Journal of Neurophysiology*, *97*, 4296–4309.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, *43*, 59–69.

Kröger, B. J., Birkholz, P., Kannampuzha, J., & Neuschaefer-Rube, C. (2007). Modeling the perceptual magnet effect and categorical perception using self-organizing neural networks. In *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS XVI)*, (pp. 789–792).

Kruschke, J. K. (1992). ALCOVE: An Exemplar-based Connectionist Model of Category Learning. *Psychological Review*, *99*, 22–44.

Li, W., Piëch, V., & Gilbert, C. D. (2004). Perceptual Learning and Top-down Influences in Primary Visual Cortex. *Nature Neuroscience*, *7*(6), 651–657.

Linsker, R. (1988). Self-organization in a perceptual network. *Computer*, *21*(3), 105–117.

McClelland, J. L., & Elman, J. L. (1986). The trace model of speech perception. *Cognitive Psychology*, *18*, 1–86.

Miikkulainen, R., Bednar, J. A., Choe, Y., & Sirosh, J. (2005). *Computational Maps in the Visual Cortex*. New York: Springer Science+Business Media.

Mollon, J. D., & Danilova, M. V. (1996). Three remarks on perceptual learning. *Spatial Vision*, *10*, 51–58.

Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research*, *45*(2), 205–231.

Notman, L. A., Sowden, P. T., & Özgen, E. (2005). The nature of learned categorical perception effects: A psychophysical approach. *Cognition*, *95*(2), B1–B14.

Olshausen, B. A., & Field, D. J. (2005). How close are we to understanding v1? *Neural Computation*, *17*(8), 1665–1699.

Op de Beeck, H. P., Deutsch, J. A., Vanduffel, W., Kanwisher, N. G., & DiCarlo, J. J. (2008). A Stable Topography of Selectivity for Unfamiliar Shape Classes in Monkey Inferior Temporal Cortex. *Cerebral Cortex*, *18*(7), 1676–1694.

Padgett, C., & Cottrell, G. W. (1998). A simple neural network models categorical perception of facial expressions. In *Proceedings of the Twentieth Annual Cognitive Science Conference*, (pp. 806–811).

Pavlou, A., & Casey, M. C. (2009). Identifying emotions using topographic conditioning maps. In M. Koeppen, N. Kasabov, & G. Coghill (Eds.) *Advances in Neuro-Information Processing: Proceedings of the 15th International Conference on Neuro-Information Processing, Lecture Notes in Computer Science 5506*, (pp. 40–47). Springer-Verlag.

Petrov, A. A., Dosher, B. A., & Lu, Z. (2005). The dynamics of perceptual learning: an incremental reweighting model. *Psychological Review*, *112*, 715–743.

Pollack, I., & Norman, D. A. (1964). A non-parametric analysis of recognition experiments. *Psychonomic Science*, *1*, 125–126.

Rumelhart, D. E., & Zipser, D. (1986). Feature discovery by competitive learning. In D. E. Rumelhart, & J. L. McClelland (Eds.) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. Volume 1: Foundations, (pp. 151–193). MIT Press.

Schoups, A., Vogels, R., Qian, N., & Orban, G. (2001). Practising orientation identification improves orientation coding in v1 neurons. *Nature*, *412*(6846), 549–553.

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences*, *104*(15), 6424–6429.

Shi, C., & Davis, M. (2001). Visual pathways involved in fear conditioning measured with fear-potentiated startle: Behavioral and anatomic studies. *The Journal of Neuroscience*, *21*(24), 9844–9855.

Sigala, N., & Logothetis, N. K. (2002). Visual categorization shapes feature selectivity in the primate temporal lobe. *Nature*, *415*, 318–320.

Sowden, P. T., & Schyns, P. G. (2006). Channel surfing in the visual brain. *Trends in Cognitive Sciences*, *10*, 538–545.

Spratling, M. W., & Johnson, M. H. (2004). A feedback model of visual attention. *Journal of Cognitive Neuroscience*, *16*(2), 219–237.

Spratling, M. W., & Johnson, M. H. (2006). A feedback model of perceptual learning and categorization. *Visual Cognition*, *13*(2), 129–165.

Stein, B. E. (2005). The development of a dialogue between cortex and midbrain to integrate multisensory information. *Experimental Brain Research*, *166*(3-4), 305–315.

von der Malsburg, C. (1973). Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik*, *14*(2), 85–100.

Yang, T., & Maunsell, J. H. R. (2004). The effect of perceptual learning on neuronal responses in monkey visual area v4. *The Journal of Neuroscience*, *24*(7), 1617–1626.

List of Figures

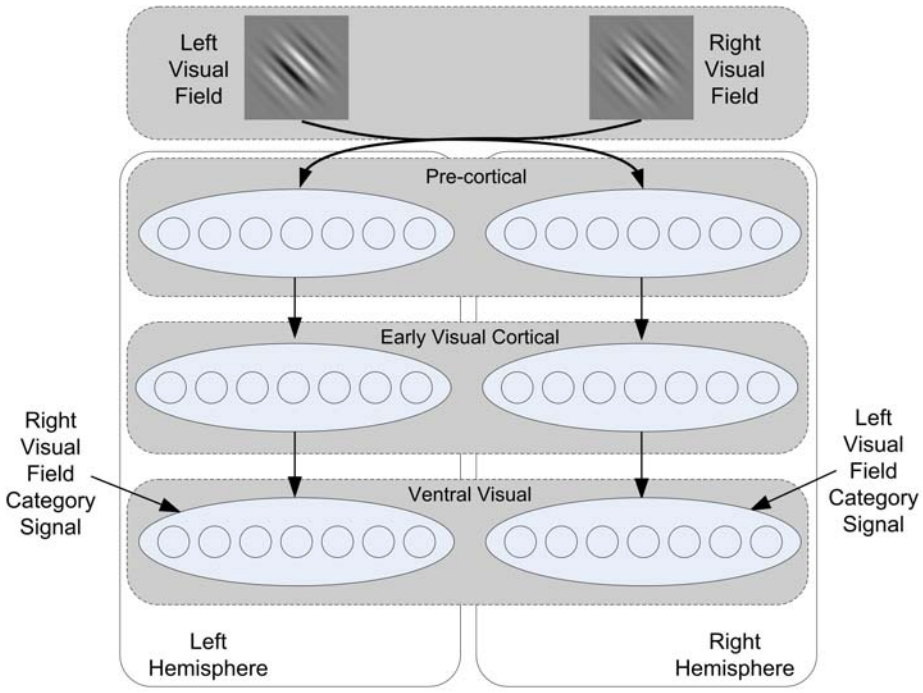


Figure 1: Schematic of the model of learned categorical perception in human vision. Left and right processing streams are shown with neural modules representing pre-cortical (PC), early visual cortices (EC) and ventral visual (VV) processing. When the model is being trained on the categories, the category signal is activated in VV.

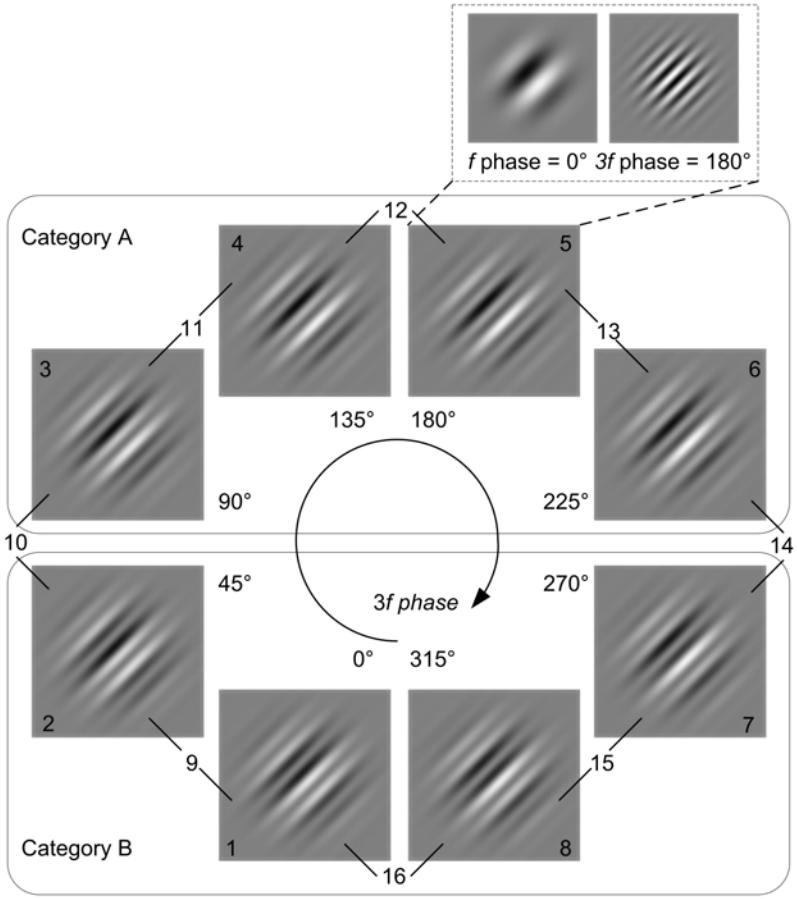
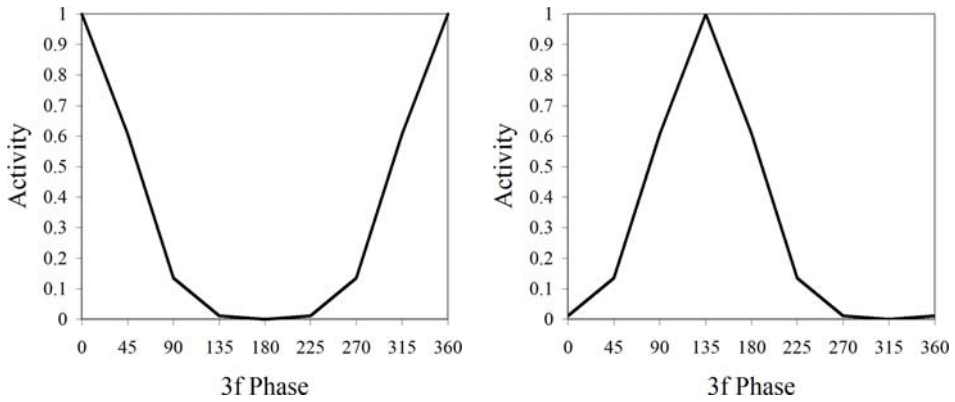


Figure 2: The set of eight images used for the human experiments, each constructed from two sets of Gaussian windowed compound gratings (Notman et al., 2005). The gratings consisted of sinusoids with spatial frequency $f = 0.32$ and $3f$ cycles per degree (cpd) each. The eight images were then constructed with a fixed f phase and varying $3f$ phase to form compound gratings, so that category A images had $3f$ phase {90°, 135°, 180°, 225°} and category B images {0°, 45°, 270°, 315°}. For convenience, each image is numbered (clockwise from 1 to 8), and also each image pair (again clockwise from 9 to 16). During discrimination testing, images are presented as pairs, in random order, with either the same image on the left and right visual field (1 to 8), or pairs made up of consecutive images (9 to 16), noting the difference in within (9, 11, 12, 13, 15, 16) and between category (10, 14) pairs.



a) 0° b) 135°

Figure 3: The visual field inputs representing a 3f phase of a) 0° and b) 135°, both for orientation 45°. Each input consists of an f component (not shown) and components for each of the $3f$ phases. The values for each 3f phase component are calculated using a Gaussian with mean corresponding to the input's $3f$ phase and bandwidth 106° to match the human data on phase selectivity. Note that a phase of 360° is equivalent to a phase of 0°, hence the values wrap around.

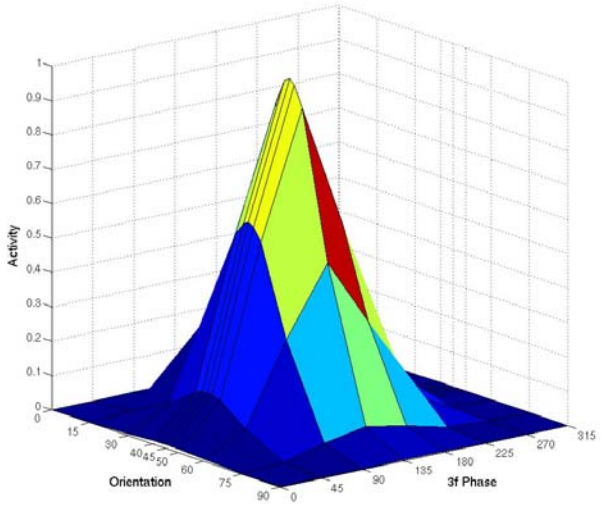
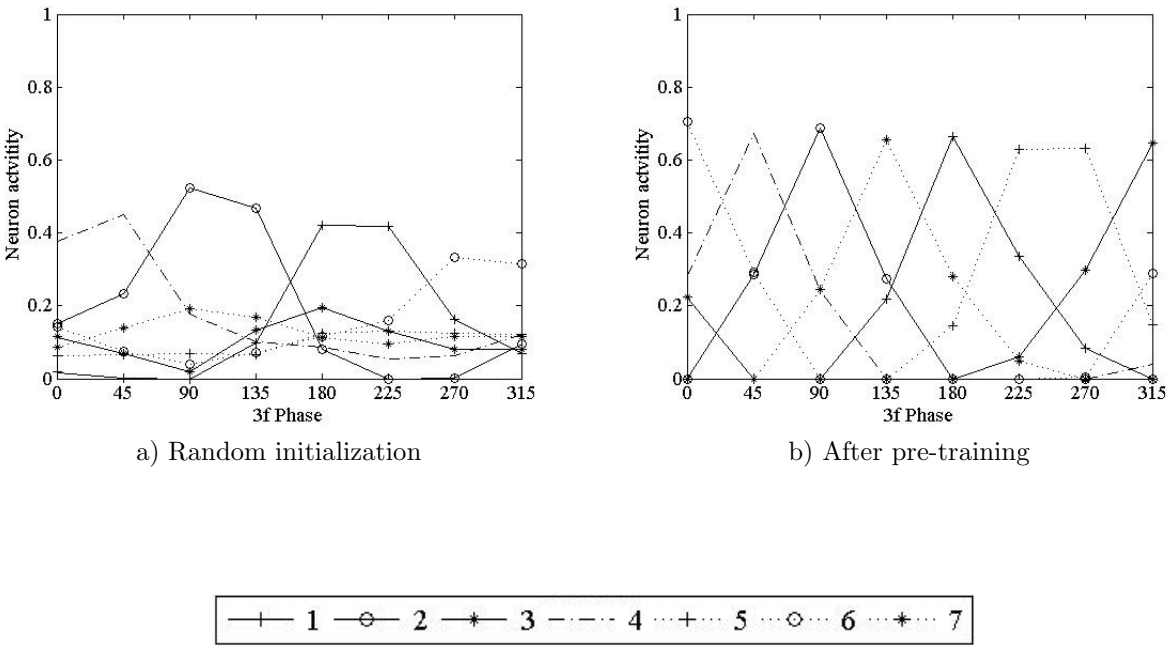


Figure 4: Surface plot showing an example input with a $3f$ phase of 135° and orientation 45°, spanning orientations 0° to 90°. The constant f phase value is not shown. The values for each phase and orientation are calculated using a Gaussian with mean corresponding to the input's $3f$ phase and orientation, phase bandwidth of 106° and orientation bandwidth of 30° to match human data.



a) Random initialization

b) After pre-training

Figure 5: Responses from the left PC of an example model which has 7 neurons in each module. Activity is shown for each $3f$ phase test pattern: a) using the randomly initialized weights, and b) after pre-training for 10000 epochs.

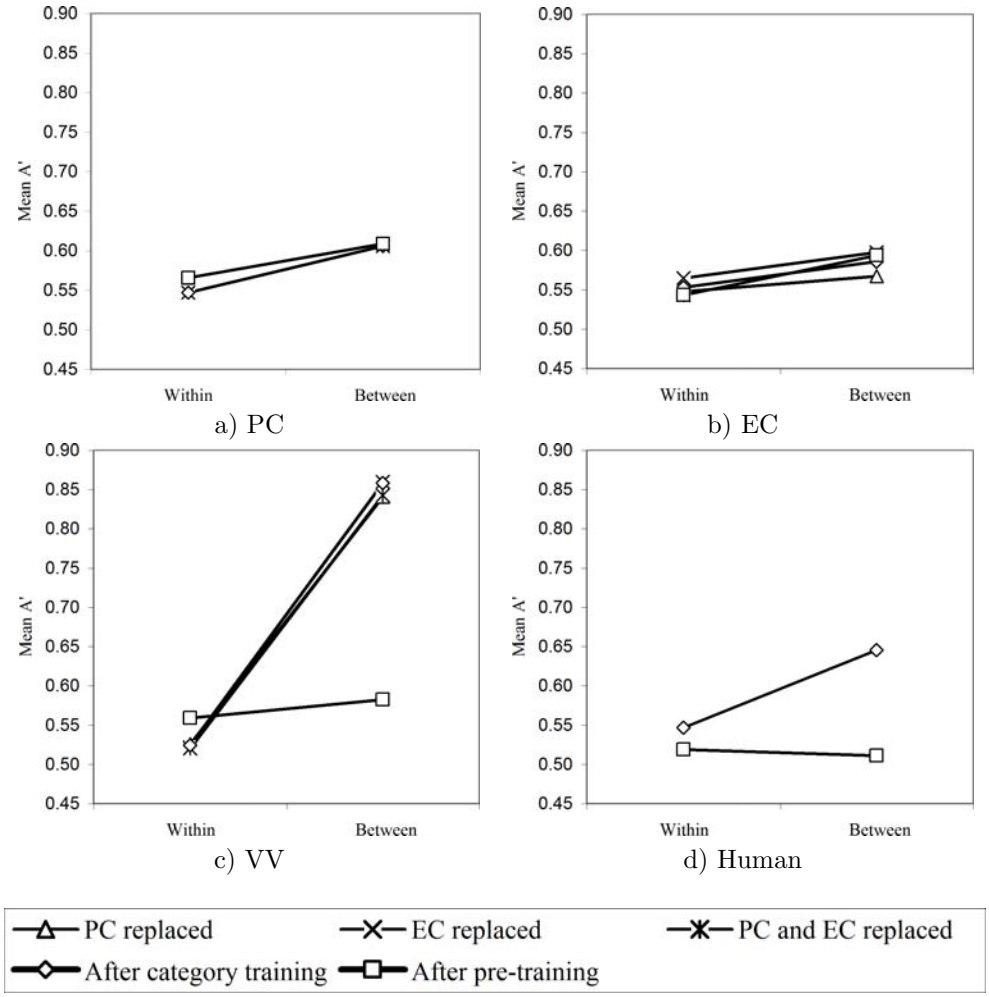
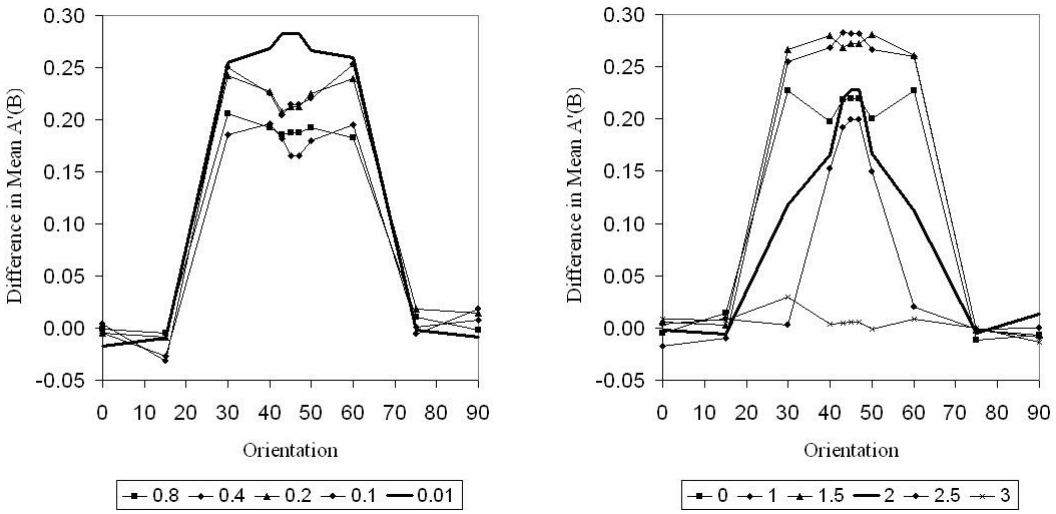
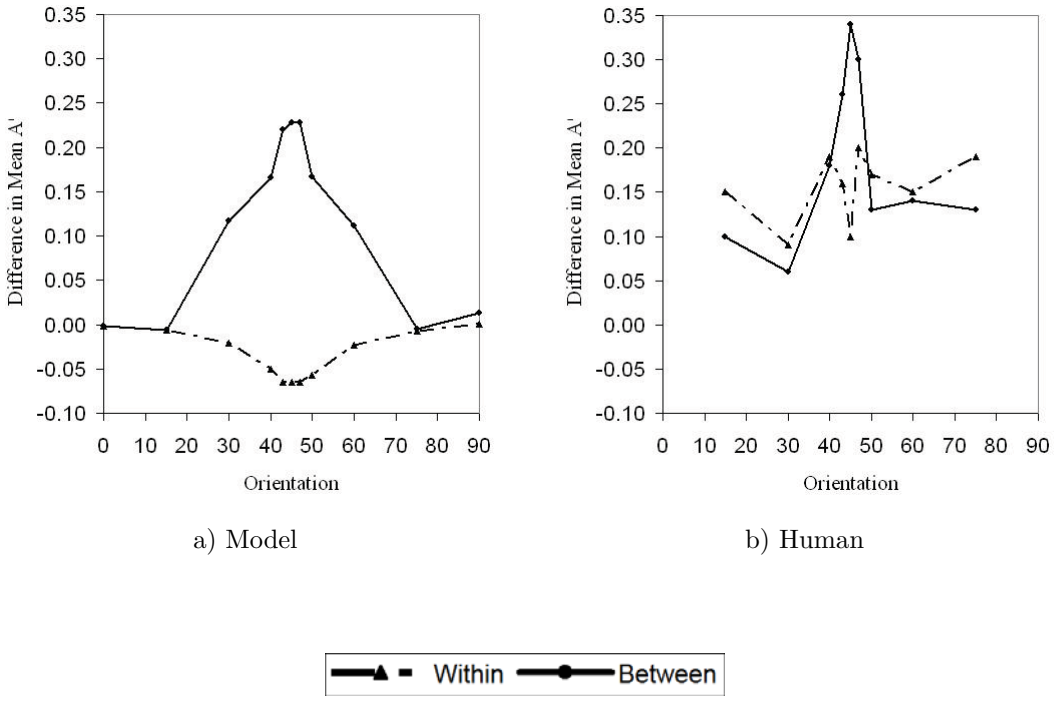


Figure 6: Mean $A'(W)$ and $A'(B)$ values for the computational model over 100 trials: a) PC, b) EC and c) VV, and b) human results over 16 observers (Notman et al., 2005). The lines depict the change in the A' values from those obtained after pre-training, to those after category training. Values are also shown for when the PC, EC, and PC and EC together were replaced in the model.



a) Difference threshold b) Weight change threshold

Figure 7: Difference in mean $A'(B)$ values before and after category training spanning orientations 0° to 90° with varying a) difference threshold $\delta = \{0.8, 0.4, 0.2, 0.1, 0.01\}$, and b) weight change threshold $\rho = \{0, 1, 1.5, 2, 2.5, 3\}$ when $\delta = 0.01$. The thick line denotes the selected values δ =0.01 and $\rho = 2$.



a) Model b) Human

Figure 8: Difference in mean $A'(W)$ and $A'(B)$ values for the a) computational model over 100 trials with difference threshold $\delta = 0.01$ and weight change threshold $\rho = 2$, and b) human experiments over 12 subjects (Notman et al., 2005).

List of Tables

Parameter	Value	
Neurons per module	M_{EC}	7
	M_{PC}	7
	M_{VV}	7
Inhibition rate	μ_{EC}	0.6
	μ_{PC}	0.4
	μ_{VV}	0.2
Category input fixed weight	W_c	0.4
Learning rate	η	0.1
Pre-training epochs	N_p	10000
Category training epochs	N_c	11
Weight change threshold	ρ	1
Difference threshold	δ	0.2

Table 1: Parameters used for the computational experiments to explore whether the model learns the categorical perception effect.

Effect	F	dF		Partial η^2
Model	18.29	2.40	237.54	0.16
Module	88.98	1.85	182.76	0.47
Category	581.39	1.00	99.00	0.85
Model × Module	23.76	4.33	428.71	0.19
Model × Category	43.13	2.30	227.72	0.30
Module × Category	181.93	1.78	176.51	0.65
Model × Module × Category	48.41	4.04	400.21	0.33

Table 2: Results from analysis of variance to explore A' scores for within and between category discriminations, in each visual area as a function of training. Here, we show analyses for five different variants of the model (after pre-training, after category training, then with the PC, EC and the PC and EC replaced), the three different modules in each processing stream (PC, EC and VV) and the category position (within and between). The Mauchley Sphericity Test showed that the assumption of sphericity was violated for all effects. Consequently, the Greenhouse Geisser correction was made to the degrees of freedom in every case. In all cases, $p < 0.0000001$ and the observed statistical power is very high (1.0) because each model was run a large number of times (100). With such high power even a very small difference will be statistically significant. A separate question is whether the difference is meaningful. An indication of this is given by measures of effect size such as partial η^2. Values for this can range between 0 and 1 with larger values indicating a bigger effect and a value of less than 0.2 a very small effect. In the present case most effects were moderate to large in size and importantly, this is true of the crucial interactions with category position.