

# Virtual Brain Reading: A Connectionist Approach to Understanding fMRI

**Rosemary A. Cowell (r.a.cowell@kent.ac.uk)**

Computing Laboratory, University of Kent,  
Canterbury, CT2 7NF, UK

**David E. Huber (dhuber@psy.ucsd.edu)**

Department of Psychology, University of California San Diego, 9500 Gilman Drive # 0109  
La Jolla, CA 92093-0109, USA

**Garrison W. Cottrell (gary@cs.ucsd.edu)**

Department of Computer Science and Engineering, University of California San Diego, 9500 Gilman Drive # 0404  
La Jolla, CA 92093-0404, USA

## Abstract

We present a neurocomputational model of visual object processing, which takes photographic inputs and creates topographic stimulus representations on the hidden layer. We perform multi-voxel pattern analysis on the activations of hidden units and simulate contradictory findings from Haxby et al. (2001) and Spiridon and Kanwisher (2002) within a single model. With no special processing mechanism or architecture for faces in our model, we obtain the same results as Spiridon & Kanwisher, who interpreted their results as evidence for a “face module” – something our model does not possess.

**Keywords:** MVPA; face processing; visual cortex; fMRI; visual expertise; visual perception; fusiform face area.

## Introduction

A region in the fusiform gyrus of the temporal lobe, dubbed the “Fusiform Face Area” (FFA), has been shown to be differentially activated when subjects in an fMRI scanner view faces compared to other classes of visual stimuli (Kanwisher et al. 1997). The FFA has been the subject of fierce debate. Opinions are divided between those who believe this area constitutes a specialized module for visual processing of faces as a unique class of visual stimuli (e.g., Kanwisher et al., 1997) and those who believe, instead, that the existence of a region in visual cortex dedicated to faces can be accounted for with alternative hypotheses, such as the “visual expertise” hypothesis (Gauthier, et al., 1999; Tarr & Gauthier, 2000), or the “object form topography” account (Haxby et al., 2001).

A number of studies attempting to resolve this debate have drawn upon multi-voxel pattern analysis techniques (MVPA) for analyzing fMRI data collected while subjects view images of objects and faces (Grill-Spector et al., 2004; Haxby et al., 2001; O’Toole et al. 2005; Spiridon & Kanwisher, 2002). MVPA techniques typically use statistical or machine-learning techniques to draw inferences from fMRI data by comparing the patterns of activation across multiple voxels in different experimental conditions.

Unlike traditional subtraction methodologies that focus on spatially contiguous voxels within a single patch of cortex, MVPA considers patterns across potentially disparate voxels that maximally discriminate between behavioral

conditions. This can be done within an anatomical region or even across regions.

Central to the debate concerning the existence of a face module are two seemingly contradictory sets of findings from two fMRI studies. Both studies used MVPA techniques to analyze patterns of response that were recorded while subjects viewed photographs of faces and objects. Haxby et al. (2001) (hereafter referred to as H01), reported that information about the category membership of the object being viewed was distributed across visual cortex, rather than confined to regions that were maximally active in response to that category. In contrast, Spiridon and Kanwisher (2002) (hereafter referred to as SK02), carried out a different set of analyses on a similar set of fMRI data and reported evidence for specialized processing of faces in regions preferentially activated by faces.

In this paper, we present a neurocomputational model of object processing in the ventral visual pathway, and analyze the activation patterns in the model using a method we have termed “virtual MVPA”. We account for data from both H01 and SK02, providing an explanation for both sets of findings within a single framework. The model contains no specialized mechanism for the processing of faces, casting doubt on the interpretation of the findings in SK02 as evidence in favor of a “face module”.

Object representations in the model are organized topographically, which can be related to the topographic organization of stimuli seen in much of visual cortex. Because of this property of the network, we are able to analyze model behavior with a method analogous to the analysis of activation patterns across voxels in MVPA of fMRI data. More specifically, we sort hidden layer units into those that are “selective” for faces, houses, chairs and so on, allowing breakdown of the activation patterns by region, as was carried out in H01 and SK02.

## A Neurocomputational Model of Object Processing in Visual Cortex

The model architecture is shown in Figure 1. Input images are first subject to Gabor wavelet filtering and Principal Components Analysis (PCA); these stages are designed to mimic the processing in early visual cortex.

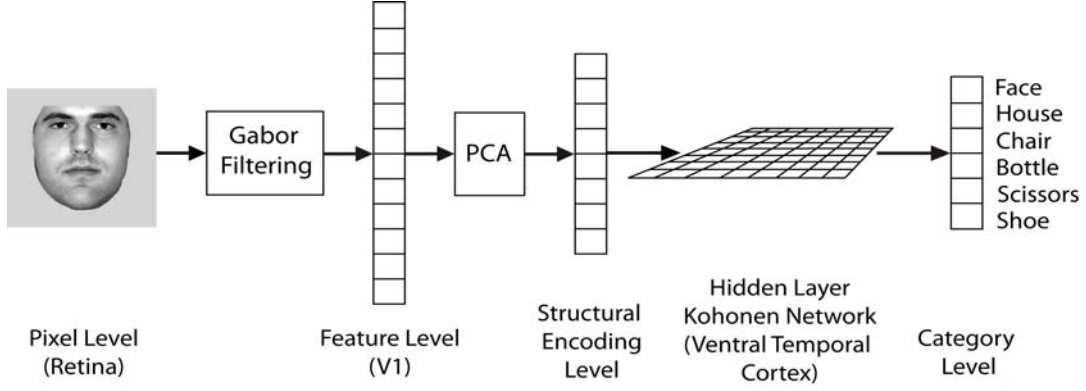


Figure 1. The Model Architecture.

Pre-processed stimuli are then input to a Kohonen layer, in which topographically organized representations of objects develop through a self-organization process. Finally, representations in the topographic layer are associated with output nodes corresponding to the six classes of object, via weights that are learned using the delta rule. The use of a Kohonen network at the hidden layer is a departure from our previous models of face and object processing (Dailey & Cottrell, 1999; Tong et al., 2008). Whereas these models used a hidden layer trained with the backpropagation algorithm, learning in the present model is unsupervised and neurophysiologically plausible. Kohonen networks are designed to model cortex; the learning algorithm is a computational abstraction of cortical mechanisms such as Hebbian learning and lateral inhibition. The stimulus representation layer in a Kohonen network is typically a two-dimensional array in which all units occupy a fixed position with respect to their neighbors. The network is trained by incremental update of the weights from input to hidden units over successive presentation of stimuli; the weight update is governed by a neighborhood function ensuring that neighboring units receive similar updates in response to a given stimulus, with the result that neighboring units come to represent items that are close together in stimulus space. These properties make Kohonen networks suitable for computational investigations of object processing in visual cortex (see, e.g., Cowell et al., 2006).

### Visual Categorization Training

**Stimuli** We used stimuli from the fMRI study of H01: grayscale photographs of faces, houses, chairs, bottles, scissors and shoes (we did not use the cats or scrambled images from H01). Images were scaled and cropped to 64x64 pixels. There were 240 training images: 40 in each category, 4 views of 10 exemplars. In addition, 48 images were assigned to a “hold-out” set: 8 per category, 4 views of 2 individuals. The hold-out set was reserved for testing networks’ classification performance after every 10 epochs.

**Image Pre-processing** Before presentation to the neural network, stimuli were subject to Gabor filtering to extract representations suitable for object recognition (Dailey and Cottrell, 1999). We applied the Gabor filters to a 32x32 grid of points in each image resulting in a vector of size 40,960

(32x32 sample points, with 8 orientations and 5 scales of the filter at each point).

These patterns were reduced in dimensionality using Principal Components Analysis (PCA), to make them suitable for input to a neural network. PCA was performed on all 288 images used in the present study. We used the first 20 components and did not normalize them to have equal variance, a difference from previous work that enabled better Kohonen network learning.

**Training** Networks were trained to classify objects into one of six categories. Note that learning of the feed-forward weights from the input layer to the topographic hidden layer was independent from learning of the weights from the hidden layer to the output nodes. Sigmoidal units were used throughout. Eight networks were trained for 10 epochs at a time, and then checked for performance on the hold-out set. The only effect of the output layer on the Kohonen layer was that training ended when hold-out set accuracy at the output was >90% in 3 successive presentations.

The input-Kohonen layer weights were trained by the standard Kohonen learning rule:

$$w_{ji}(t+1) = w_{ji}(t) + f(dist) * (a_i - w_{ji})$$

where  $w_{ji}$  is the weight from input  $i$  to unit  $j$ ,  $a_i$  is the activation of input  $i$ , and  $f$  is a neighborhood weighting function on the learning rate, centered on the most highly-activated unit ( $dist$  is the distance from unit  $i$  to the maximally activated unit in the grid).  $f$  takes the form:

$$f(dist) = \eta e^{-\left(\frac{dist}{G}\right)^2}$$

in which  $\eta$ , the learning rate, starts at 1 and reduces over epochs according to  $\eta = \text{epoch}^{(-0.2)}$ .  $G$  is a Gaussian width parameter, which also decreases over epochs, according to  $G = 0.5 + 10 * \text{epoch}^{(-0.3)}$ , but stops reducing at epoch 50.

This neighborhood learning rule ensures that nearby units learn similar patterns, which is what gives the Kohonen network its map-like quality. The output units were trained by a simple delta rule, with learning rate 0.01. We used 2500 units on the Kohonen layer for these simulations.

### MVPA Analyses

We simulated four findings from the fMRI literature: three demonstrations from H01 and one finding from SK02. For

all MVPA analyses, we examined activation patterns in the hidden layer of the network. For each of the eight trained networks, we recorded the activation patterns elicited by presentation of all stimuli in the 240-image training set, using the input-Kohonen weight values that the network had attained at the final epoch of training. Since the output units do not affect hidden unit representations in the model, the output weights were simply used to ensure that the network had formed a useful representation of the data.

**Haxby et al. (2001) MVPA Methods** Subjects viewed photographic images from seven different categories: faces, houses, cats, bottles, scissors, shoes and chairs, plus scrambled “nonsense” images, while patterns of response were measured with fMRI. The authors first identified object-selective cortex by selecting those voxels for which the response to the different object categories that subjects viewed differed significantly. Data for each subject were split into two sets: odd and even numbered scans. By examining the similarity of patterns of response evoked by each category on odd and even scans, the authors determined whether the category being viewed by the subject could be identified, and measured the discriminability of the responses to different categories.

Similarity was measured by the correlation between responses in odd and even runs. To determine the discriminability of, say, faces and shoes, the correlation between the mean response to faces across odd runs and the mean response to faces across even runs was compared to the correlation between the mean response to faces on odd runs and the mean response to shoes on even runs. If the within-category (face-face) correlation was higher than the between-category (face-shoe) correlation, that pairwise comparison was deemed correct. In comparing faces and shoes, there are four pairwise comparisons to be made: 1.  $\text{face}_{\text{odd}}\text{-face}_{\text{even}}$  vs.  $\text{face}_{\text{odd}}\text{-shoe}_{\text{even}}$ , 2.  $\text{face}_{\text{odd}}\text{-face}_{\text{even}}$  vs.  $\text{face}_{\text{even}}\text{-shoe}_{\text{odd}}$ , 3.  $\text{shoe}_{\text{odd}}\text{-shoe}_{\text{even}}$  vs.  $\text{face}_{\text{even}}\text{-shoe}_{\text{odd}}$ , 4.  $\text{shoe}_{\text{odd}}\text{-shoe}_{\text{even}}$  vs.  $\text{face}_{\text{odd}}\text{-shoe}_{\text{even}}$ . Thus, for each discrimination of two categories, there were four binary comparisons, yielding a possible score of 0, 25%, 50%, 75% or 100% on that pairwise discrimination, for every subject.

H01 presented all photographic images to subjects several times each. However, the manner of assigning photographs to scans and sorting data from scans (by odd or even number) into two halves meant that the evoked activation patterns that were pooled by averaging over each half of the data were not elicited by an identical set of images. So the within-category correlation values – derived from activation to, e.g., faces in the first half and faces in the second half – were partly determined by the reproducibility of brain responses to different images from the same category.

**Spiridon and Kanwisher (2002) MVPA Methods** The methods were broadly the same as in H01. However, SK02 assayed different ways of splitting the data into two halves. In the ‘identical images’ condition, images were allocated to scans and sorted into two halves in such a way that each half

of the data contained responses evoked by an identical set of images. In the ‘different views’ condition, scans were sorted such that each half contained responses evoked by the same exemplars, but different views appeared in each half. In the ‘different exemplars’ condition, responses to a given category in each half of the data were evoked by images of different exemplars. When SK02 performed the same type of correlational analysis as H01, they found no significant differences between the three conditions in the percentage of correct pairwise category discriminations.

**Virtual MVPA Methods** There is considerable noise in fMRI BOLD responses, which can be attributed to both endogenously-generated noise within the brain responses of subjects and externally-generated noise such as variability in scanner measurements. Owing to these sources of noise, correlations between neural responses reported by H01 were never perfect; even for within-category correlations, mean values ranged from 0.28 (bottles) to 0.81 (houses), measured across all of object-selective ventral temporal cortex. In contrast, there is no noise in the activation values of our model: we can access the exact activation evoked in a given unit by a given pattern at the time of test. If we present an image to the network twice, we obtain the identical activation pattern on both occasions. Presentation of exactly the same images in the two halves of the data, as in the ‘identical views’ condition of SK02, would invariably yield within-category correlations of 1. In order to exploit the within-category variability in the model of activation patterns due to different images (and avoid perfect within-category correlations), we followed the methods of H01 and the ‘different views’ condition of SK02: we sorted our activation patterns so that the two halves contained images from different views. Thus, both within-category and between-category correlations ( $r_{\text{within}}$  and  $r_{\text{between}}$ ) were free to vary between -1 and 1, depending on the degree of similarity of the two activation patterns being compared.

In practice, because of the lack of noise in the model, within-category correlations were always higher than between category correlations, therefore performing four binary comparisons, as in H01 and SK02, would always yield scores of 100%. Therefore we devised a measure of pairwise discrimination accuracy using the within- and between-category  $r$  values and the Luce Choice rule:

$$P(\text{correct}) = \frac{e^{\beta * r_{\text{within}}}}{\sum_r e^{\beta * r}}$$

where the sum over  $r$  means  $r_{\text{within}}$  and  $r_{\text{between}}$ , and  $\beta = 2$  (the  $\beta$  value was chosen for discrimination of 95% on fictitious high within-category and low between-category  $r$ ’s before applying to the network data). Accuracy of discrimination was equated with probability of the correct choice.

## Results

All eight networks trained on the classification task reached criterion in 50 - 330 epochs. Figure 2 shows the topographic

organization of stimulus representations that emerged on the Kohonen layer for a typical simulation.

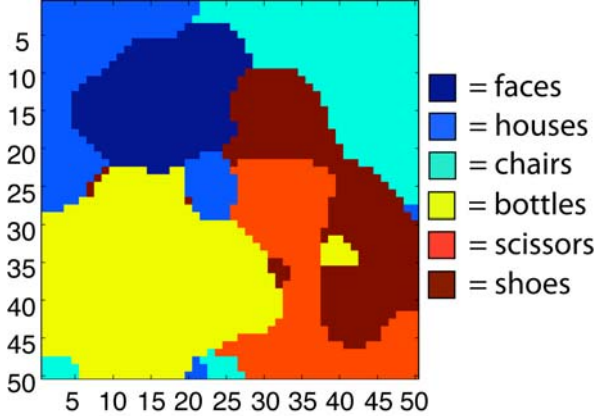


Figure 2. Bird's eye view of the Kohonen network hidden layer. Each hidden unit is represented by a colored square; the color indicates which category maximally activated the unit. The position of a unit in the 2D grid is shown on the x- and y-axes.

**Haxby Simulation 1** H01 found that patterns of response across all object-selective voxels in ventral temporal cortex contained sufficient information to perform pairwise discriminations of object categories. The correct category was identified in 96% of pairwise comparisons.

Like Haxby et al., we also found that correlations between activation patterns on the hidden layer of our model correctly identified the category being viewed, with a mean probability of 86% that the correct category would be chosen in a pairwise comparison (see Table 1). This result was obtained without optimizing the constant in the Luce Choice Rule. Interestingly, H01 found that faces and houses were best discriminated by human subjects' activation patterns, and of the categories on which we trained our model (i.e., excluding cats), they reported that chairs were the next best discriminated, with scissors, shoes and bottles performing more poorly. We report the exact same trend, with faces and houses best discriminated, followed by chairs, and then by the other three categories. In H01, there were two further analyses demonstrating that activations patterns are distributed across cortex; here we simulate both.

**Haxby Simulation 2** H01 reported that information about a given category is not solely contained in voxels in regions that respond maximally to stimuli from that category. They showed this by measuring accuracy of pairwise category

discriminations using responses across object-selective cortex from which the voxels maximally responsive to the two categories being discriminated had been removed. For example, when discriminating houses and shoes, they removed the house- and shoe-selective voxels from the analysis. They found that the category being viewed was correctly identified in 94% of pairwise comparisons – a score barely diminished from that obtained using all object-selective voxels. We replicated this analysis using activation patterns from the hidden layer in which all units maximally activated by either category in the discrimination were removed. Like H01, we found that the mean probability of correct identification of a category in a pairwise discrimination was 87%, similarly undiminished from the case where all object-selective units were used. The results of this analysis are shown in line 2 of Table 1.

**Haxby Simulation 3** Since voxels outside of the region responding maximally to a category contained information about that category (shown in H01 finding of Simulation 2), Haxby et al. proposed that each category-selective region contains information about other categories, too. To investigate this, they examined the discriminability of all object categories in each region that was maximally (and differentially) activated by one category, i.e., in each “category-selective” region. H01 showed that, within only cortex that responded maximally to one object class, good identification of all categories was possible. They reported mean accuracies for all pairwise category discriminations ranging from 83% to 94% across the different regions selective for faces, houses, or small man-made objects.

We replicated this analysis by using activations from each “category-selective” region of the hidden layer for pairwise discriminations of all categories. The results are shown in lines 3-5 of Table 1. Like H01, we found that accuracy values for discriminations based on these regions of the hidden layer were similar to those obtained in Simulations 1 and 2, with mean probability of correct identification ranging from 83% to 85% across the face, house and small man-made object regions of the hidden layer.

The results of Analyses 2 and 3 from H01 suggested distributed cortical representations. SK02 performed similar analyses to H01, but in doing so, they found results that they interpreted as evidence for specialized processing of faces. Hence, we simulate those results here.

Table 1: Simulation of Haxby et al. (2001) results. Accuracy of identification of the category being viewed based on activation patterns in the topographic hidden layer of the model.

| Region of hidden layer  | Identification accuracy |            |            |            |            |            |
|---|-------------------------|------------|------------|------------|------------|------------|
|   | Faces                   | Houses     | Chairs     | Bottles    | Scissors   | Shoes      |
| All object-selective units                                      | 91.4 ± 0.2              | 87.9 ± 0.2 | 87.5 ± 0.3 | 85.4 ± 0.4 | 82.3 ± 0.2 | 82.1 ± 0.2 |
| Minus regions maximally responsive to categories being compared | 92.0 ± 0.4              | 87.7 ± 0.2 | 89.2 ± 0.5 | 86.8 ± 0.4 | 82.2 ± 0.3 | 82.2 ± 0.3 |
| Regions maximally responsive to:                                |                         |            |            |            |            |            |
| Faces   | 93.6 ± 0.2              | 83.1 ± 0.5 | 84.8 ± 0.5 | 84.5 ± 0.9 | 79.7 ± 0.6 | 79.1 ± 0.4 |
| Houses  | 87.6 ± 0.4              | 89.9 ± 0.3 | 87.8 ± 0.3 | 82.1 ± 0.5 | 80.9 ± 0.4 | 80.9 ± 0.3 |
| Small Objects   | 85.1 ± 1.0              | 84.7 ± 1.2 | 84.5 ± 1.4 | 81.7 ± 0.9 | 81.2 ± 1.5 | 80.2 ± 0.8 |

**Simulation 4: Spiridon & Kanwisher (2002)** One finding from SK02 was particularly pertinent to the debate regarding a specialized module for faces. This finding emerged from a replication of H01’s Analysis 3, but in which the size of the cluster of voxels used in the MVPA was equated across categories. The authors measured the accuracy of pairwise category discriminations between activation patterns across clusters of 30 voxels most selective for each category. They found that the cluster of voxels most selective for faces yielded better accuracy in discriminating face from non-face categories than in discriminating non-face from non-face categories. Spiridon and Kanwisher took this as evidence for specialized processing of faces by face-selective neurons. Since no other cluster of voxels exhibited such an advantage for discrimination of the category to which it responded maximally, the authors concluded that this specialized processing does not exist for any other object category.

The results from our replication of this analysis, using 30-unit clusters, are shown in Table 2. Using the same methods of analysis as SK02, we also find that, in the cluster of units most selective for faces, faces are better discriminated than any other category. As in SK02, this is true for no other category.

The relationship between the number of units in our simulation and the number of voxels in the brain is unclear; there is no reason to expect a one to one mapping between voxels and hidden units. Therefore we performed analyses using cluster sizes of 60 and 100 units. We found that the simulation results generalize across different spatial scaling assumptions; that is, for both 60- and 100-unit clusters we found the same selective advantage for discriminating faces in the face cluster, and no such advantage for discriminating the category to which the cluster is maximally responsive for any other cluster.

## Discussion

In this paper we presented a neurocomputational model of visual object processing in which representations of objects on the hidden layer develop, without supervision, into a topographically organized map of stimulus space. Learning of hidden layer representations is governed by the Kohonen algorithm, which mimics processing in mammalian sensory cortex. The topography of object representations in the hidden layer allowed us to develop a method we term

“virtual MVPA”, in which we analyze patterns of activation across units in the hidden layer with statistical techniques similar to those used in MVPA of fMRI data.

We have replicated three findings from H01 and an important result from SK02. In the replications of H01, we demonstrated that activation patterns across units in the model can be used to determine the category of stimulus being presented to the model, under three conditions: (1) using all object-selective units in the discrimination, (2) using all object-selective units minus those that are maximally activated by the categories being discriminated, and (3) using only the units maximally responsive to a single stimulus category. In the simulation of SK02, we showed, using a model containing no specialized processing for faces, that the ‘face region’ shows an advantage for category discriminations involving faces (i.e., identifying that an object is a face) but this is not true for other category selective regions. In other words, the region maximally responsive to shoes is not best at shoes.

These simulations showcase our virtual MVPA method, demonstrating that introducing topographic representations into a model of face and object processing allows analysis of the model in the way that brain scan data from human subjects is analyzed using MVPA. Further, our successful simulation of H01 demonstrates that we have a neurocomputational model of visual object processing in which representations contain the type of information about object categories that is contained in neural representations in ventral temporal cortex, as assessed by fMRI.

The three findings from H01 support an account of visual processing in which there is no face module, since they suggest that information about other categories exists in the face area, and information about faces is contained in other areas. For example, Haxby Simulations 2 and 3 demonstrate that a unit maximally activated by bottles is still useful in discriminating, say, faces from houses, because it responds consistently with characteristic (albeit non-maximal) activation values to face and house stimuli. Information about stimulus identity is not solely carried by the units responding maximally to the stimulus. However, Spiridon and Kanwisher performed a similar analysis to Haxby Simulation 3 and found that information was *not* equally distributed across cortex for all categories. For this reason, we simulated SK02.

Table 2: Simulation of Spiridon and Kanwisher (2002) results. Accuracy of identification of the category being viewed based on activation patterns across the 30 units most selective for each category.

| Region of hidden layer       | Identification accuracy |                |                |                |                |                |
|------------------------------|-------------------------|----------------|----------------|----------------|----------------|----------------|
|                              | Faces                   | Houses         | Chairs         | Bottles        | Scissors       | Shoes          |
| 30 units most selective for: |                         |                |                |                |                |                |
| Faces                        | 90.3 $\pm$ 2.1          | 80.4 $\pm$ 1.6 | 82.7 $\pm$ 2.5 | 84.0 $\pm$ 1.0 | 76.2 $\pm$ 1.7 | 74.1 $\pm$ 2.2 |
| Houses                       | 85.6 $\pm$ 1.1          | 83.1 $\pm$ 2.0 | 85.8 $\pm$ 1.7 | 79.3 $\pm$ 1.2 | 77.2 $\pm$ 1.6 | 81.0 $\pm$ 1.5 |
| Chairs                       | 87.2 $\pm$ 2.1          | 84.8 $\pm$ 1.6 | 85.6 $\pm$ 1.6 | 77.3 $\pm$ 1.7 | 77.6 $\pm$ 2.9 | 75.9 $\pm$ 1.2 |
| Bottles                      | 90.9 $\pm$ 1.5          | 81.7 $\pm$ 1.4 | 82.0 $\pm$ 1.7 | 82.9 $\pm$ 1.2 | 83.4 $\pm$ 1.5 | 86.3 $\pm$ 1.0 |
| Scissors                     | 84.6 $\pm$ 2.0          | 81.6 $\pm$ 1.4 | 84.8 $\pm$ 1.8 | 86.4 $\pm$ 1.5 | 81.4 $\pm$ 1.2 | 82.9 $\pm$ 2.4 |
| Shoes                        | 85.8 $\pm$ 1.7          | 86.4 $\pm$ 0.6 | 84.8 $\pm$ 1.7 | 86.9 $\pm$ 0.9 | 86.0 $\pm$ 2.7 | 79.6 $\pm$ 2.5 |



We replicated the SK02 finding that there is no cluster of voxels *other than for faces* that selectively discriminates one object category from the alternatives. We were able to replicate this finding simply because the similarity structure of the representations in the model reflects the similarity structure of the stimuli (echoing the match between the similarity of brain scans and the similarity of the images eliciting the brain scans revealed by O’Toole et al., 2005). Among the object categories, the images of faces show the greatest within-category similarity and the greatest dissimilarity from other categories; this is also true of the representations of faces in the model (Figure 3A). This means that individual face stimuli elicit highly reproducible patterns (yielding a high within-category correlation for face activations) that are quite different from the patterns due to other objects (yielding low between-category correlations for face vs. non-face comparisons). This is particularly true in the region of the hidden layer maximally responsive to faces, where all units are activated maximally by a face, but yield highly variable activations in response to other objects. Thus in the face region, the face-face correlation is high and the face-nonface correlation is low (giving good face discrimination) but, e.g., the shoe-shoe correlation is not high (giving poor shoe discrimination). By contrast, the within-category similarity of other categories is much lower, so that individual scissor stimuli elicit very variable responses (Figure 3B), even in the region maximally responsive to scissors. Here the within-category correlation for activation patterns due to scissors is not high, and the between-category correlation for scissors compared to, say, shoes, may be moderate (since the average activation over a variable set of scissor patterns may be similar to the average over a variable set of shoe patterns), therefore discrimination of scissors is poor.

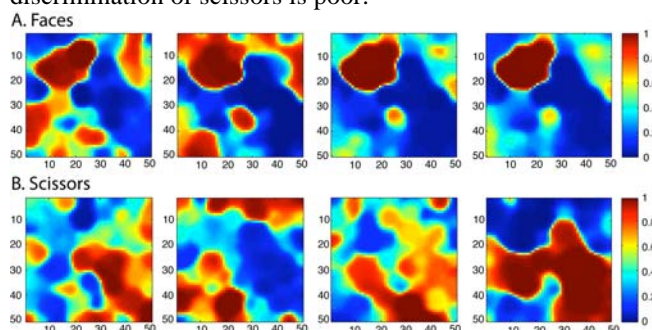


Figure 3. Kohonen layer activations evoked by four individual face stimuli (top row) and four individual scissor stimuli (bottom row). Individual exemplars were chosen randomly. Compare to Figure 2.

In simulating the SK02 result, we have helped resolve the inconsistency of this finding with Haxby et al.’s interpretation of their own results within a distributed processing account. In essence, using a neurocomputational model that has no special anatomy or processing mechanism for faces, we have accounted for MVPA results that both suggest the face specific region of cortex is special (because only the face area is better at discriminating faces from the alternatives) and that faces are not special (because they can be discriminated in regions not selective for faces). This

demonstration was made using equal amounts of training for all object categories (that is, ignoring any influence of the social significance of, or extra exposure to, faces). The success of this computational account suggests that the differences between faces versus other objects may lie primarily in the visual characteristics of faces themselves as objects (O’Toole et al., 2005). Moreover, since our virtual MVPA findings map onto the results of real brain scan MVPA, we provide an existence proof that one can obtain the SK02 results from a model that has no specialized processing for faces. This calls into question use of these results in arguing for a specialized “face module” in ventral temporal visual cortex.

## Acknowledgments

RAC was supported by the RCUK, a BBSRC ISIC Grant and a Royal Society travel grant. GWC was supported by NSF grant #SBE0542013 to the Temporal Dynamics of Learning Center.

## References

- Cowell, R.A., Bussey, T.J., & Saksida, L.M. (2006). Why does brain damage impair memory? A connectionist model of object recognition memory in perirhinal cortex. *J Neurosci.*, **26**(47):12186-12197.
- Dailey, M. & Cottrell, G.W. (1999) Organization of face and object recognition in modular neural network models. *Neural Networks*. **12**(7-8) :1053-1074
- Gauthier, I., Tarr, M.J., Anderson, A.W., Skudlarski, P. & Gore, J. C. (1999) Activation of the middle fusiform “face area” increases with expertise in recognizing novel objects. *Nat Neurosci.*, **2**:568-573.
- Grill-Spector, K., Knouf, N. Kanwisher, N. (2004) The fusiform face area subserves face perception, not generic within category identification. *Nat Neurosci.*, **7**:555-562.
- Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., & Pietrini, P. (2001). Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. *Science*, **293**:2425-2430.
- Kanwisher, N., McDermott, J. & Chun, M.M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci.*, **17**, 4302-11.
- Kohonen, T. (1984). *Self-organization and associative memory*. Springer-Verlag, Berlin.
- O’Toole, A.J., Jiang, F. Abdi. H. & Haxby, J.V. (2005). Partially Distributed Representations of Objects and Faces in Ventral Temporal Cortex. *J Cog Neuro*, **17**(4):580-590.
- Spiridon, M., & Kanwisher, N. (2002). How distributed is visual category information in human occipito-temporal cortex? An fMRI study. *Neuron*, **35**:1157-116.
- Tarr, M.J. & Gauthier, I. (2000) FFA: A flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nat Neurosci.*, **3**:764-769.
- Tong, M.H., Joyce, C.A., & Cottrell, G.W. (2008) Why is the fusiform face area recruited for novel categories of expertise? A neurocomputational investigation *Brain Res.* **1202**:14-2.