Are Greebles Special?
Or, why the Fusiform Fish Area would be recruited for Sword Expertise
(If we had one)

Matthew H. Tong (mhtong@cs.ucsd.edu)
Carrie A. Joyce (cjoyce@cs.ucsd.edu)
Garrison W. Cottrell (gary@cs.ucsd.edu)
UCSD Department of Computer Science and Engineering,
9500 Gilman Drive, Dept. 0123
La Jolla, CA 92093-0114 USA

Abstract

The fusiform face area (FFA) has commonly been deemed an area specialized for face processing. Many recent studies have challenged this view by showing that the FFA also responds to stimuli from domains in which the subject is an expert. We have developed neurocomputational models to explore the question of why an area initially devoted to face processing would later be recruited for other domains. Previous studies showed that the FFA could be a fine-level discriminator, spreading apart similar stimuli in its representational space. These characteristics would make it ideal for recruitment to other fine-level discrimination tasks. These initial findings have been challenged on several accounts. Here we introduce new work showing that the expertise effect remains despite additional controls on the type and difficulty of the task presented.

Keywords: Perceptual expertise; connectionist models; face recognition; object classification

Introduction

As its name suggests, the fusiform face area (FFA) has been identified as an area that is specific to processing faces. Imaging scans show increased activity of this area when face stimuli are presented when compared to activation when non-face stimuli are presented (Kanwisher, McDermott & Chun, 1997). Damage to the FFA tends to lead to prosopagnosia (an inability to recognize faces) and similar deficits, further implicating this area in facial recognition (De Renzi et al., 1994). There is also electrophysiological evidence in the form of a maximum amplitude of a negative going wave about 170 ms following a stimulus (the N170) that shows that faces may be handled differently at a neurological level (Eimer, 2000). Based on this evidence and psychological evidence showing that faces are processed differently at a behavioral level, many have concluded that the FFA is a face-specific processing module.

The view of the FFA as a module for face processing has been challenged by evidence reporting that the fusiform face area seems to be recruited for other tasks as well. Experts in recognizing cars, birds, dogs, and Greebles (a class of artificial objects developed by Isabel Gauthier) also show increased activation of the fusiform face area (Gauthier et al., 1999; Gauthier et al., 2000). Furthermore, visual experts show an increased amplitude N170 when shown images from their field of expertise when compared to items for which they are not experts (birds for bird experts, versus dogs for dog show judges) (Tanaka & Curran, 2003). In these studies, subjects are defined as being experts when their reaction times are as fast when verifying the category of an object at the subordinate or individual level as they are when verifying the category at the basic level.

These findings suggest a different account of the function of the FFA, that the area is actually a subordinate-level, fine-grained visual discrimination area. Humans become face experts at a young age, and later in life they may pick up additional areas of visual expertise as well. Under this account, earlier studies showing the FFA being highly active only in face processing did so because they did not choose non-face stimuli with which the subjects were experts. Recent studies (Grill-Spector, Knouf, & Kanwisher, 2004; Rhodes et al., 2004; Yovel & Kanwisher , 2004) that purport to refute the expertise claim never seem to achieve the stringent reaction time requirement of expertise (defined above) in their work. For example, although Grill-Spector et al. recruited car experts, they then tested their subjects on antique cars, which would presumably not be named quickly by their "experts." Indeed, this would seem to be an example of an "other race effect" for cars, as it is already known that other race faces also produce a lower response in the FFA (Golby et al., 2001).

In any case, if one assumes the expertise hypothesis, this leads to the question of what characteristics do these different expert-level tasks share that would cause one area of the brain to be used for them but not for basic-level tasks. It has been suggested that the FFA implements a process of fine level discrimination (Tarr & Gauthier, 2000). As we have shown elsewhere (Joyce & Cottrell, 2004), and elaborate on here, we believe that the main feature of this process is a transformation that takes similar visual items and magnifies the differences between them. This transformation generalizes to novel domains, and tuning of the transform allows the model FFA to acquire new expertise faster than an area that simply categorizes objects at the basic level (e.g., Lateral Occipital Complex (LOC)). This suggests that in a competition between cortical areas for processing a new task, the FFA would have a distinct advantage, thus explaining why the FFA would become recruited for novel expertise tasks – it simply learns them faster than an area that only

performs basic level processing. Computational models allow us the ability to examine such issues in more detail than can be achieved in human subjects, or even in monkeys. The main prediction such models make is that the variance of the responses of neurons in the FFA will be greater to objects of expertise (Sugimoto & Cottrell, 2001; Joyce & Cottrell, 2004).

These early models only examined the effects of expertise when learning Greebles, a class of objects designed specifically to have some special properties. Here we replicate and expand the earlier work, showing that the effect is not unique to Greebles. The original experiments also had the expert networks performing a greater number of discriminations and a more difficult task. Hence there were a larger number of outputs in the FFA model network compared to the basic level categorizer. Perhaps it was simply the number of distinctions being drawn by the expert networks that made them better at differentiating new categories (Michael Tarr, personal communication). In more recent work, Tran, Joyce, and Cottrell (2004) controlled for the number of classes being identified in a letter/font task, showing that the effect remained. However, the domains were letters rather than object categories, and the relative difficulty of the two tasks, recognizing letters versus recognizing fonts, was not controlled. This paper introduces a new set of controls and stimuli that address these concerns.

In what follows, we first report on an experiment that suggests there is nothing special about Greebles as a novel expertise task; it turns out that an area that is a cup expert is faster at learning faces than one that is not. The second experiment demonstrates that the expertise advantage persists when the basic and expert tasks are equal in their difficulty and number of classifications being made.

Experiment One

The stimulus set consisted of the 300 64 x 64 8-bit grayscale images comprising five basic-level classes (books, cans, cups, faces, and Greebles) that were used in previous experiments (Sugimoto & Cottrell, 2001; Joyce & Cottrell, 2004). Each basic-level class of sixty images had twelve subordinate level categories, each composed of five instances of that class. Each subordinate class of faces was composed of images of a human face producing different facial expressions. Subordinate classes from other classes had individual instances formed by rotating an image by as much as 3 degrees and shifting in a cardinal direction by a pixel. In earlier work Greebles were set aside for phase two training, during which the networks learned them as novel area of expertise (Joyce & Cottrell, 2004). To eliminate the possibility that the observed effects were unique to Greebles, the experiment was repeated with each of the object classes serving as the phase two stimulus set.

Images went through several stages of preprocessing before being presented to the networks. First they were run through a bank of Gabor wavelet filters of five different scales and eight orientations to approximate the processing that occurs in visual brain area V1. The result is z-scored and then submitted to principal component analysis in order to reduce the dimensionality to forty PCA projections. It should be noted that the phase two stimulus objects were not included in the PCA in order to have them be truly novel to the model. The results of the PCA were z-scored again before being fed to the networks. The networks themselves were standard feed-forward networks trained using back-propagation and initialized with random weights. There were forty input neurons (one for each principal component), sixty hidden units, and one output unit for each basic or subordinate level category. The learning rate was set to 0.01 with a momentum of 0.5 and the networks were initialized with random weights.

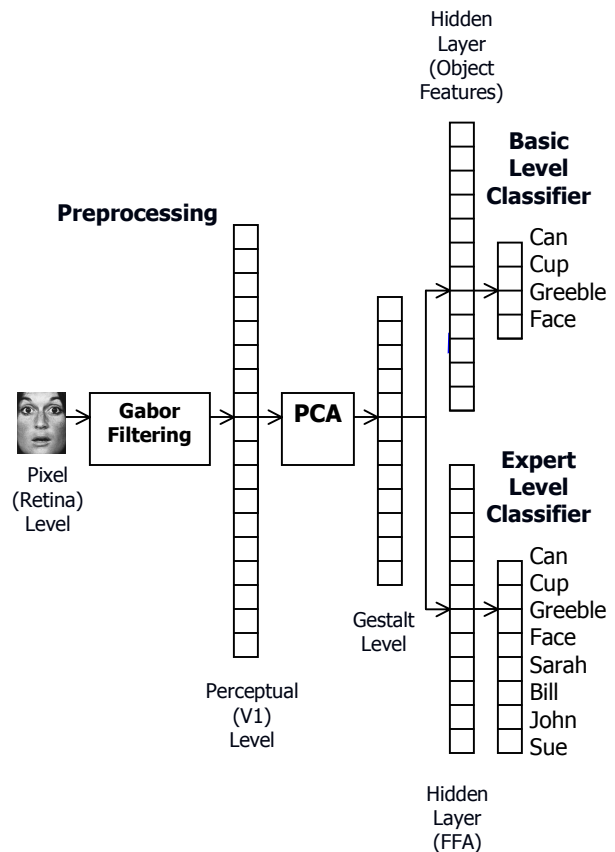Figure 1: Sample stimuli for experiment 1

Figure 2: Network architecture

During phase one, all networks were trained to discriminate between the four basic-level classes (ex. book, can, cups, faces). Expert networks were additionally trained to discriminate ten subordinate level classes for one of the four

phase one classes (ex. Bob, Carol, Anne). Thus, basic-level networks had 4 output units, while expert-level networks had 14. Five networks of each type were trained, making 25 networks in all. Training continued well past the point of leveling off, to 5120 epochs of training. At various points in the training, the network's weights were saved (after 1, 10, 20, 40, 80, 160, 320, 640, 1280, 2560, and 5120 epochs).

During phase two, the weights of each of these saved networks were used as the initial weights of a network that was then trained on the phase two stimuli at the subordinate level. This required an additional 11 output neurons for the networks (1 for basic-level phase two discrimination, 10 for the subordinate-level), The total number of neurons during phase two were therefore 15 for basic-level networks (5 basic-level categories, 10 subordinate level categories) and 25 for expert networks (5 basic-level categories, 20 subordinate-level categories). Training with the set of phase one stimuli continued during phase two. The second phase continued until error rates fell below .05 for the phase two stimuli.

Results revealed that basic-level discriminations were learned more quickly than expert ones during phase one, showing that the expert-level tasks were more difficult (see Figure 3). More interestingly, however, the expert-level networks learned the new expert-level task in less time during phase two than the basic-level networks (see Figure 4). Although this was true even when the amount of training occurring in phase one was minimal or nonexistent (due to the fact that phase one training is still ongoing in phase two), the advantage increased as the amount of phase one training increased. These results follow the pattern displayed in earlier studies, showing that the effect is not unique to Greebles.
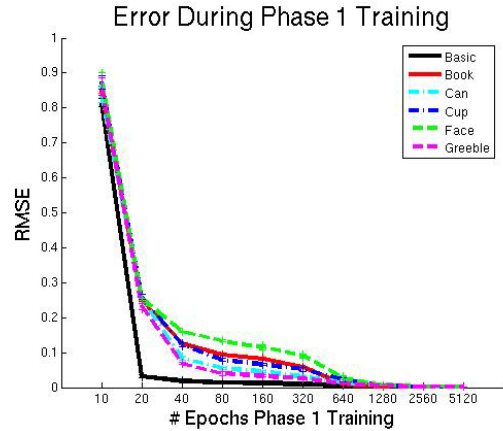

Figure 3: Error rates during phase 1 as a function of time

Experiment 2

Several factors could be influencing the results of the first experiment. The subordinate level discriminations being made in this experiment are simply more difficult, possibly causing the phase one training of expert networks to extract more discriminating information, which then generalizes to the new categories. Secondly, the number of discriminations being made by the expert is far larger than the basic-level network (14 categories versus 4). Finally, the expert-level network is currently being trained to make all the basic-level distinctions in addition to their expert-level specialty.
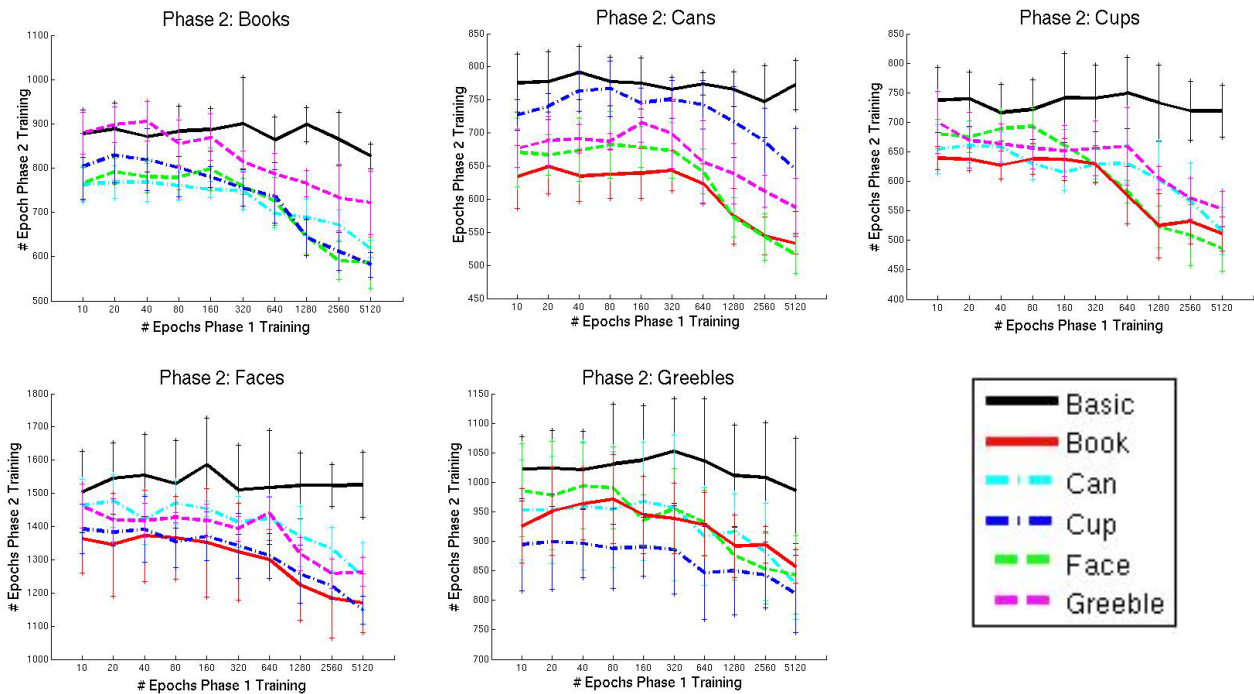

Figure 4: Time to learn new task (books, cans, cups, faces, and Greebles) as a function of phase one training. Note that the basic-level network (denoted by black) is consistently slower to learn the new task.

Figure 5: Six of the object classes for experiment two (cars, glasses, lamps, hats, chairs, and fish)

To address these concerns, a new data set was constructed that would have additional controls in place. This new data set consisted of 13 basic level object classes (balls, brass instruments, cars, chairs, doughnuts, fish, fruits, glasses, guitars, hats, lamps, ships, and swords). Each basic class was composed of 13 subordinate-level classes with 5 instances per subordinate class. Images were gathered from Hemera.com and were shrunk to 64 x 64 pixels, grayscaled, and normalized for luminance and contrast. As in the earlier experiment, object instances were made by rotating the objects by up to 3 degrees and shifting by up to a pixel. Three of these classes (lamps, ships, and swords) were set aside for use during phase 2. The remaining 10 classes were used in phase one training for basic classification. The set of faces used in the first experiment were re-used for expert-level training.

The basic network architecture and training procedures were identical to those described in the first experiment. Basic networks had one output unit corresponding to each of the ten basic-level classes. Expert networks had one output unit corresponding to each of ten faces. To model similar early stages of processing, the principle components were calculated with the combined data set but omitting the three object classes reserved for phase two. The total number of images in the basic and expert data sets was the same, but
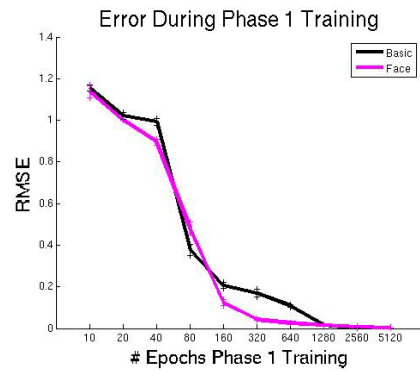

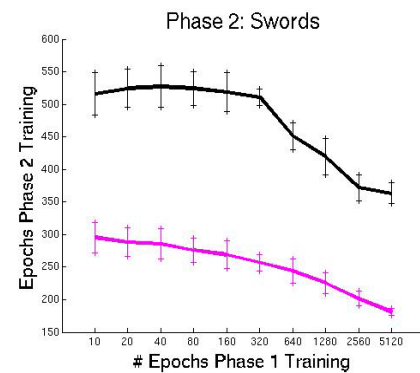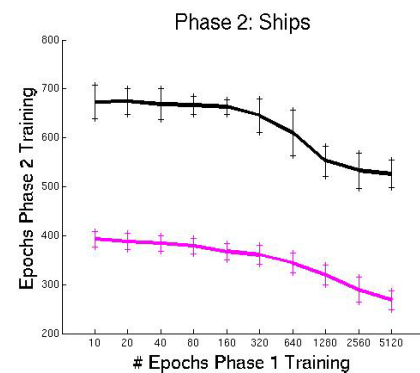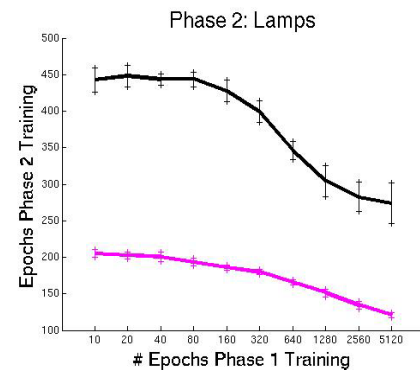Figure 6: Error rates during phase 1 as a function of time

Figure 7: Time to learn new task (lamps, ships, or swords) as a function of phase one training. Basic-level networks' performance is shown in black, while expert networks' performance is shown in pink.

2199

during training a network was only exposed to one of the data sets. During phase two, an additional ten output units were added to discriminate between ten subordinate-level classes of a new class (lamps, ships, or swords), bringing the total number of outputs for both basic-level and expert networks to twenty. As in the previous experiment, training on the phase one stimuli was continued in phase two, with training ending when error rates on the phase two stimuli fell below a threshold.

In this experiment, the two types of networks were matched in the number of discriminations they must make. Expert networks are no longer trained to do the basic level task in addition to the expert level one. The number of classes is larger and contains objects unlikely to be considered face-like. Comparing Figures 3 and 6, we see that the two tasks are now approximately matched in their difficulty. Despite these additional controls, the expertise effect remained strong (Figure 7).

Analysis

Examining the activation of the hidden units, there is a clear increase in the within-class variance for the expert-level networks compared to the basic-level networks. Intuitively, this makes sense, as an expert network must spread out its internal representation to differentiate between similar exemplars. A basic-level network, on the other hand, tries to map differing images into the same class, compressing its representation of each class. As the basic level networks become trained as experts during phase two, this difference diminishes. This is shown in Figure 8.

Interestingly, this increased within-class variance for expert networks carried through to the unseen phase two stimuli (Figure 9). This means that the different subordinate classes for the new task are already differentiated to some extent in the representational space for the expert-level networks prior to the networks ever being exposed to phase two stimuli. This early differentiation makes the new task easier to learn.

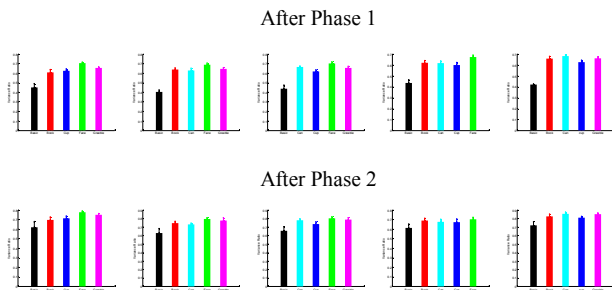After Phase 1

After Phase 2

Figure 8: Ratios of within-class variance to total variance averaged over all stimuli. From left to right these are graphs where the phase two stimuli are: cans, cups, books, Greebles, and faces. Each bar represents a particular network in that training condition. Basic networks (shown in black) consistently have the least within-class variance during phase one, but they become more like the expert networks after receiving expert training in phase two.
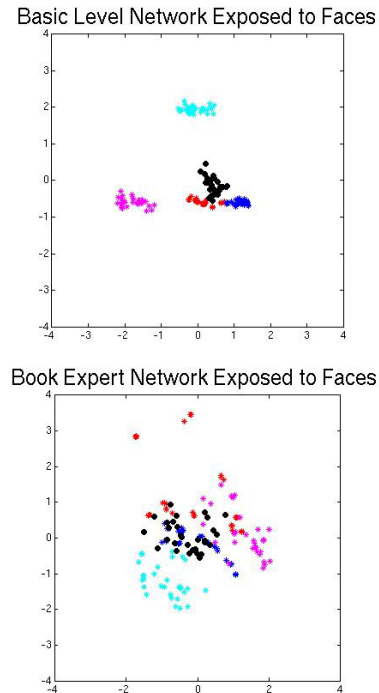

Figure 9: PCA of hidden unit activation for each class of networks after phase one training. Faces were the novel stimulus and are shown as black circles. The first two components are displayed.

Conclusions

Neural networks trained to perform subordinate level discrimination in one class of objects show an advantage when learning a new class of objects at the subordinate level when compared to networks trained at the basic level. This is because learning an expert-level task causes the network to spread out objects from the same class in its representational space, increasing the within-class variance of all classes, even unseen ones. The results were not specific to one object class and were robust to additional controls on the difficulty of task and number of discriminations being made.

This provides further evidence for our contention that the FFA is recruited for new expertise tasks because of the way it treats its inputs: it magnifies small differences between homogeneous objects. Furthermore, this paper also confirms our position that it is the nature, not the number, of the distinctions that matter when visual objects are processed. That is, in the expert networks, they are given many objects *of the same sort*, and are required to distinguish among them. This problem clearly requires magnifying within class variance, in contrast to the basic level categorization process, which requires magnifying between class variance while minimizing within class variance.

This paper therefore furthers our investigation into the issue of how the FFA, if one believes that it is a fine level discrimination area, becomes recruited for novel expertise tasks (the "visual expertise mystery", Joyce & Cottrell, 2004). Here we have shown that again, visual expertise is a general skill – being a "fish expertise area" is a good prerequisite for

becoming a "sword expertise area," as counterintuitive as that might seem.

Acknowledgements

References

De Renzi, E., Perani, D., Carlesimo, G.A., Silveri, M.C., and Fazio, F. (1994). Prosopagnosia can be associated with damage confined to the right hemisphere – An MRI and PET study and review of the literature. *Neuropsychologia*, **32**(8):893-902.

Eimer, M. (2000). The face-specific N170 component reflects late stages in the structural encoding of faces." *Cognitive Neuroscience Vol 2 No 10: 2319-2324.*

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., and Gore, J. C. (1999). Activation of the middle fusiform "face area" increases with expertise in recognizing novel objects. *Nature Neuroscience,* **2**(6):568-573.

Gauthier, I., Skudlarski, P., Gore, J.C., and Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, **3**(2):191-197.

Golby, AJ, Gabrieli, JDE, Chiao, JY and JL Eberhardt (2001) Differential responses in the fusiform region to same-race and other-race faces. *Nature Neuroscience,* **4**(8):845-850.

Grill-Spector, K., Knouf, N. and Kanwisher, N. (2004) The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience* **7**(5):555-562.

Joyce, C. A. and Cottrell, G. W. (2004). Solving the Visual Expertise Mystery. *Neural Computation and Psychology Workshop*. Howard Bowman and Christophe Labiouse (Eds.), World Scientific.

Kanwisher, N., McDermott, J., and Chun, M. M. (1997). The Fusiform Face Area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, 17:4302-4311.

Rhodes, G., Byatt, G., Michie, PT and Puce, A. (2004) Is the Fusiform Face Area Specialized for Faces, Individuation, or Expert Individuation? *Journal of Cognitive Neuroscience* **16**(2):189–203.

Sugimoto, M. and Cottrell, G. W. (2001). Visual expertise is a general skill. *Proceedings of the 23rd Annual Conference of the Cognitive Science Society* (pp. 994-999). Mahwah, NJ: Lawrence Erlbaum.

Tanaka, J. W. and Curran, T. (2003). A neural basis for expert object recognition. *Psychological Science* **1**(12):43-47.

Tarr, MJ, and Gauthier, I. (2000). FFA: a flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nature Neuroscience* **3**(8):764-769.

Tran, B., Joyce, C. A., and Cottrell, G. W. (2004). Visual expertise depends on how you slice the space. *Proceedings of the 26th Annual Conference of the Cognitive Science Conference.* Mahwah, NJ: Lawrence Erlbaum.

Yovel, G. and Kanwisher, N. (2004) Face perception: Domain specific, not process specific. *Neuron*, **44**:889–898.