Why is the Fusiform Face Area Recruited for Other Domains of Expertise?

Garrison Cottrell (gary@ucsd.edu)
Department of Computer Science and Engineering
University of California, San Diego, CA 92093

Abstract

What is the role of the Fusiform Face Area (FFA)? Is it specific to face processing, or is it a visual expertise area? The expertise hypothesis is appealing due to a number of studies showing that the FFA is activated by pictures of objects within the subject's domain of expertise (e.g., cars for car experts, birds for birders, etc.), and that activation of the FFA increases as new expertise is acquired in the lab. However, it is incumbent upon the proponents of the expertise hypothesis to explain how it is that an area that is initially specialized for faces becomes recruited for new classes of stimuli. We dub this the visual expertise mystery. This paper summarizes a decade of research on this topic. We show that a neurocomputational model trained to perform subordinate-level discrimination within a visually homogeneous class develops transformations that magnify differences between similar objects, in marked contrast to networks trained to simply categorize the objects. This magnification generalizes to novel classes, leading to faster learning of new discriminations. We suggest this is why the FFA is recruited for new expertise. The model predicts that individual FFA neurons will have highly variable responses to stimuli within expertise domains. **Keywords:** Neurocomputational models; Fusiform Face Area; Expertise hypothesis

Introduction

There has been a great deal of progress in understanding how complex objects, in particular, human faces, are processed by the cortex. At the same time, there is a great deal of controversy about the role of various cortical areas, especially the Fusiform Face Area (FFA) (Kanwisher et al., 1997; Kanwisher, 2000; Tarr and Gauthier, 2000). Is the FFA a module, specific to the domain of faces, or is it instead specific to the process of fine level discrimination? Several fMRI studies have shown high activation in the FFA only to face stimuli and not other objects (Kanwisher et al., 1997; Kanwisher, 2000). Gauthier et al. (1997) have challenged the notion of the face specificity of the FFA by pointing out that the earlier studies failed to equate the level of experience subjects had with non-face objets with the level of experience they had with faces.

Gauthier et al. (2000) showed that the FFA was activated when car and bird experts were shown pictures of the animals in their area of expertise. Furthermore, they illustrated that, if properly trained, individuals can develop expertise on novel, non-face objects (e.g., Greebles), and subsequently show increased FFA activation to them (Gauthier et al., 1999). Crucially, the same 2 or 3 voxels that are most active for faces also show the largest increase in activity over the course of expertise training on non-face stimuli, suggesting that the FFA is recruited as subjects learn to visually discriminate novel homogeneous stimuli, and is automatically engaged when the subject is an expert (Tarr and Gauthier, 2000). Hence the theory is that the FFA is a fine level discrimination area. However, this idea still does not answer the question of what mechanism would explain how an area that presumably starts life as a face processing region is recruited for these other types of stimuli. This is a job for modeling.

The Model

In short, we trained two networks (Figure 1), one to be a"basic level" categorizer, that simply mapped images into their categories, and one to be an "expert level" discriminator, that mapped images into their subcategory (Bob, Carol, Ted, or Alice, book1, book2, etc.) (Tong et al., 2008). The hidden layers of the two networks then represent the features learned by the Lateral Occipital Complex (LOC), and the FFA. We then trained the two networks on Greebles, and put them into a race to see who won.

Our working hypothesis was that cortical areas *compete* to solve tasks, and that the region of cortex that learned a task faster would be used for that task. In Figure 2, we plot the amount of epochs required to learn Greebles as a function of the number of epochs of training on the primary task. As can be seen from the Figure, when the initial domain of expertise was faces, the expert network always learned the Greeble task faster than the basic network (representing the task performed by LOC).

One advantage of modeling is that one can perform experiments that are impossible or unethical to perform on humans. One aspect of this in Figure **??** is that at every point in the figure, the network is "xeroxed", and trained on Greebles, resulting in the points shown in the graph. That is, we can do this as a within-subjects experiment, and show that the longer we train on the primary task, the faster the network learns the Greebles.

Another aspect of the modeling advantage is the non-Face expert curves in Figure 2. These curves are instances of training the network initially to be a cup, can, or book expert. In all three cases, the expert network learned the Greeble task faster than the basic network. This experiment is impossible to perform on humans, but the model results lead us to suggest, somewhat facetiously, that if our parents were cans, the Fusiform Can Area would be recruited to perform the Greeble task.

Our model thus suggests that there is nothing special about faces *per se* any primary expertise domain resulted in faster Greeble training. It is our contention that it is the *task* that is the important variable here, not the domain of expertise.

The next question is, why does this happen? Surely, cups and Greebles don't share features, do they? It turns out that in this case, they do. In Figure 3 we plot the internal representation of the penultimate hidden layer by plotting the 2nd and 3rd components of the PCA of the hidden unit acti-

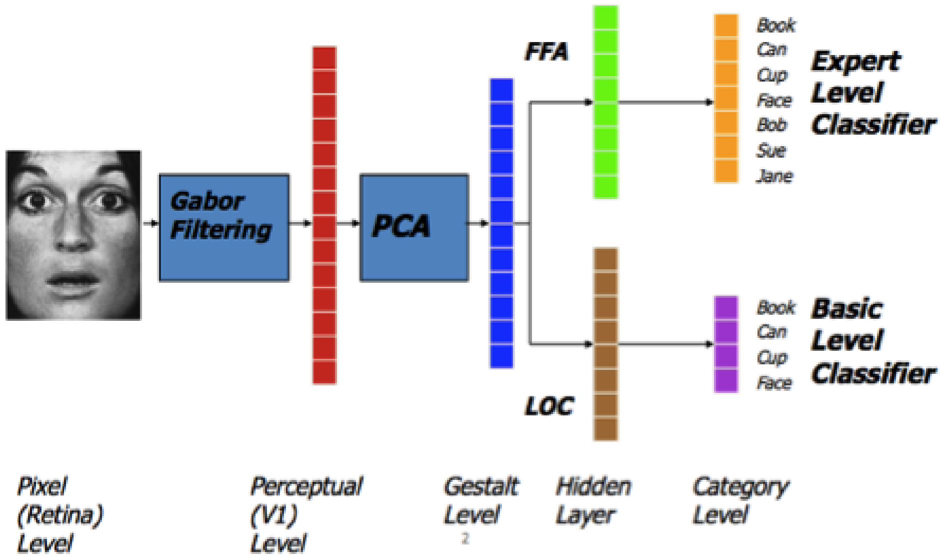

Figure 1: The Model. The first layer is convolutional, using Gabor filters, followed by PCA, followed by a trained hidden layer. The two branches reflect the putative tasks of the LOC and FFA.

vations (the first just expresses the growth in weights). The expert network (top row), displays the "spreading transform" that separates not only categories, but individuals in those categories, including the basic level categories it was trained on. The basic level network "clumps" all of the inputs into one or two regions of the space. The plot also shows the Greebles (red dots) projected into the space before training. As can be seen, the Greebles are already spread out in face space prior to training. The spreading transform generalizes to novel inputs. The Basic network, on the other hand, clumps all of the Greebles together, as it does the other categories.

Another way to think about this is that a basic level network has to take a set of similar-looking things, e.g., books, and minimize the variance between them, while maximizing the variance between the categories, similarly to Fisher's Linear Discriminant. On the other hand, the Expert networks must take a set of similar-looking things, in this case faces, and maximize the variance between them. This is what we refer to as the spreading transform. This allows the final layer to separate out individuals.

An interesting observation here is in Figure 2. The expert networks have actually learned their primary task (data not shown) early in training, in less than 100 epochs. At that point, they are already performing near-optimally on the expert task. The networks are therefore *overtraining* on the primary task. Yet, this overtraining leads to *faster* training on the Greeble task. This is the opposite of most experience in machine learning, where transfer and generalization is made more difficult from overlearning on the original task. As can be seen in Figure 3, the network at 1280 epochs has almost already solved the Greeble task, as they are already spread out in representational space.

A visualization of the receptive fields of two hidden units in the face network before and after Greeble training is shown in Figure 4. As can be seen in the Figure, the representations have hardly changed after training, supporting the interpretation of the Greeble representations in Figure 2, that the features learned by the network have almost already solved the Greeble task by spreading them out in representational space.

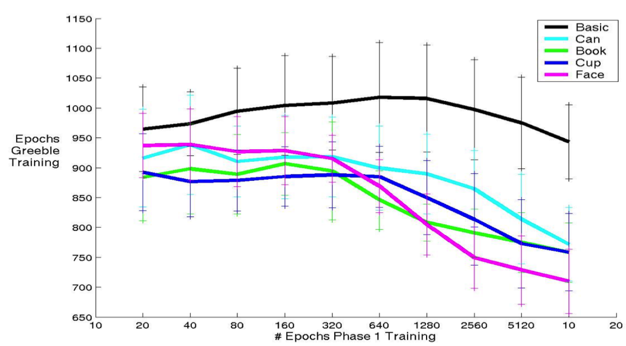

Figure 2: Amount of training time required to learn the Greeble task as a function of training time on the first domain of expertise.

We have replicated this experiment many times, using a basic-level task that was just as difficult as the expert task, and using identical inputs and number of outputs for the two networks, cutting the space in different ways to reflect basic and expert categorization (identify the letters (basic) or identify the fonts (expert)). We have also shown that this network can explain individual differences in expertise, and can account for the result that when experience with non-face categories is high, face performance correlates highly with object performance (Wang et al., 2016).

Conclusion

This work shows that a neurocomputational model can be used to explain what otherwise might be mysterious: Why would the face area get recruited for new tasks? The answer is that the task requires a transform of the data, the spreading transform, that generalizes to new categories. Again, the network representing the task carried out by the FFA must learn features that discriminate between similar-looking things (faces), and that representation is useful for separating elements from other categories as well.

The other broad conclusion we take from this work is that there is nothing special about faces, despite the many papers to the contrary. What is special is what we do with them - individuate them - and that task requires learning a representation that spreads the items out in representational space. This occurs no matter if the domain is cups, cans, books, or even letters vs. fonts.

References

Gauthier, I., Anderson, A.W., Tarr, M.J., Skudlarski, P., Gore, J.C., 1997. Levels of categorization in visual recognition studied with functional MRI. *Curr. Biol.* 7, 645651.

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, 568573.

Gauthier, I., Skudlarski, P., Gore, J.C., Anderson, A.W., 2000. Expertise for cars and birds recruits brain areas involved in face recognition. Nat. Neurosci. 3, 191197.

Kanwisher, N., 2000. Domain specificity in face perception. Nat. Neurosci. 3, 759762.

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, 43024311.

Tarr, M.J., Gauthier, I., 2000. FFA: a flexible fusiform area for subordinate-level visual processing automatized by expertise. Nat. Neurosci. 3, 764769.

Tong, M.H., Joyce, C.A., and Cottrell, G.W. (2008) Why is the fusiform face area recruited for novel categories of expertise? A neurocomputational investigation. Brain Research 1202:14-24.

Wang, Panqu, Gauthier, Isabel, and Cottrell, Garrison W. (2016) Are face and object recognition independent? A neurocomputational modeling exploration. J. Cogn. Neuro. 28, 558574.

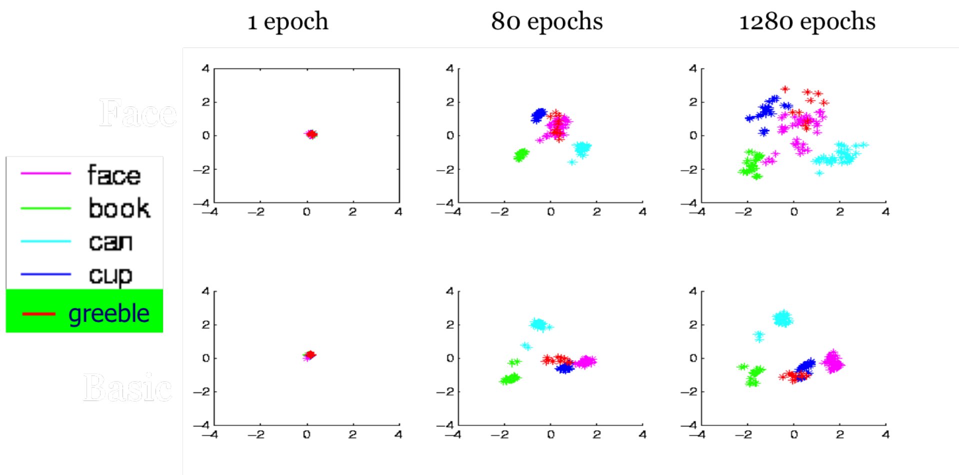

Figure 3: Plot of the representational space learned in the hidden layer of the Face expert network (top) and the Basic network (bottom). This is the 2nd and 3rd principal components of the hidden layer representations over learning.

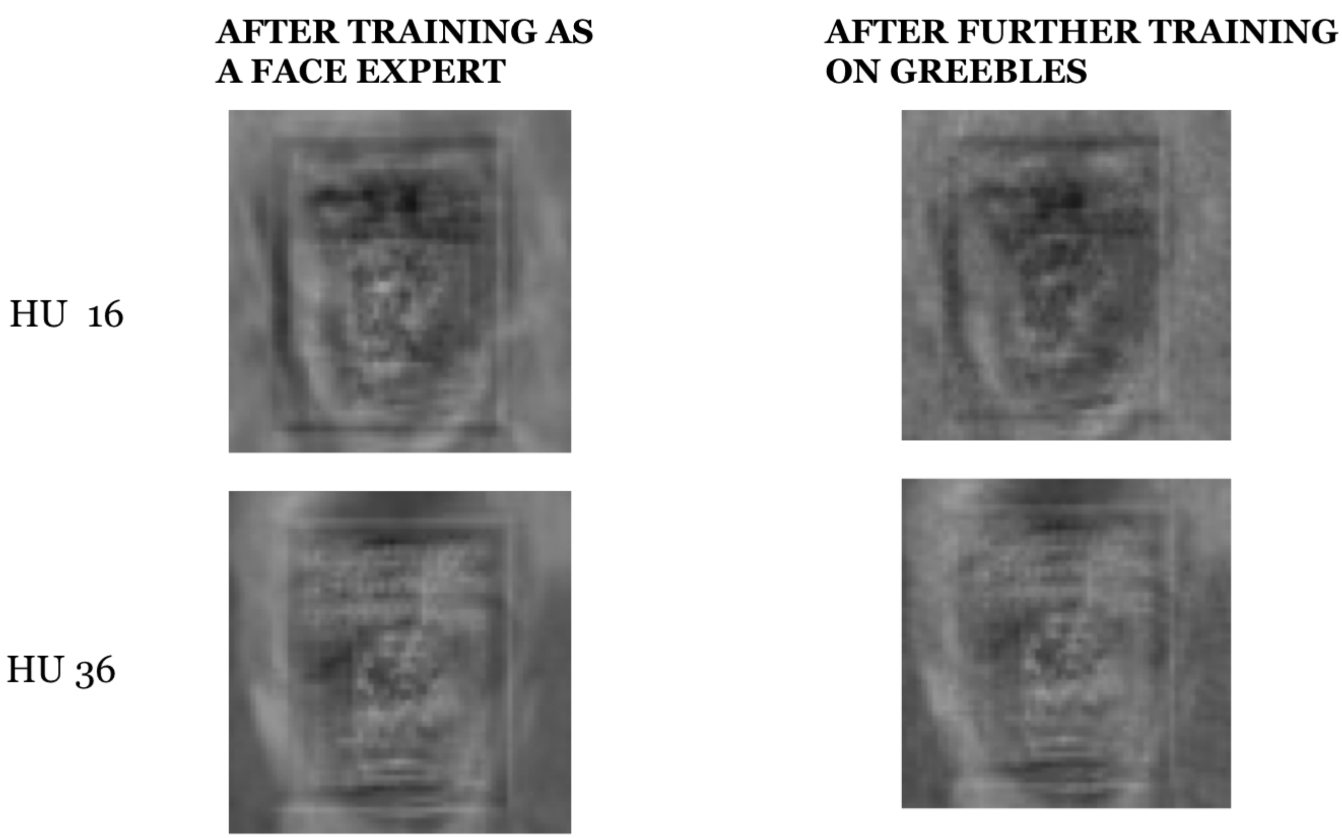

Figure 4: A visualization of two randomly-chosen hidden unit receptive fields in the face network before and after Greeble training.