# SOLVING THE VISUAL EXPERTISE MYSTERY

C. A. JOYCE AND G. W. COTTRELL

*University of California, San Diego*
*Computer Science and Engineering*
*9500 Gilman Dr.,*
*La Jolla, CA 92093-0114, USA*
*E-mail: {gary,cjoyce}@cs.ucsd.edu*

Through brain imaging studies and studies of brain-lesioned patients with face or object recognition deficits, the fusiform face area (FFA) has been identified as a face-specific processing area. Recent work, however, illustrates that the FFA is also responsive to a wide variety of non-face objects if levels of discrimination and expertise are controlled. The mystery is why an expertise area, whose initial domain of expertise is presumably faces, would be recruited for these other domains. Here we show that features tuned for fine-level discrimination within one visually homogeneous class have high-variance responses across that class. This variability generalizes to other homogenous classes, providing a foothold for learning.

## 1. Introduction

There has been a great deal of progress in understanding how complex objects, in particular, human faces, are processed by the cortex. However, there is also controversy about the roles of various cortical areas, especially the Fusiform Face Area (FFA).[1,2,3] Is the FFA a "module," specific to the domain of faces, or is it instead specific to the process of fine level discrimination? Damage to the FFA leads to prosopagnosia[4] (the inability to recognize faces), but it is unclear how face-specific this processing deficit is. Further, some researchers have shown, using fMRI, that when the level of expertise is controlled, the FFA is activated in car, bird, and Greeble (a class of fictional objects, see Figure 1, right column) experts when they view their respective categories of expertise.[5,6,7] This suggests that the FFA is a *fine level discrimination area*. The issue we address in this paper is why an area that presumably starts life as a face processing area (this being the first domain of expertise) is *recruited* for these other types of stimuli?

In addressing this question, the definition of "expertise" is critical. We adopt Gauthier and Tarr's operational definition of the term: someone is an expert if they are as fast to identify members of a category as individuals (subordinate level) as they are to verify their category membership (basic level). For example, a bird expert would be as fast/accurate at identifying a picture of a bird as an "Indigo Bunting" as at identifying it as a "bird." When training a subject in a novel category, the convergence in reaction times in these two tasks is called the "entry level shift."

This study replicates and expands on previous work[8] in which we have shown that neurocomputational models trained to make fine level discriminations learn individuation of Greebles faster than models that have not been trained to become experts in any domain. This suggests that, if there is a *competition* between cortical areas to solve tasks, as has been suggested previously,[9,10] the FFA would be primed to win the competition for a novel expertise task. Here, we show why this happens.

## 2. Experimental Methods

To investigate this issue, neural networks were trained on Greeble identification following various pretraining regimens.

The stimulus set consisted of 300 64x64 8-bit grayscale images of human faces, books, cans, cups, and Greebles (60 images per class, 5 images of 12 individuals, see Figure 1). The five images of each indivudual within each category were created by randomly moving the item 1 pixel in the vertical/horizontal plane, and rotating up to +/-3 degrees in the image plane.

Images were preprocessed by applying Gabor wavelet filters as a simple model of complex cell responses in visual cortex, extracting the magnitudes (which makes them nonlinear and somewhat translation invariant), normalizing via z-scoring, and reducing dimensionality to 40 via principal component analysis (PCA)[11]. Greeble images were not used to generate the principal components in order to model subjects' lack of experience with this category.

A standard feed-forward neural network architecture (40 input units, 60 standard logistic-sigmoid hidden units, variable numbers of linear output units) was used. Networks were trained using a learning rate of 0.01 and momentum of .5.

During pretraining, *all* networks (basic and expert) learned to perform basic level categorization on all 4 non-Greeble categories. *Expert* networks were additionally taught to perform subordinate level categorization of one of the four categories. Basic level networks had 4 output nodes correspond-

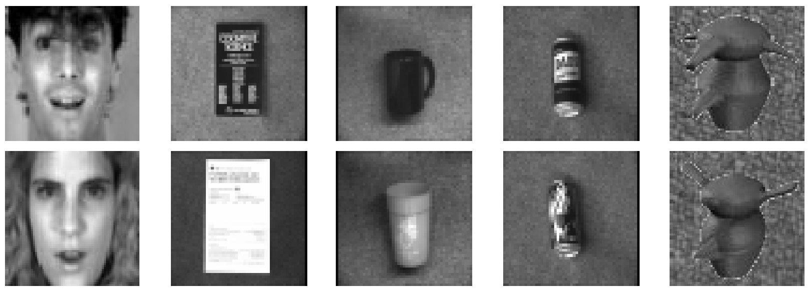

Figure 1. Example stimuli

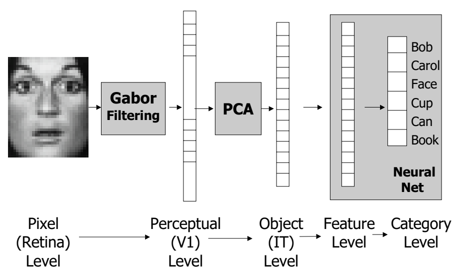


Figure 2. The expertise model. The feature level is where task-specific features are developed and variance is measured in Figure 7.

ing to book, can, cup, and face. Expert networks had 14 outputs: 4 for the basic categories, and 1 for each of the 10 individuals (e.g. can1, can2, ... can10, for a can expert). In phase two, the pretrained networks learned subordinate level Greeble categorization along with their original task. Eleven output nodes were added: 1 for the basic level Greeble categorization, and 1 for each Greeble individual. The network then performed a 15-way (basic network) or 25-way (expert network) classification task. All networks were trained on 30 images (3 images of 10 individuals) per class during pretraining and 30 more images of Greebles in phase 2. Thus any differences in representation are due to the task, not experience with exemplars. To test for generalization, 29 images were used (one new image of each of the expert category individuals (10 + 10), plus 3 images of novel basic level exemplars per category).

Ten networks, each with different random initial weights, were trained on each of the 5 pretraining tasks (basic, or face/can/cup/book expert) for 5120 epochs. Image sets were randomized. Intermediate weights of each network were stored every $5 * 2^n$ epochs, for n=1:10. Phase 2 training was performed at each of these points ("copying" the network at that point) to

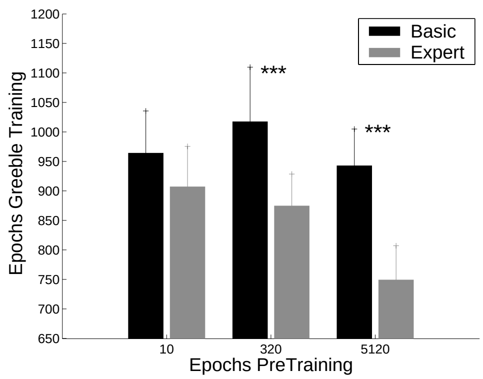


Figure 3. Number of epochs to learn the new task based on number of pretraining epochs. Error bars denote +/-1 standard deviation.

observe the time course of expertise effects. Training concluded when the RMSE of the Greebles fell below .05. Thus, there were a total of 550 phase 2 networks.

## 3. Results

All networks reached an RMSE of less than .0012 by the completion of 5120 pretraining epochs, with basic networks learning faster than expert networks.

Figure 3 shows the average number of epochs required for networks of each type to learn the subordinate Greeble task at three levels of pretraining epochs. The basic level networks took by far the longest to learn the Greeble task, obtaining no benefit from more pretraining cycles. All of the expert networks learned the Greeble task significantly faster if they were given more pretraining on their initial expert task, with faces benefitting the most from additional pretraining (data not shown).

### 3.1. *Entry Level Shift*

Training paradigms with human subjects use the reaction time entry level shift to determine a subject's expert status. Example data from a human Greeble expert is shown in Figure 4a. In networks, reaction time is modelled as the amount of uncertainty in the output of the network. This uncertainty is measured by taking 1 minus the logistic of the output activation on the node corresponding to the correct category or individual classification for each output pattern. Figure 4b shows the Greeble entry level shift for a network pretrained as a book expert. Note that response time to subordinate level classification of books is as fast as basic level classification prior to Greeble training.

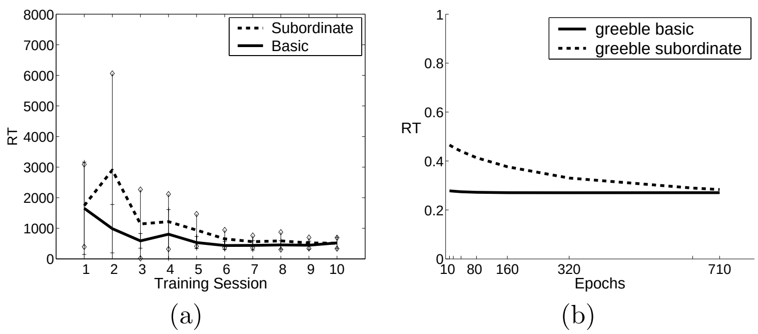


Figure 4. Entry level shift for the Greeble task. (a) Human data from one of our experts. (b) Network data.

### 3.2. *Network Plasticity*

In previous work[8], we hypothesized that the hidden units in the expert networks would tend to stay in the linear range, in order to better perform the fine level discrimination task. We suggested that this would lead to faster learning of the new task, since the higher slope of the hidden units would result in faster weight changes. The slope of the hidden units has been called a measure of plasticity in previous work.[12]

Plasticity to a stimulus category is measured as the average value of the slope of the activation function (here the logistic sigmoid) across all hidden units for all input patterns from that category. Unexpectedly, results indicated that lower plasticity networks learned the new task faster. Figure 5 shows the plasticity of the pretrained networks in response to the stimuli used during pretraining (left), and to the new set of untrained Greeble patterns (right). For all patterns (pretrained and novel), non-expert networks retained their plasticity better across pretraining epochs than experts. Furthermore, plasticity to the (untrained) Greebles *decreased* over training on the expert task. This paradox may be resolved in part if the plasticity measure is viewed as a measure of *mismatch* between the stimuli and the weight vectors – the closer the weight vectors line up with the stimuli (either in

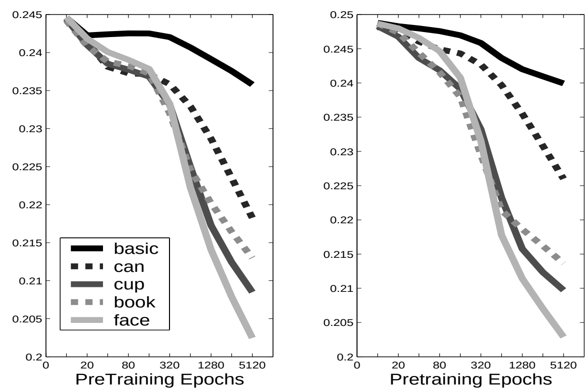


Figure 5. Average plasticity of the hidden units over training to learned categories (left), and novel Greebles (right).

the same or opposite direction), the more the hidden units will be activated or inactivated. Thus, here the weight vectors are simply becoming more aligned with the stimuli, and, perhaps surprisingly, also more aligned with the Greeble images. This is not the whole story, however, as we will see in the next section.

### 3.3. *Hidden Unit Activation*

Since expert network representations become less plastic with training, how does the network actually discriminate one individual from another within and across categories?

The activation of the hidden units in response to each category of stimulus provides some explanation. Figure 6 shows the activation levels of 3 representative hidden units from a basic level (a,b) and a face expert (c,d) network in response to individual training patterns both prior to (column 1) and after (column 2) Greeble training.

Prior to Greeble training (column 1), the hidden units in subordinate level networks (Figure 6c) show more variability of response across input patterns than do basic level networks (Figure 6a). After Greeble training, both basic and expert level networks show more variability in hidden unit activation across input patterns (Figure 6b,d).

These results suggest that correct discrimination requires a representation that is distributed across multiple hidden units which modulate in different ways in response to different input patterns from the same class.

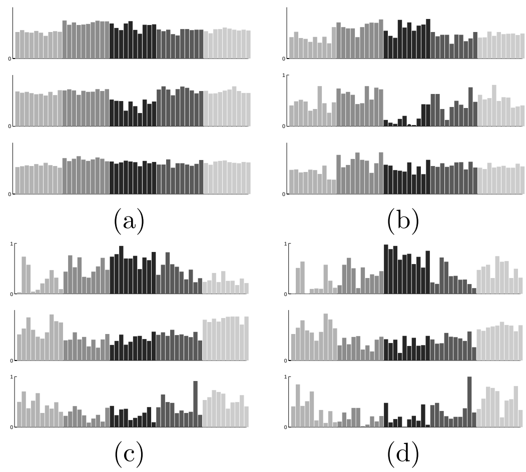


Figure 6. Single unit recordings from networks for face, book, can, cup, and greeble patterns, respectively. a) basic network, pre-Greeble training; b) basic network, post-Greeble training; c) face expert, pre-Greeble training; d) face expert, post-Greeble training.

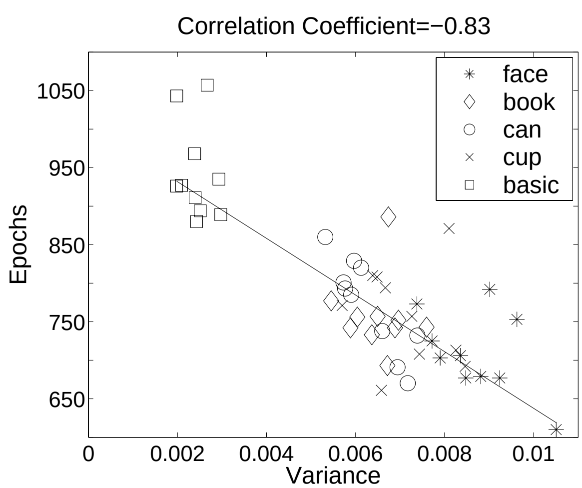


Figure 7. A regression of Greeble pre-training variance versus training time.

## 3.4. *Relationship of Variability to Learning*

There appears to be a provocative relationship between learning and hidden unit variability: networks that have learned a subordinate level task and exhibit hidden unit variability, also learn a secondary subordinate level task faster than basic level networks which exhibit little hidden unit variability. This suggests two things: 1) variability should increase with experience, particularly when making a subordinate level discrimination, and 2) the amount of variance a network exhibits in response to a category prior to training on that category should be predictive of the speed with which that category is learned. The first hypothesis is addressed by examining how variability changed over the course of pretraining: 1) variability increases for all categories in all networks as the number of training epochs increases; 2) increases in variability are much larger for expert networks than basic networks, and are largest for the category being learned at the subordinate level; 3) expert networks show more variability to all categories than basic networks, even to categories being learned at the basic level; 4) even variability to Greebles, which the network has never been trained on in any manner, increases with pretraining epochs, although not as much as the categories being trained (at both subordinate and basic levels). These results support the conclusion that pretraining causes networks, particularly those making a subordinate level discrimination, to learn features which generalize well to new categories.

Figure 7 illustrates the second hypothesis: that amount of variability to Greebles, prior to training on them (x-axis), should be predictive of how fast the network can learn the Greeble task (y-axis). There is a strong negative linear correlation between these two variables for expert networks such that those exhibiting the lowest variance also take the longest to learn the Greeble task ($r^2 = -.53, p < .001$). For basic networks, there is no significant correlation between variance and learning time ($r^2 = -.21, p = .557$). Those expert networks exhibiting the highest variance and lowest Greeble learning time, are the networks that initially learned faces, the task that was the most difficult to learn in pretraining. This suggests a relationship between the difficulty of the pretraining task, and the ease with which subsequent subordinate discriminations can be learned.

## 4. Conclusions

The results of these simulations are indicative of a system in which expertise results from the flexible use of fine-tuned feature representations. Further, the types of features learned through subordinate level discrimination of visually different categories seem to generalize well to new categories. Finally, learning difficult perceptual discriminations enables faster learning of new discriminations. These results suggest that the FFA fine–tunes its sensitivity to small differences in homogeneous stimuli when given a novel, fine–level discrimination task.

It might be considered counter-intuitive that an expert network with low plasticity at the hidden layer should yield more variable responses across hidden units. The measures themselves explain how this can occur. Maximum plasticity occurs when there is a large mismatch between inputs and weights (i.e., they are orthogonal). As the network becomes more expert, the inputs and weights become more similar/matched (i.e., it loses plasticity). Basically, the weights become more tuned to the specific input vectors and, for expert networks, more responsive to the small differences between them. Thus, the resulting hidden unit activations become more variable because they correspond more closely to the fine-level differences between the input patterns (for the expert networks).

A critical question, then, is what exactly are the features the FFA uses? More research is required to address this question, but clearly these features must be broad enough to encompass vastly visually different stimuli. In further work we will investigate the possibility that these features result from combinations of lower level visual sensitivities of the cells that feed into FFA - for example, cells which are sensitive to low spatial frequencies. Thus the features coded in this area could be reflections of early, lower-level visual processing biases.

## Acknowledgments

This work was supported by the McDonnell Foundation (Perceptual Expertise Network, 15573-S6) and NIMH MH57075 grant to GWC.

## References

1. N. Kanwisher, J. McDermott, and M. M. Chun. The fusiform face area: A module in human extrastriate cortex specialized for face perception. *J Neurosci*, 17:4302–4311, 1997.
2. N. Kanwisher. Domain specificity in face perception. *Nat Neurosci*, 3 (8):759–763, August 2000.
3. M. J. Tarr and I. Gauthier. Ffa: A flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nat Neurosci*, 3(8):764–769, August 2000.
4. E. De Renzi, D. Perani, G.A. Carlesimo, M.C. Silveri, and F. Fazio. Prosopagnosia can be associated with damage confined to the right

hemisphere — An MRI and PET study and a review of the literature. *Psychologia*, 32(8):893–902, 1994.

5. I. Gauthier and M.J. Tarr. Becoming a “greeble” expert: Exploring mechanisms for face recognition. *Vision Res*, 37(12):1673–1682, 1997.
6. I. Gauthier, M. J. Tarr, A. W. Anderson, P. Skudlarski, and J. C. Gore. Activation of the middle fusiform “face area” increases with expertise in recognizing novel objects. *Nat Neurosci*, 2(6):568–573, June 1999.
7. I. Gauthier, P. Skudlarski, J. C. Gore, and A. W. Anderson. Expertise for cars and birds recruits brain areas involved in face recognition. *Nat Neurosci*, 3(2):191–197, 2000.
8. Maki Sugimoto and Garrison W. Cottrell. Visual expertise is a general skill. In *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, Mahwah, New Jersey, 2001. Lawrence Erlbaum Associates.
9. G.M. Edelman. *Neural Darwinism: The theory of neuronal group selection*. Basic Books, Inc., New York, NY, 1987.
10. Robert A. Jacobs, Michael I. Jordan, Steven J. Nowlan, and Geoffrey E. Hinton. Adaptive mixtures of local experts. *Neural Comput*, 3:79–87, 1991.
11. M. N. Dailey and G. W. Cottrell. Organization of face and object recognition in modular neural network models. *Neural Netw*, 12(7–8): 1053–1074, 1999.
12. A.W. Ellis and M.A. Lambon Ralph. Age of acquisition effects in adult lexical processing reflect loss of plasticity in maturing systems: Insights from connectionist networks. *J Exp Psychol Learn Mem Cogn*, 26(5): 1103–1123, 2000.