Published in final edited form as:
Brain Res. 2008 April 2; 1202: 14–24. doi:10.1016/j.brainres.2007.06.079.

Why is the Fusiform Face Area recruited for novel categories of expertise?: A neurocomputational investigation

Carrie A. Joyce[1], **Matthew H. Tong**[2], and **Garrison W. Cottrell**[2]

[1]*ID Analytics, Inc., San Diego, CA*

[2]*Computer Science and Engineering, University of California, San Diego, 9500 Gilman Dr., La Jolla, California 92093-0114*

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." One suggested answer to this mystery is that the FFA is used simply *because* it is a fine discrimination area, but this account has historically lacked a mechanism describing exactly how the FFA would be recruited for novel domains of expertise. In this study, 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.

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 & 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 showed high activation in the FFA only to face stimuli and not other objects (Kanwisher et al., 1997; Kanwisher, 2000). Furthermore, studies involving patients with associative prosopagnosia, the inability to identify individual faces (Farah et al., 1995), and visual object agnosia, the inability to recognize non-face objects (Moscovitch et al., 1997), seem to indicate a clear double dissociation between face and object processing. Prosopagnosic patients had lesions encompassing either right hemisphere or bilateral FFA, while object agnosic patients' lesions did not (De Renzi et al., 1994).

Corresponding Author: Garrison W. Cottrell, CSE Dept. 0404, UCSD, La Jolla, CA 92093-0404, Tel: 858-534-6640, Fax: 858-534-7029, gary@ucsd.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Gauthier and colleagues 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 objects with the level of experience they had with faces (Gauthier et al., 1997; Gauthier et al., 1999a). Gauthier and colleagues showed that the FFA was activated when car and bird experts were shown pictures of the animals in their area of expertise (Gauthier et al., 2000). Further, 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 & Tarr, 1997; Gauthier et al., 1999b). 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 & Gauthier, 2000). Hence the theory is that the FFA is a *fine level discrimination area* (this is still controversial – see (Grill-Spector et al. 2004; Rhodes et al. 2004) for competing evidence). However, the idea that the FFA is a fine level discrimination area 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.

Before addressing this question it is important to define the notion of an “expert.” We use Gauthier’s operational definition of the term: experts are as fast to verify that a picture of an object is a particular individual (subordinate level) as they are to verify their category membership (basic level). For example, a bird expert would be as fast and as accurate at verifying that a picture of a bird is an “Indigo Bunting” as at identifying it as a “bird.” On the other hand, a novice will show the fastest reaction time at the basic level, and is slower at both subordinate and superordinate level (Tanaka & Taylor, 1991). The basic level was first identified by Rosch as the level at which objects tend to share the same shape and function, and tends correspond to the first word we use to describe an object (a picture of a chair is labeled “chair” rather than “furniture” of “office chair”). When training a subject in a novel category, the downward shift in reaction times in these two tasks is taken as evidence of expertise.

Previously, we have demonstrated that developmentally appropriate conditions (low spatial frequency input and learning subordinate/individual level classification) are sufficient for our neurocomputational model to specialize for faces (Dailey & Cottrell, 1999) Here, we investigate what properties the FFA might possess that would result in its recruitment for non-face, subordinate level discrimination tasks.

We compare the properties of two kinds of cortical models: “expert networks” trained to make subordinate level categorizations (“Is this Bob, Carol, Ted or Alice?”, top path of Figure 1), and “basic networks” trained to make category level classifications (“Is this a face, cup, can, or book?”, bottom path of Figure 1) on the stimuli shown in Figure 2. We then show that expert networks learn individuation of *novel* categories faster than basic networks. Thus, if cortical networks compete to solve tasks, this learning advantage suggests that the FFA, as a fine level discrimination network, would be recruited to perform novel fine-level discrimination tasks over a network that has no previous experience with such processing. An advantage of computational modeling is that the “first expertise” domain of the networks need not be faces: our results do not depend on the order in which domains are learned, suggesting there is nothing special about faces.

Similar to previous work (Dailey & Cottrell, 1999; Dailey et al. 2002; Palmeri & Gauthier, 2004; Reisenhuber & Poggio, 1999), the model uses layers of processing from low level features to high level categories: 1) a Gabor filter layer models cortical responses of early visual cortex (Daugman, 1985); 2) a principal components layer (learnable via Hebbian methods

(Sanger, 1989)) models object representations as correlations between Gabor filter responses; 3) a hidden layer models a task-specific feature representation (representing subordinate or basic level processing, depending on the task), trained by back-propagation (Rumelhart, et al., 1986); and 4) a categorization layer that controls the level of discrimination between the stimuli, either subordinate or basic level. Minor variations of this model have accounted for a variety of behavioral face processing data (Cottrell et al., 2002; Dailey & Cottrell, 1999; Dailey et al., 2002; By analyzing the hidden layers of the two types of networks, we found that expert networks spread out the representations of similar objects in order to distinguish them. Conversely, basic networks represent invariances among category members, and hence compress them into a small region of representational space. The transformation performed by expert networks (i.e. magnifying differences) generalizes to new categories, leading to faster learning. The simulations predict that FFA neurons will have highly variable responses across members of an expert category.

Results and Discussion

Network Training

Training of the networks occurred in two phases. During the pre-training phase, two kinds of networks were trained. Basic-level networks were trained to differentiate a set of stimuli (cups, cans, books, and faces) (see Figure 2) at the category level. Expert-level networks also had to perform this basic-level categorization, but were also required to differentiate one of these classes at the subordinate level. Hence there were four kinds of expert networks – “cup experts,” “can experts,” “book experts,” and “face experts.” During the second phase of training, a novel stimulus type, “Greebles[1]” was introduced and both basic and expert networks were trained to identify Greebles and to recognize individual Greebles. Training was also continued on the prior tasks. This reflects the fact that exposure to the new area of expertise is added to the daily routine of interacting with the world. This is also true in human experiments in creating experts in the lab, where training typically occurs for an hour a day over one to two weeks (Gauthier & Tarr, 1997). Not performing this interleaving would be equivalent to taking a human subject “out of the world,” and allowing them only visual exposure to the objects of expertise, a situation that seems unrealistic at best. If our model was not trained in such an interleaved fashion, face expertise would decay over the course of training. This may seem like an unrealistic prediction of the model. However, it is worth noting that it has recently been reported that for one class of experts, this prediction would seem to hold up. Kung et al. (2007) examined bird experts’ FFA activity with respect to their degree of expertise. Expertise was measured by d’ on a same/different species test with bird images. As might be expected from previous studies of visual expertise, they found that with increasing levels of bird expertise, the FFA was more activated by bird images. However, they also found that with increasing levels of bird expertise, the FFA was *less activated by faces.* This finding suggests that the FFA is plastic in its responsiveness depending on the kind of expertise that is most prominent in a particular subject. Our model would exhibit similar differences if it were trained more frequently on Greebles than on its original domain of expertise.

Basic networks learned their pre-training task the fastest and maintained the lowest error (RMSE, see Methods) until between 1280 and 5120 training epochs (one pass through the training set), when the various expert networks caught up (can, cup, book, and face experts in that order) (see Figure 3). Conversely, the basic-level networks took by far the longest to learn the novel task (Figure 4), obtaining no significant benefit from additional pre-training cycles. A linear trend analysis shows that all of the expert networks (but not the basic networks) learned

[1] Greebles are a fictitious category of objects created by Isabel Gauthier for her Ph.D. thesis. They were constructed to have some properties similar to human categories – they have family resemblances, they have a “gender,” and are symmetric. They have gender labels, family labels, and individual names. Two examples are shown in the last column of Figure 2.

the novel task faster if they were given more pre-training on their initial expert task, with faces benefiting the most from additional pre-training (an F-test for non-zero slope with n=100 for each test (10 networks at 10 time steps) yields p = 0.2962 for basic networks and p < 0.0001 for expert networks). Thus, for the networks learning a harder pre-training task (expert-level classification), more pre-training lead to faster learning on the secondary, expert-level task. In this study, we alternately used faces, cups, cans and books as the primary expertise task, and Greebles as the novel (secondary) expertise task. However, we have replicated these results consistently with a variety of primary and secondary expertise tasks. For example, a network with prior expertise with books learns expertise with faces faster than a network with only basic level experience with books.

The networks learn both the primary and secondary tasks, but are they experts? We model human subjects' reaction time as the uncertainty of the maximally activated output (see Methods). Figure 5b shows the entry-level shift for Greebles in a network that was trained to be a face expert during pre-training (note that subordinate face model reaction times are already as low as basic level face reaction time). This curve is quite similar to the entry-level shift shown by a human subject trained in our lab to individuate Greebles (Figure 5a). Therefore, according to the criterion used for human subjects, the networks have attained expert status.

Internal Representations

We hypothesized that the learning advantage for expert networks was due to the larger amount of information that must be carried by the internal representations formed during training. We can visualize the representations by performing a Principal Components Analysis (PCA) of the hidden unit activations over the data and then project the data onto a two-dimensional subspace. We perform this over the training time of the network in order to see how the representations develop. This is shown in Figure 6, in which the second and third principal components of the hidden unit activation to each input pattern are plotted against one another (the first PC just captures the magnitudes of the weights growing over time). Note the larger separation for the expert network on both subordinate and basic level categories as pre-training progresses. On the other hand, while the basic network separates the classes, it also compresses each class into a small blob in the space. Furthermore, we can project the (so far untrained) Greeble patterns into the same space, and the plot shows that these are also more separated by the expert network – the spread of representations of homogeneous classes generalizes to a novel category. This is the fundamental reason for speeded learning of Greebles: it is easier to "pick off" each Greeble if they are in different locations in feature space to begin with.

A neurophysiological correlate to the above results is that the spread of representations will correspond to increased variability of single-unit responses across a homogeneous category in an expert network, and hence, in the FFA. Referring to the PCA visualization in Figure 6, the two dimensions in that graph correspond to two "virtual unit" responses to the stimuli. Since the points are more spread out in expert networks, this means that these units have higher variability of response across a class. We can visualize this in the single unit recordings shown in Figure 7, which shows the actual activation levels of several hidden units in basic and expert networks to individual stimuli. As is clear from the figure, there is greater variability across a single class of stimuli in an expert network versus a basic network, and the greatest variability is for the class being discriminated. An analysis of variance with 5 levels of category (Expert networks shown stimuli from their domain of expertise (called Expert), Expert networks shown stimuli outside their domain of expertise (but trained at the basic-level, called Expert-basic), Expert networks shown the untrained Greeble stimuli (Expert-Greeble), Basic networks shown stimuli from the trained basic set (Basic), and Basic networks shown the untrained Greeble stimuli (Basic-Greeble)) and 11 levels of training epoch (0, 10, 20, 40, 80, 160, 320, 640, 1280, 2560, 5120) was performed to determine effects. For this ANOVA, the mean variance over the relevant stimuli was the observation; thus for the expert networks 40 observations of each mean variance were available (4 types of networks, 10 runs of each), while for basic networks 10 observations were used; this yields a total of 1540 points in the ANOVA. There was a main effect of category [F(4,1485) = 992.91, $p < 0.0001$] such that the Expert category showed the most variance followed in order by the Expert-Basic, Expert-Greeble, Basic, and Basic-Greeble categories in order. There was also a main effect of epoch [F(10,1485) = 1216.73, $p < 0.0001$], with the least variance exhibited initially with variance significantly increasing across training epochs. There was also a significant interaction of category with epoch [F(40,1485) = 43.51, $p < 0.0001$].

To examine how this develops over time, we plot the average variance of response of the hidden units across a class over training in Figure 8. As expected based on the PCA visualization, the greatest variability is to the category learned at the subordinate level, and this variability of response extends to the non-expert categories as well. That is, in expert networks, there is more variability of response to every stimulus category than in networks that simply do basic-level categorization. Furthermore, this variability in response extended to the completely novel Greeble category. Note that Figure 8 shows the response to untrained Greeble stimuli. When Greebles are then trained, the variance of response to them then increases above the levels shown in Figure 8 (data not shown).

A post-hoc right-tailed two-sample t-test was performed on the final epoch to determine the significance of the final ordering; all orderings were significant ($p < .00001$, with $n = 40$ or $n = 10$ measures of mean variances for expert and basic networks respectively) except for expert networks shown basic and Greeble stimuli ($p = 0.8840$). All networks were initialized with weights drawn from the same distribution and show only the small differences in variance of response due to differences in stimuli classes, so this result is due to the effects of training with the pre-training stimulus prior to (and during) training the novel stimulus. Finally, (data not shown), becoming a Greeble expert increased variability in all networks. This caused the originally basic-level networks to resemble the other expert networks in that now their variability was higher to all categories. Based on these results, the model predicts that neurons involved in fine level discrimination, as is the hypothesis concerning the FFA, will show greater variability across stimuli that the subject possesses expertise in. This variability of response will be greater than in areas outside the FFA.

It is possible that these results are simply due to a scaling difference between the two types of networks, if the weights in a basic level network are simply smaller overall than in an expert network. To control for this possible artifact, we computed the variance of the object classes *relative* to the variance between classes of the internal representation. We find that the relative variance of the representation of discriminated classes in expert networks is significantly higher than in networks where these same stimuli are simply being categorized. As the PCA visualization suggests, we find that if we average together the variability of all classes categorized at the basic level, and compute the ratio of this to the between class variance (the variance of the means), there is still a significant difference (using a right-tailed paired t-test with $n = 10$, $p < 0.0001$ for all pairings of expert to basic). This demonstrates that the expert networks are unnecessarily spreading out the classes they do not need to discriminate. Finally, we find that objects that are novel to the network (Greebles) also have a higher spread in expert networks (again using a right-tailed paired t-test with $n = 10$, $p < 0.0001$ for all pairings of expert to basic).

In the simulations discussed above, networks that learned a subordinate level task, and therefore exhibited a high degree of hidden unit variability, learned a secondary subordinate level task faster than basic level networks that exhibit little hidden unit variability. This suggests that the amount variance a network exhibits in response to a category prior to training on that category should be predictive of how fast that network will learn to discriminate that category.

To test this hypothesis, we performed a regression on the amount of variability of feature responses to Greebles prior to Greeble training, versus the number of epochs it takes the network to learn the Greeble task. There is a strong negative linear correlation between these two variables ($r = -0.6317$, $p < 0.0001$), such that those networks exhibiting the lowest variance also take the longest to learn the Greeble task (Figure 9).

At this point the careful reader will have noticed that the main effect of being an expert network is a higher variability of response to stimuli from the categories of expertise, and then wondered how this could possibly account for the increased BOLD signal seen in fMRI experiments in the FFA for expertise stimuli. One might assume we should be measuring increases in mean firing rates, rather than variance. However, we suggest that an increased variance in firing rates for neurons over a class of stimuli should correlate with higher mean firing rate, by the following argument: Biological neurons find encodings of the world that tend to maximize sparsity in order to minimize their firing rates while maintaining high levels of discriminability. In the interests of simplicity, our model contains no such bias for sparsity, and our artificial neurons utilize their full range of firing rates with equal probability. Furthermore, since both positive and negative weights are allowed, the actual activation of a neuron says nothing about its sensitivity to a particular type of stimuli; sensitivity is instead displayed by changes in activation, which is related to variance in sparse encodings. In the case of biological neurons with base rates near zero, an increased firing rate will result in a net increase in the variance. In particular, if the probability density function of a neuron’s firing rate *r* follows a steep exponential distribution (one possible model of sparse coding), as in:

$$f\,(r,\lambda) = \lambda e^{-\lambda r}\ (r \geq 0)$$

then as the variance increases (given by λ^{-2}), so does the mean (λ^{-1}). While our model’s activations do not follow this distribution, we argue that a more realistic model that did use sparse coding would also show the same increase in variance to stimuli of expertise. Indeed, it seems obvious now that within-class variability in such a model directly corresponds to having a different pattern of activation for different stimuli, an essential component of the ability to discriminate. Hence, while the goal of our model was to describe the recruitment of the FFA to other domains of expertise due to the FFA’s relative fitness for such tasks compared with other areas, our model also does show the kind of sensitivity to domains of expertise that correlate with findings from the fMRI literature. A more literal correlation of mean firing rates would require several additional assumptions in our model that go well beyond the scope of this paper.

A second concern may arise due to the fact that, at least in the principal components plots of Figure 6, it would appear that the same dimensions that are sensitive to faces are also sensitive to other stimuli. Is there really such an overlap in representation in the FFA? Recent work by Grill-Spector et al. (2006;2007) suggests that there is. After localizing the FFA using standard fMRI, high-resolution fMRI was used to measure the BOLD response from 1 mm^2 voxels in the FFA. These voxels were assessed for their selectivity to faces, cars, animals, and abstract sculptures. In the original paper, it appeared that voxels were highly selective for each of these categories, but that face voxels were simply more numerous. However, in response to critiques of the analysis technique (Baker et al., 2007;Simmons et al., 2007), a more accurate assessment of sensitivity was applied. The result was that, while most voxels were most selective for faces, they were also sensitive to other categories as well. While this does not prove anything about individual neuron tuning, it does suggest that the FFA is not just responsive to faces; it is a much more hetergogeneous area than was originally thought. This analysis is also consistent with the idea that individual neurons may respond to faces and other categories of stimuli, and hence is consistent with our model's suggestion that minor re-tuning of the neural responses in the FFA is sufficient to account for the responses to new areas of expertise.

Finally, there is still a great deal of controversy whether there is a "face area" at all. Work by Haxby and colleagues (Haxby et al. 2001; Hanson et al., 2004) has shown that it is possible to accurately classify the stimulus class being observed by a subject using a standard machine learning pattern classifier applied to several different regions of cortex, that may or may not include the FFA. However, these experiments do not address the foremost role that we hypothesize for the FFA – fine level discrimination of homogeneous categories. It is not surprising that one can determine *at a basic level* what is being observed from multiple brain areas. Indeed, we would predict that from our model. What has not been shown that one can determine *who* is being observed from widely distributed brain activations. Thus this data is not inconsistent with the putative role of the FFA as a fine level discrimination area.

Conclusions

Several effects were observed in these simulations: 1) networks can become experts, by the behavioral definition of the entry-level shift in reaction times; 2) expert networks learn the Greeble expertise task faster than basic-level categorizers; 3) this can be attributed to the spread of representations in expert networks: Greebles are more separated by these features than by the basic-network features; and 4) this feature variability to the Greeble category prior to training on it is predictive of the ease with which it will be learned. The results imply some specific hypotheses about phenomena that might be observable in human and/or primate subjects. First, though, let us be clear about what these results do *not* imply. We interpret these results to be relevant to competing cortical areas, not to different subjects learning different tasks. Thus, our results should not be interpreted to mean that subjects that have just learned a hard discrimination task should be more successful at learning a new discrimination task than subjects who have learned a simple discrimination task. Indeed, it is usually the case that it takes longer to learn novel categories of visual stimuli like these than it would if the network were starting from initial random weights. The point is rather that fine level discrimination areas are better at learning new fine level discriminations than simple object categorization areas.

What the results do suggest is that if the FFA is performing fine-level discrimination, then that task requires it to develop representations of the stimuli that separate them in representational space – the neural responses are highly differentiated. That is, similar objects have the differences between them magnified by the expert networks. On the other hand, networks that simply categorize objects map those objects into small, localized regions in representation space (this is in the space of neural firing patterns, and should not be confused with spatially localized representations). The magnifying transform of the expert networks generalizes to a novel category, and this generalization leads to faster learning; hence, the recruitment of the FFA for Greeble expertise. We have suggested that the hidden layer of the expert networks of our model corresponds roughly to the FFA based on the equivalency of tasks and have shown that the nature of the task is sufficient to cause the recruitment of the FFA based on a shared need for fine-level discrimination; however, the actual brain is of far greater complexity than our model, and some of the changes observed in the hidden layer may turn out to be distributed among several brain areas.

An advantage of using simulations is that we were also able to show that this expertise effect is not limited to face experts. To put it in a somewhat fanciful way, the results suggest that if our parents were cans, then the Fusiform Can Area would be recruited for Greeble expertise. Furthermore, other simulations show that this learning advantage is not limited to novel Greeble expertise, nor is it dependent on the difference in the number of distinctions the two networks are making (Tong et al., 2005; Tran et al., 2004).

These simulations also make a prediction concerning the physiological responses of FFA neurons. They predict that at the physiological level (perhaps using intracranial electrode arrays), cells in the FFA should show more variability across stimuli within a category than cells in other high-order visual object areas, and that this variability would be particularly high for categories for which the viewer possesses expertise (e.g. human and/or monkey faces). This is a falsifiable prediction of the model, and hence we look forward to our model being put to the test.

Methods

Training and testing

Neural networks were trained on a subordinate level classification task following various pre-training regimens. The image preprocessing steps, network configurations, and simulation procedures are described below.

The stimulus set consisted of 300 64×64 8-bit grayscale images of human faces, books, cans, cups, and Greebles (60 images per class, 5 images of 12 individuals, see Figure 2). The five images of each Greeble were created by randomly moving the Greeble 1 pixel in the vertical/horizontal plane, and rotating up to +/-3 degrees in the image plane. Pictures of objects were taken under constant lighting and camera position, varying object position slightly over different images. Pictures of faces were frontal images of people making different facial expressions while camera angle and lighting remained constant (Cottrell & Metcalfe, 1991).

The images were preprocessed by applying Gabor wavelet filters of five scales and eight orientations as a simple model of complex cell responses in visual cortex, extracting the magnitudes, and reducing dimensionality to 40 via principal component analysis (PCA). We have found that the particular number of principal components used does not make any significant differences in our results for ranges from 30-50. 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 hidden units) was used (see Figure 1). The hidden layer units used the standard logistic sigmoid function while the outputs were linear. Networks were trained using backpropagation of error with a learning rate of 0.01 and a momentum of 0.5.

During pre-training all networks were trained to perform basic level categorization on all 4 non-Greeble categories. The expert networks were additionally taught to perform subordinate level categorization of one of the four categories. Non-expert networks (basic level task only) had 4 output nodes corresponding to book, can, cup, and face. Expert networks (subordinate level task) 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 2, the pre-trained 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 learned a 15-way (basic network) or 25-way (expert network) classification task. All networks were trained on the same 30 images (3 images of 10 individuals) per class during pre-training. 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 new basic level exemplars per category). All networks generalized well.

Ten networks, each with different random initial weights, were trained on each of the 5 pre-training tasks (basic, face expert, can expert, cup expert, 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 observe the time course of expertise effects. Training concluded when the RMSE of the Greebles fell below .05. Thus, there were a total of 50 phase 2 networks on which to perform the analyses.

Analysis

The linear trend analysis on the time to learn the novel Greeble identification task as a function of phase one training time was performed using an F-test on a least-squares linear regression to test for non-zero slopes. For each of the five networks there were 10 points at each of the 10 sampled epochs, yielding n=100. The time scale used was logarithmic. Although the data was non-linear, this nevertheless quantified the trend of the networks as they were exposed to additional training.

Reaction times of the networks were modeled as the uncertainty of the appropriate output. I.e., for the Greeble basic versus Greeble subordinate comparison in Figure 4b, we used RT = 1 – *activation*, where activation refers to the Greeble output unit for the basic RT, and activation refers to the output corresponding to the ith Greeble for the subordinate RT. Both of these are averaged over all 10 Greebles for one network chosen at random for the graph in Figure 2.

The principal components analysis of the hidden layer was performed on a network by recording the hidden unit activations for every training pattern at every point during which weights were saved (the initialization and the 10 stages of phase one training). The 60×60 covariance matrix of this data was formed, and the eigenvectors computed. A randomly chosen set of examples from each class at each time point was then projected onto the second and third eigenvector and plotted. A representative set of Greeble stimuli were also presented to the network (without training them), and their hidden unit vectors were projected into the subspace.

The variability plots were formed by computing the variance of each of the 60 hidden unit activations over the appropriate class of stimuli at different training epochs. Five levels of category were of interest: Expert networks shown stimuli from their domain of expertise, Expert networks shown stimuli outside their domain of expertise (but trained at the basic-level), Expert networks shown the untrained Greeble stimuli, Basic networks shown stimuli from the trained basic set, and Basic networks shown the untrained Greeble stimuli. The variance of these was tracked over eleven time samples (the variance of the randomly initialized networks and the ten stages of training). The variance over the sixty hidden units was then averaged for each of the 10 networks in a given category and epoch. As there were four categories of experts, there were 40 samples for each epoch for the expert networks, while there were only 10 for the basic networks, yielding a total of 1540 samples of average variance. To compensate for uneven cell sizes, an ANOVA using type 3 sum of squares was performed to measure the effects of these 5 categories and 11 epochs. We also computed the ratio of the average variability within a class to the variability between classes, to measure the spread of representations in the two types of networks, performing a two sample t-test on the variance ratio after phase 1 training was complete (n = 40 for expert networks, n = 10 for basic).

Acknowledgements

We would like to thank the reviewers for their helpful comments on an earlier version of this paper. This work was supported by a grant from the McDonnell Foundation to the Perceptual Expertise Network (15573-S6), NSF grant DGE-0333451 to GWC, NSF grant SBE-0542013 to GWC, and NIMH Grant R01 MH57075 to GWC.

References

Baker CI, Hutchinson TL, Kanwisher N. Nat Neurosci 2007;10:3–4. [PubMed: 17189940]

Buhmann, J.; Lades, M.; von der Malsburg, C. Proceedings of the International Joint Conference on Neural Networks (IJCNN). San Diego: 1990. Size and distortion invariant object recognition by hierarchical graph matching; p. IIp. 411-416.

Cottrell, GW.; Branson, KM.; Calder, AJ. Proceedings of the 24th Annual Conference of the Cognitive Science Society. Lawrence Erlbaum; Mahwah: 2002. Do expression and identity need separate representations?.

Cottrell, GW.; Metcalfe, J. Empath: face, gender and emotion recognition using holons. In: Lippman, Richard P.; Moody, John; Touretzky, David S., editors. Advances in Neural Information Processing Systems. 3. 1991. p. 564-571.

Dailey MN, Cottrell GW. The organization of face and object recognition in modular neural network models. Neural Networks 1999;12:1053–1074. [PubMed: 12662645]

Dailey MN, Cottrell GW, Padgett C, Adolphs R. EMPATH: A neural network that categorizes facial expressions. Journal of Cognitive Neuroscience 2002;14(8):1158–1173. [PubMed: 12495523]

Daugman JG. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. J Optical Soc Amer A 1985;2:1160–1169.

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

Farah MJ, Levinson KL, Klein KL. Face perception and within-category discrimination in prosopagnosia. Neuropsychologia 1995;33(6):661–674. [PubMed: 7675159]

Gauthier I, Anderson AW, Tarr MJ, Skudlarski P, Gore JC. Levels of categorization in visual recognition studied with functional MRI. Current Biology 1997;7:645–651. [PubMed: 9285718]

Gauthier I, Behrmann M, Tarr MJ. Can face recognition really be dissociated from recognition? Journal of Cognitive Neuroscience 1999a;11:349–370. [PubMed: 10471845]

Gauthier I, Skudlarski P, Gore JC, Anderson AW. Expertise for cars and birds recruits brain areas involved in face recognition. Nature Neuroscience 2000;3(2):191–197.

Gauthier I, Tarr MJ. Becoming a ”greeble” expert: Exploring mechanisms for face recognition. Vision Res 1997;37:1673–1682. [PubMed: 9231232]

Gauthier I, Tarr MJ, Anderson AW, Skudlarski P, Gore JC. Activation of the middle fusiform “face area” increases with expertise in recognizing novel objects. Nat Neurosci 1999b;2:568–573. [PubMed: 10448223]

Grill-Spector K, Knouf N, Kanwisher N. The fusiform face area subserves face perception, not generic within category identification. Nat Neurosci 2004;7:555–562. [PubMed: 15077112]

Hanson SJ, Matsuka T, Haxby JV. Combinatorial codes in ventral temporal lobe for object recognition: Haxby (2001) revisited: is there a face area? Neuroimage 2004;23(1):156–66. [PubMed: 15325362]

Haxby, James V.; Ida Gobbini, M.; Furey, Maura L.; Ishai, Alumit; Schouten, Jennifer L.; Pietrini, Pietro. Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. Science September 28;2001 293(5539):2425–2430. [PubMed: 11577229]

Joyce, CA.; Cottrell, GW. Solving the visual expertise mystery. In: Bowman, Howard; Labiouse, Christophe, editors. Connectionist Models of Cognition and Perception II: Proceedings of the Eighth Neural Computation and Psychology Workshop. World Scientific; 2004.

Kanwisher N, McDermott J, Chun MM. The fusiform face area: A module in human extrastriate cortex specialized for face perception. J Neurosci 1997;17:4302–4311. [PubMed: 9151747]

Kanwisher N. Domain specificity in face perception. Nat Neurosci 2000;3:759–762. [PubMed: 10903567]

Kung, Chun-Chia; Ellis, Colin; Tarr, Michael J. Dynamic reorganization of Fusiform gyrus: Long-term bird expertise reduces face selectivity. Poster presented at the 2007 Meeting of Cognitive Neuroscience Society; May 2007; New York. 2007.

Moscovitch M, Winocur G, Behrmann M. What is special about face recognition? Nineteen experiments on a person with vusual object agnosia and dyslexia but normal face recognition. Journal of Cognitive Neuroscience 1997;9(5):555–604.

Palmeri T, Gauthier I. Visual object understanding. Nature Reviews Neuroscience 2004;3:291–303.

Reisenhuber M, Poggio T. Hierarchical models of object processing in cortex. Nat Neurosci 1999;2:1019–1026. [PubMed: 10526343]

Rhodes G, Byatt G, Michie PT, Puce A. Is the fusiform face area specialized for faces, individuation, or expert individuation? J Cogn Neuro 2004;16:189–203.

Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature 1986;323:533–536.

Sanger TD. Optimal unsupervised learning in a single-layer linear feedforward neural network. Neural Networks 1989;2:459–453.

Simmons WK, Bellgowan PSF, Martin A. Measuring selectivity in fMRI data. Nat Neurosci 2007;10:4–5. [PubMed: 17189941]

Sugimoto M, Cottrell GW. Visual Expertise is a General Skill. Proceedings of the 23rd Annual Cognitive Science Conference 2001:994–999.

Tanaka JW, Taylor M. Object categories and expertise: is the basic level in the eye of the beholder? Cognitive Psychology 1991;23:457–482.

Tarr MJ, Gauthier I. FFA: A flexible fusiform area for subordinate-level visual processing automatized by expertise. Nat Neurosci 2000;3:764–769. [PubMed: 10903568]

Tong, MH.; Joyce, CA.; Cottrell, GW. Proceedings of the 27thAnnual Conference of the Cognitive Science Society. Lawrence Erlbaum; Mahwah: 2005. Are Greebles special? Or, why the Fusiform Fish Area (if we had one) would be recruited for sword expertise.

Tran, B.; Joyce, CA.; Cottrell, GW. Proceedings of the 26thAnnual Conference of the Cognitive Science Conference. Mahwah, NJ: Lawrence Erlbaum; 2004. Visual expertise depends on how you slice the space.

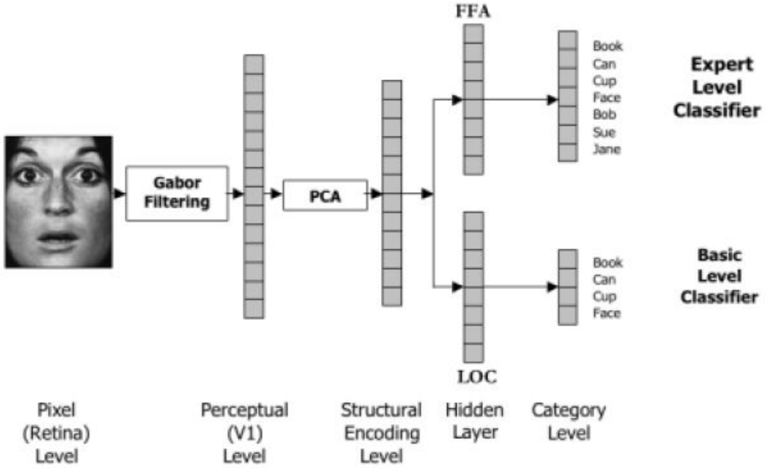

Figure 1.

Network Architecture. Input images are 64×64 grayscale images. The first layer of processing consists of Gabor filters (wavelets) at 8 different orientations (0, π/8, π/4, 3π/8, π/2, 5π/8, 3π/4 and 7π/8) and 5 different scales (see ref 9 for details). We keep the magnitudes of these filters (i.e., 40 numbers) from an 8×8 grid of 64 points, resulting in a 2560-dimensional representation of the image, which we term the perceptual level. The filter magnitudes are z-scored (shifted and scaled so they have 0 mean and unit standard deviation) on an individual basis across the data set before applying PCA. The top 40 components, again z-scored, were then used as input to a one hidden layer network. The hidden layer models the representations used for basic level categorization or fine-level discrimination, depending on the task. For basic networks, classification at the output nodes was at the basic level (i.e., four outputs, one per category) for all stimuli during pre-training and at the subordinate level (10 additional outputs) for Greebles following pre-training. For expert networks, one category (cars, cups, books, faces) was learned at the subordinate level and all other at the basic level during pre-training. Following pre-training Greebles were learned at the subordinate level.

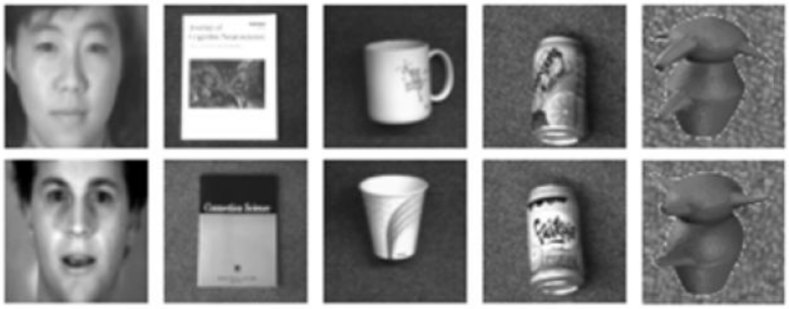

Figure 2.

Example Network Stimuli[9] 64×64, 8-bit, grayscale photos of books, cans, cups, faces, and Greebles used in the network simulations. Greebles are a created class of objects often used in studying expertise due to their novelty to subjects[5]. Three different images of each individual were used in training. Faces of the same person varied in expression, while images of other individual objects varied slightly in placement of the object in the image.

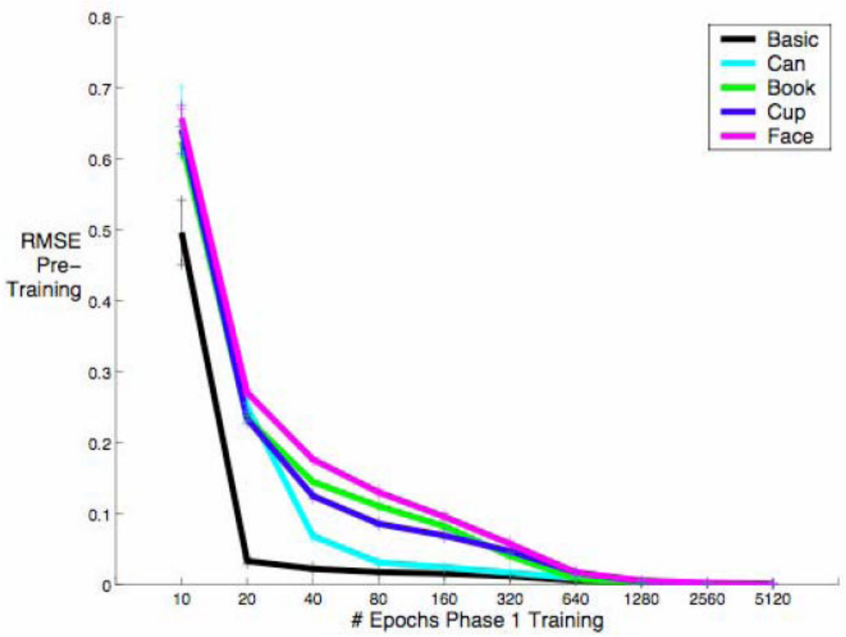

Figure 3.

Root Mean Squared Error (RMSE) on the training set over training time for the primary task. The basic level categorization task is the easiest.

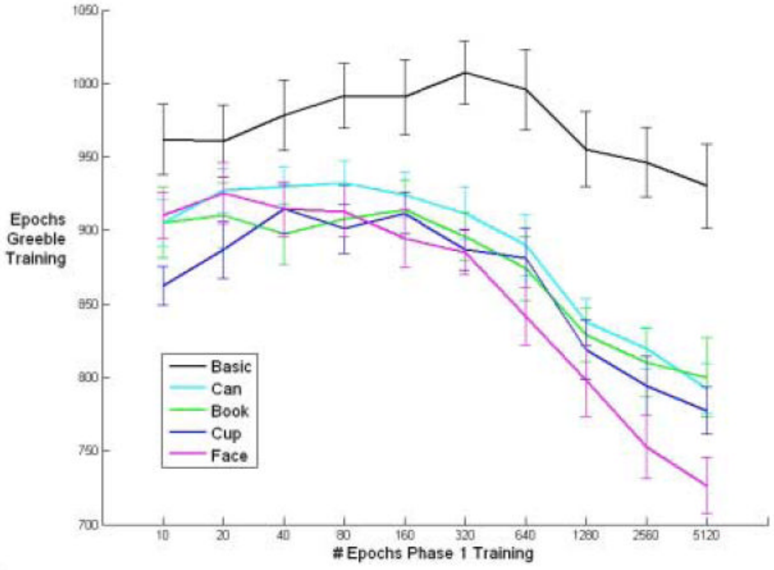

Figure 4.

Amount of time to learn Greebles as a function of number of epochs of pretraining on the first task. Training concluded when the RMSE of the Greebles fell below .05. Networks at the basic level always took longer to learn Greebles than all other networks and did not benefit significantly from increased experience with the basic level task. All expert level networks benefited from more pre-training, especially faces. Error bars denote +/-1 standard error.

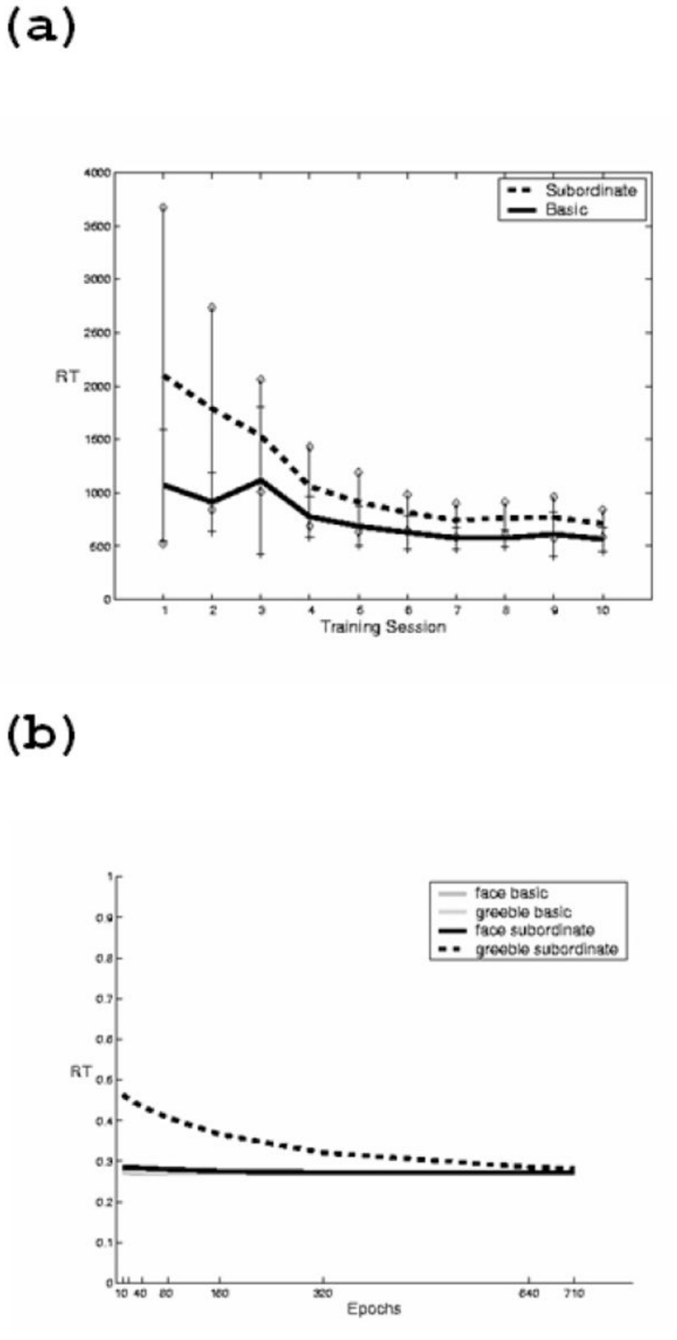

Figure 5.

Entry level shift. (a) Typical reduction in reaction time for basic vs. subordinate level judgments with training in a human subject learning to discriminate Greebles. (b) Reduction in reaction time in a neural network over training for subordinate versus basic level categorization. Reaction time is measured as the uncertainty of the maximum output (1-max$_{output}$).

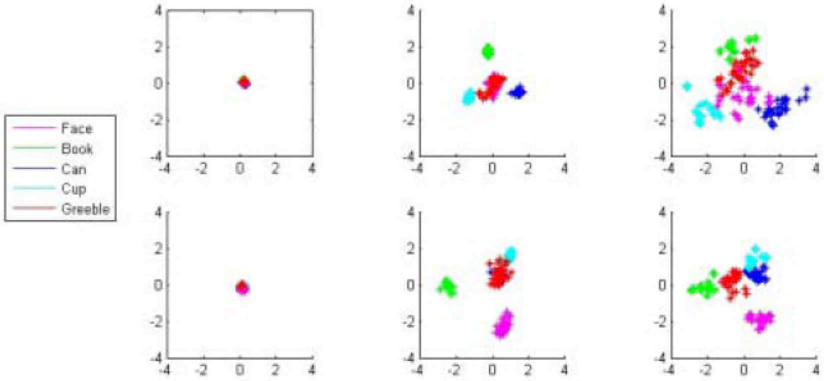

Figure 6.

Visualization of the hidden unit representation. The figure shows the second and third principal components (the first PC simply describes a growth in activation magnitude) of the hidden unit activation to images from the training set of two types of networks, a face expert (top row) and a basic-level network (bottom row) over training time. Samples are taken at 0 epochs (column 1), 80 epochs (column 2), and 5120 epochs (column 3) of training on the first task. Colors correspond to different object categories. Both networks separate the categories over training, but the face expert (top) also spreads out the representations within each class, with the largest spread for the category learned at the subordinate level (faces). This difference in representation corresponds to a difference in variability of response of the hidden units between the expert networks and the basic networks: The farther apart each point is, the larger the difference in unit response. To demonstrate the spread of the unseen, novel stimuli (shown in red), Greebles were presented to the networks and their hidden unit activations were projected onto the principal components.

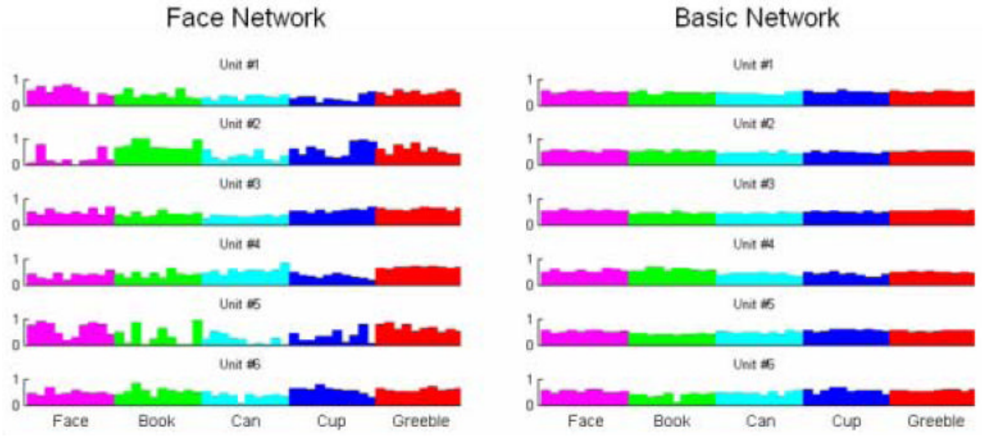

Figure 7.

Single unit recordings of randomly chosen units from the hidden layer of an expert (face) network (left) and a basic network (right), showing the higher variability of the expert network feature responses. Each histogram shows the response of one unit to 10 stimuli from five different categories.

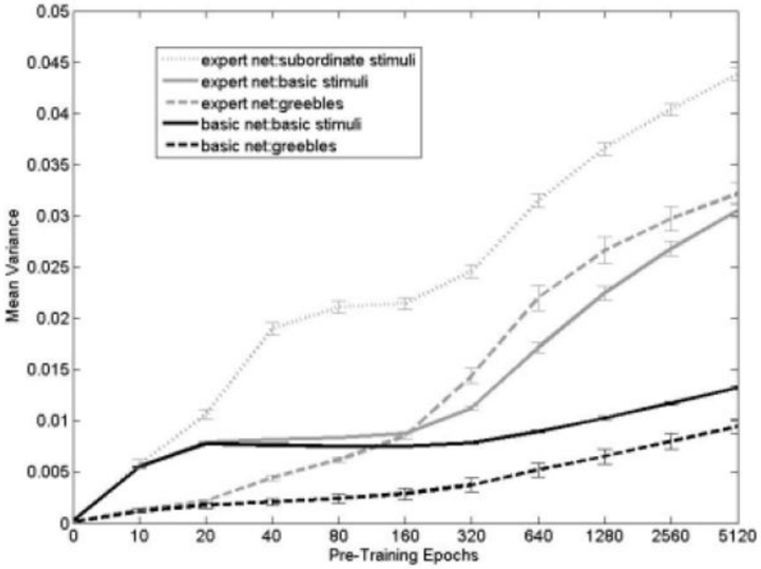

Figure 8.

Mean variance of hidden unit activations over training. While variance in all networks increased with training, the increase was largest for expert networks, and for categories learned at the subordinate level. This variability transferred to the unlearned Greeble category. Error bars denote +/-1 standard error.

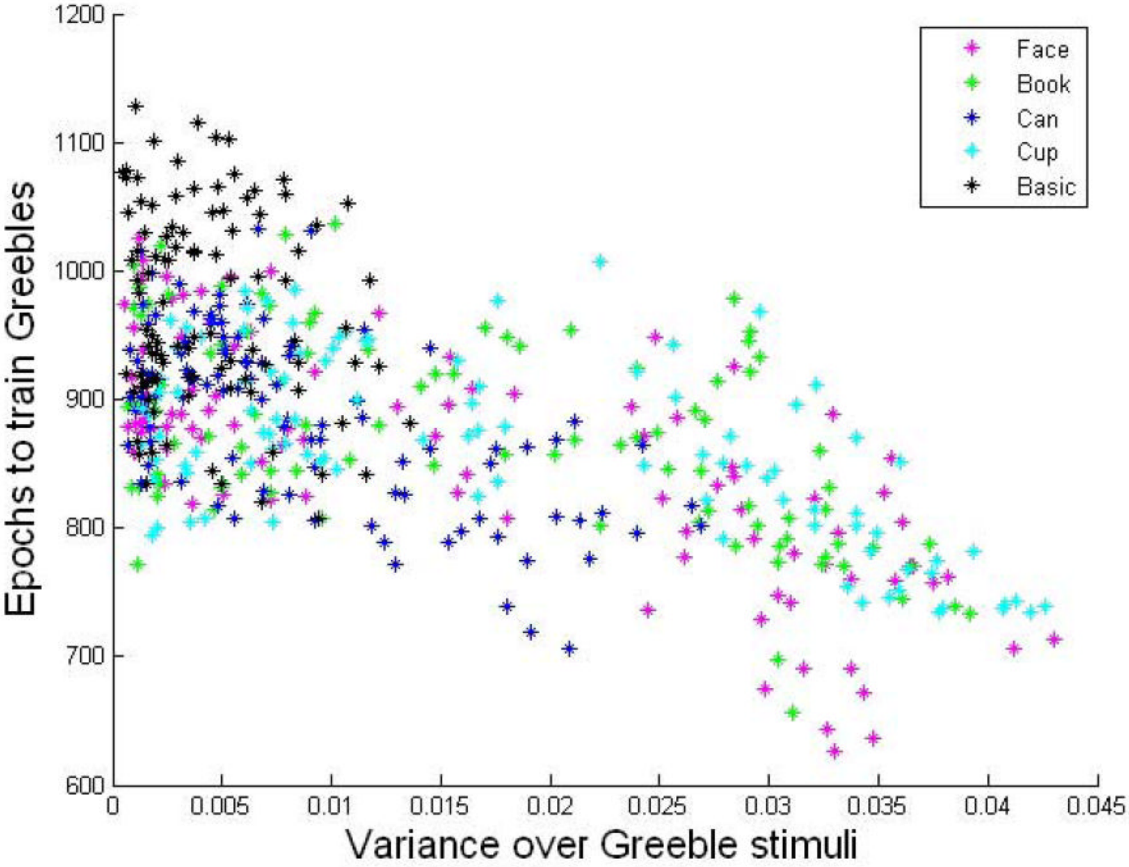

Figure 9.

Time to learn Greebles over Greeble pre-training activation variance. As the variance of the hidden layer activations over the Greeble stimuli increases, the training required to learn Greebles decreases. This correlation is strong (r= -0.6317, p < 0.0001). This variance is taken before the networks are trained with Greebles and represents the initial spread of Greebles in representational space.