Owls and Wading Birds: Generalization Gradients in Expertise

Nam H. Nguyen Garrison W. Cottrell
{nhnguyen, gary}@ucsd.edu
UCSD Department of Computer Science and Engineering, 9500 Gilman Drive
La Jolla, CA 92093-0114 USA

Abstract

Recently, Tanaka et al. (in press) have shown that subjects trained at the subordinate level (henceforth, “experts”) exhibit an advantage over subjects trained at the basic level on the same stimuli in performing discrimination tasks within their domain of expertise, showing that “mere exposure” to a category is not enough to induce discrimination behavior consistent with expertise. In addition, experts generalize their discrimination performance in a graded fashion to novel exemplars from known classes, as well as novel exemplars from novel, but related classes. We applied our two-component neurocomputational model of perceptual expertise to this domain (Sugimoto & Cottrell, 2001; Joyce & Cottrell, 2004). To our surprise, we found that we could not match the data with our original model. Ironically, we needed to add a new component that models “mere exposure” in order to account for the discrimination performance on basic level categories.

Introduction

What is required for perceptual expertise? Is it simple exposure to lots of examples of a class, or is more required? Tanaka and colleagues have shown recently that frequent exposure to exemplars of a domain alone is not enough to reach expert levels of discrimination. Instead, they found, subordinate level labeling of stimuli is required, or at least, is sufficient.

Tanaka et al. (in press) trained subjects on images of owls and wading birds. Half the subjects learned to label owls at the species level (e.g., “Great Grey Owl”), and the wading birds at the basic level (all were labeled “wading birds”). Half did the opposite. For simplicity, we will describe the results from the point of view of the subjects who were trained to discriminate owls at the species level. Subjects were tested before training on their ability to discriminate species in a same/different task. The discrimination test was repeated after training. The results were first, that the subordinate level training produced greater post-training gains in species discrimination relative to the basic level training. The advantage in discrimination also transferred to novel exemplars of trained owl species by a smaller amount, and finally, a small, but significant advantage for novel exemplars of novel species of owl was obtained. The trained images of wading birds only showed small gains over their pre-training baseline, and there was no advantage for untrained exemplars or novel species of wading birds. Transfer of discrimination thus only occurred in the category learned at the subordinate level, and was graded by similarity to previously learned items.

In our previous studies, we have developed a neurocomputational model that can explain many face and emotion discrimination and perceptual expertise effects. Our neural network model has two modules that learn basic and subordinate level classification tasks simultaneously. In this paper, we apply our model to Tanaka et al.’s domain. We find it necessary to add an auto-encoder module (an unsupervised network that extracts compact representations of its environment) to our network to account for subjects’ discrimination performance on basic-level stimuli, since it helps to spread out the representation of basic stimuli as well as subordinate stimuli. The auto-encoder module can be considered as a model for mere exposure. As Harnad (1987) pointed out, there is a need for at least three capacities in order to categorize: discrimination, identification, and an invertible description. In our model, the first capacity falls out of the second two. It is the level of naming or identification that gives rise to differing levels of discrimination, but being able to describe the object well enough to reproduce it (as in the hidden layer of an autoencoder) is necessary to account for discrimination of basic level (but frequently encountered) objects.

Tanaka’s Experimental Design and Results

The stimuli consisted of digitized photographs of owls and wading birds (Figure 1). All participants had the same exposure to stimuli. Participants first completed a pre-assessment “same/different” discrimination task to get the baseline for later comparison. In this task, participants were shown two bird images subsequently in time, and asked to respond either “same” or “different”. For "same" trials, the birds were two different images of the same species (e.g., two different images of "screech owls"). For "different" trial, the birds were images depicting two species from the same family (e.g., "screech owl" and "burrowing owl").

Figure 1: Tanaka's stimuli.

During training, participants were divided into two groups, one group learned to classify items as a wading bird (basic level), and as one of ten species of owl (subordinate level), and the other group did the reverse. The participants performed various tasks such as a naming task, category verification, and object classification. For the keyboard naming task, participants were shown a bird image, and asked to identify the stimulus at either the subordinate level or the basic level by pressing the corresponding key (e.g., “k” for eastern screech owl, “w” for wading bird). For the category verification task, participants were presented first with a subordinate level word label (e.g., “eastern screech owl”) or basic level label (e.g., “wading bird”), and then a bird image. Their task was to identify whether the label and the image were a match. For the object classification task, participants were first shown a word level which was similar to the category verification task, and then two images side by side. Their task was to select the image corresponding to the label. Following training, participants were given a same/different sequential matching task which was the same task performed in the pre-training phase. The matching task was partitioned into three conditions: 1) old instances of old species (items seen during training) (OLD/OLD), 2) new instances of old species (new exemplars of species seen during training) (NEW/OLD), and 3) new instances of new species (exemplars of species not seen during training) (NEW/NEW). These three conditions are pictured in Fig. 1.

They assessed d’ measures for each of these conditions for both the basic and subordinate level training which then was used to compare to baseline performance. The results are shown in Figure 2.

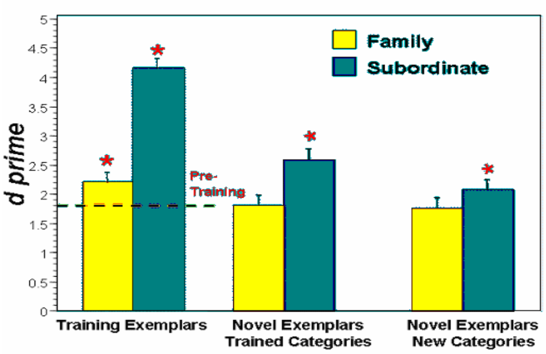


Figure 2: Post-training discrimination.

The Model

Our multi-module neural network was trained to perform two different tasks: a basic classification and expert (subordinate) classification. We began with our standard 2-module neural network then add an auto-encoder module. In this section, we will describe the input database, the image pre-processing procedure, network configurations and the simulation procedures.

Input Stimuli

To simulate Tanaka’s experiment we performed all of our simulations using symmetric and asymmetric Greebles instead of owls and wading birds. Our network model is not designed to handle variations in pose, scale and lighting so we used simpler stimuli since these variations aren’t essential to the story.

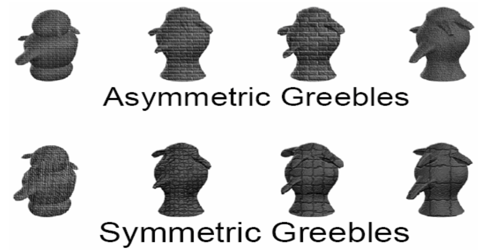


Figure 3: Greeble input stimuli.

The images were 8-bit grayscale images consisting of 5 basic classes: symmetric, asymmetric Greebles, cups, cans, and faces. Within the Greeble classes different exemplars of the same species are represented by different surface textures applied to the same Greeble. We also shifted the Greebles around by 2-3 image pixels to produce more within-species variations.

Image Preprocessing

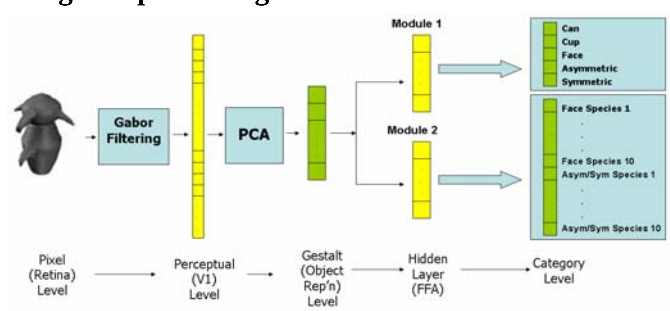


Figure 4: Image Preprocessing.

We followed the procedures introduced by Dailey and Cottrell (1999) to preprocess the images (Figure 4). First, each image was processed using 2-D Gabor wavelet filters (5 spatial frequencies at 8 different orientations each), a simple model of complex cell responses in visual cortex. The filters were applied at 64 points in an 8x8 grid, resulting in a vector of 2560 elements (Dailey & Cottrell, 1999). The vectors were then normalized via z-scoring (scaled and shifted so that they had zero mean and unit standard deviation) on a per-filter basis, a local operation. A principal component analysis (PCA) was applied to the normalized vectors. The top 40 components were saved and renormalized which constituted the input to the neural networks.

Neural Network Parameters

Networks were trained using a learning rate of 0.01 and momentum of 0.5. In all of our experiments, each module comprised of 40 input units, 20 hidden units. The number of output units was depended on the functionality of each module. Similarly, we trained all of our models until their RMSE reached 0.05.

Experiment

We ran each experiment on 50 networks. The training process is divided into 2 phases. In phase 1, the basic module and the expert module are pre-trained separately. The basic module is trained to classify cups, cans, asymmetric, and symmetric Greebles at the basic level, and the subordinate module is trained to classify faces at the subordinate level. This initial training is motivated by the prior experience of human subjects before performing the actual experiment. After phase 1, networks were tested on the symmetric and asymmetric Greeble images that will be trained in the next phase. Then we recorded the discrimination baseline for latter comparison.

During phase 2, half of the networks were trained to classify asymmetric Greebles at subordinate level and symmetric Greebles at basic level, and the other half did the reverse. Each network was trained on 100 images: 5 exemplars within 10 symmetric and 10 asymmetric Greeble species.

Networks were then tested on 300 images comprised of all images used during training, 5 new exemplars within each of the 20 learned species, and 5 exemplars within 10 new symmetric and asymmetric Greeble species.

We model discrimination performance as a function of the similarity of internal representations (hidden unit activations) between images. To measure the representational similarity, we computed the correlation between the vectors of hidden unit activations for two different input patterns.

$$similarity(v_1, v_2) = correlation(v_1, v_2)$$

In Tanaka's experiment, d-prime was used to quantify the "same/different" discrimination. As a simple model of d-prime we assume that the discrimination performance is based on the difference in similarities between images from the same species versus the similarities between images of different species. Images from the same species will be highly similar; images from different species will not have a similar representation. We assume subjects' discrimination scores are a function of the difference between these. Hence our measure of discrimination between two species is:

$$D(sp1, sp2) = \frac{1}{\binom{|sp1|}{2} + \binom{|sp2|}{2}} \sum_{i,j \in sp1,sp2} sim(i,j) - \frac{1}{|sp1| * |sp2|} \sum_{i \in sp1, j \in sp2} sim(i,j)$$

where sp1 and sp2 are the two species being compared. The first term is the average similarity of exemplars within the two species, averaged together, and the second term is the average similarity of exemplars between the species.

Experiment 1

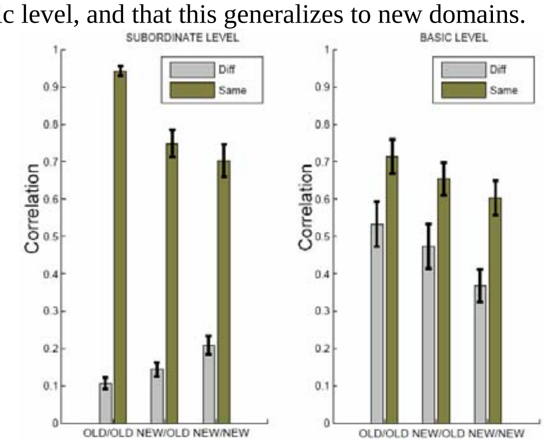
We started with a 2-module neural network model (Figure 4) which includes a basic classifier and an expert classifier.

Similarity Calculation There are two sets of hidden unit activations which come from the two modules of the network. During phase 1, the hidden layer of the basic module is used as the internal representation of the neural network to assess the baseline discrimination. This is similar to human subjects in Tanaka's experiment since they didn't have expert knowledge of owls and wading birds.

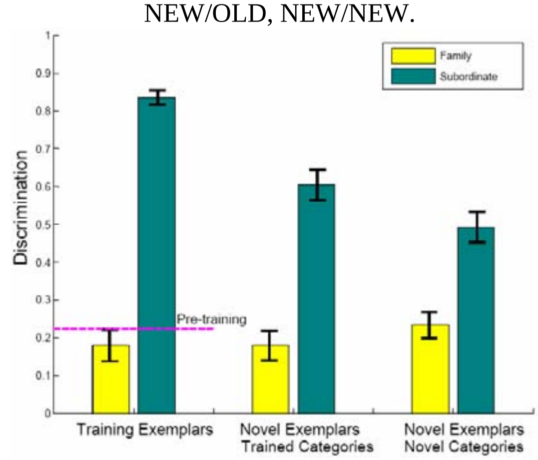
In phase 2, the hidden layer of the basic module is used to learn basic discriminations; and similarly the hidden layer of the expert module is used to learn subordinate classification. Our assumption is that the brain functionally separates the processing of basic and subordinate level information. Later we show how this might be automated. Here we assume that the discrimination is based upon the module most suited to the task:

$$Similarity(v_1, v_2) = \begin{cases} cor(basic(v_1), basic(v_2)), \text{ for basic stimuli} \\ cor(sub(v_1), sub(v_2)), \text{ for subordinate stimuli} \end{cases}$$

Result and Discussion Subordinate level training produced greater advantages in the "same/different" discrimination than basic level training. The advantages also carried over to novel exemplars of trained species and novel exemplars of novel species. This result is consistent with the effect reported by Joyce and Cottrell (2004). They have shown if a category is learned at the subordinate level, subjects are more sensitive to the differences among the individuals in that category than the category they have only learned at the basic level, and that this generalizes to new domains.



(a) "Same" and "Diff" are the average correlations between exemplars within the same species and between exemplars from different species within 3 conditions: OLD/OLD, NEW/OLD, NEW/NEW.



(b) Discrimination = (Same - Diff).

Figure 5: Post-training discrimination. Error bars denote standard deviations.

However within the basic level the "same/different" discrimination measures of the training exemplars and novel exemplars of trained species were *below* the baseline. In addition, the discrimination of the novel exemplars of novel species was higher than training exemplars and novel exemplars of trained species. This results from the basic level network compressing the representations of classes. This makes sense, since the main function of the basic module is to *ignore* the differences among individuals and group similar stimuli together. However, subjects in Tanaka's experiment not only learned to classify stimuli, but also experienced the stimuli through mere exposure. This suggests that we should have an additional module in our network modeling this phenomenon.

Experiment 2

Based on the previous experiment, we know that basic level training compressed the representation of basic stimuli. We added an autoencoder as a model of perceptual learning from mere exposure (Figure 6). Autoencoders try to reproduce their input on their output, thus learning the statistics of their environment in a passive way. The side effect of this is that the representations of items so trained will be separated in hidden unit space.

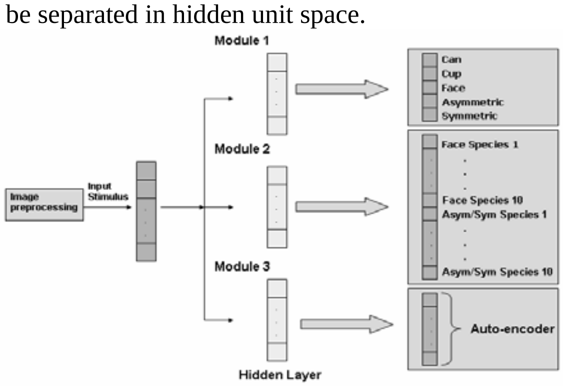


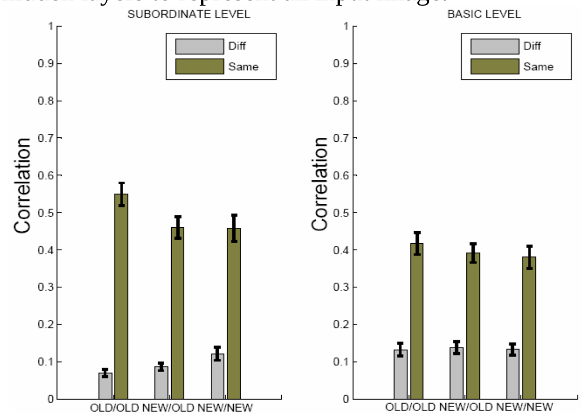
Figure 6: Multi-module neural network configuration.

Similarity Calculation Now there are three hidden layers from each of the basic, subordinate, and auto-encoder modules. To decide which hidden layer is used as the internal representation of the network, we based our decision on two motivations. First, the brain functionally processes the basic level information and the subordinate level information separately. Second, while learning a classification task, subjects are also experienced stimuli through mere exposure. Therefore, it's most likely that the brain accumulates information from the classification process and mere exposure to perform the "same/different" discrimination.

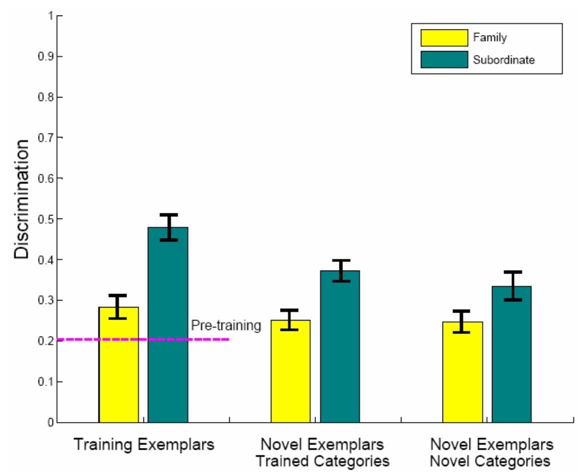
In phase 1, the concatenation of the basic and autoencoder hidden layer is used as the internal representation of the neural network to assess the baseline discrimination. In phase 2, we concatenate the hidden layers of the basic and autoencoder module and the hidden layers of the subordinate and autoencoder module to represent the internal representation of basic stimuli and subordinate stimuli respectively.

$$similarity(v_1,v_2)=\begin{cases} cor(basic(v_1)+autoencoder(v_1),basic(v_2)+autoencoder(v_2)), \text{for basic stimuli} \\ cor(sub(v_1)+autoencoder(v_1),sub(v_2)+autoencoder(v_2)), \text{for subordinate stimuli} \end{cases}$$

Result and Discussion Our multi-module neural network model has demonstrated the same results as human subjects do. Even though we have achieved our goal to model Tanaka's results with this multi-module network we would like to automate the process of choosing which combination of hidden layers to represent an input image.



(a) "Same" and "Diff" are the average correlations between exemplars within the same species and between exemplars from different species within 3 conditions: OLD/OLD, NEW/OLD, NEW/NEW.



(b) Discrimination = (Same - Diff).

Figure 7: Post-training discrimination. Error bars denote standard deviations.

Experiment 3

To automate the process of selecting which combination of hidden layers to represent an input image, we exploit a mixture network model (Figure 8). These networks implement a soft competition between the networks to process the data. The competition is based upon the output of a gating network that mediates the learning and activation flow in the two networks. Thus we set up a competition between the basic and expert level classifiers. The complete model is composed of the mixture network and a standard feed-forward network. The mixture model that we used is a modified version of the one described by Dailey & Cottrell (1999). The modification is that the output teaching signals of the gating network depend on the subordinate output activations. If any subordinate outputs are on, we turn on the teaching signal of the gating network that connects to the subordinate module. Hence, whenever subordinate classification is the goal, the system is forced to choose the expert network. We imagine that this is the role of the frontal lobe system. The mixture network learning rule is described in Appendix A.

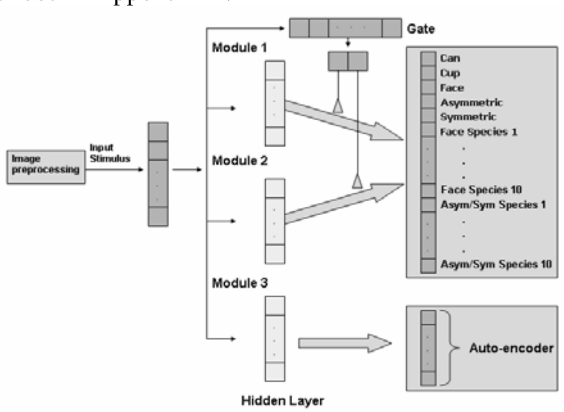


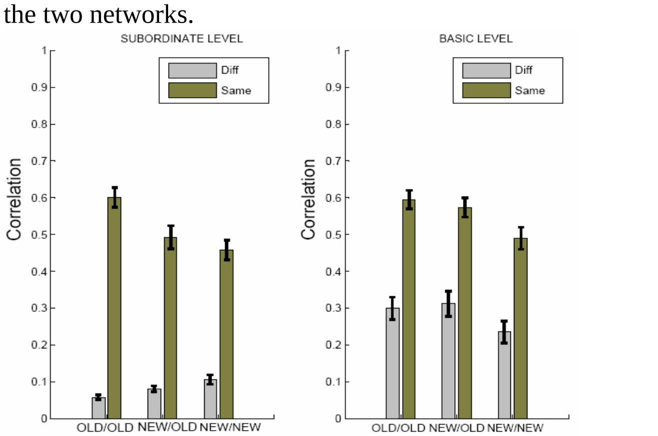
Figure 8: Multi-module neural network configuration.

Similarity Calculation In phase 1, similarly to experiment 2, the concatenation of the basic and autoencoder hidden layer is used as the internal representation of the neural network to assess the baseline discrimination.

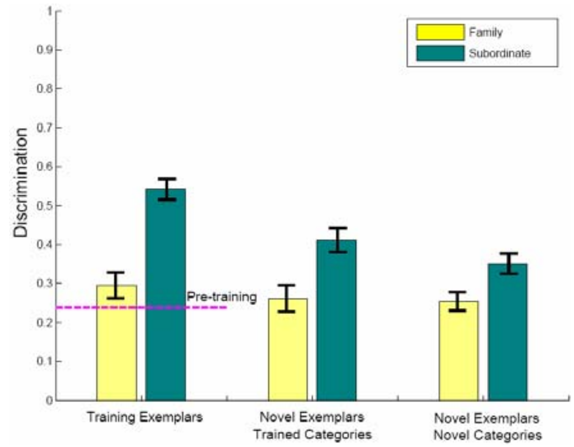
In phase 2, we use the hidden layers of the auto-encoder module with either the basic or subordinate module to represent input stimuli. To select between the hidden layers of the basic and subordinate module, we depend on the higher output activation of the gating network. The gating network has two output units which are pre-assigned to the basic and subordinate module.

$$Similarity(v_1, v_2) = cor(gating(v_1) + autoencoder(v_1), gating(v_2) + autoencoder(v_2))$$

Result and Discussion This result (Figure 9) indicates that we have again successfully modeled Tanaka's experiment. Moreover, we also have suggested a mechanism for the brain to choose between the levels of processing between the two networks.



(a) "Same" and "Diff" are the average correlations between exemplars within the same species and between exemplars from different species within 3 conditions: OLD/OLD, NEW/OLD, NEW/NEW.



(b) Discrimination = (Same - Diff).

Figure 9: Post-training discrimination. Error bars denote standard deviations.

Conclusion

Tanaka's study has shown that classification task at subordinate level, not mere exposure, is the most important ingredient in becoming an expert. Subordinate level training produced greater advantages in species discrimination relative to basic level training. Furthermore, transfer occurs in a graded fashion, to novel exemplars of known categories and novel exemplars from novel, but similar, categories.

Our multi-module network has similar performance as human subjects. Furthermore, it also supported the concept that the brain functionally separates the processing of basic and subordinate level information. Indeed, in experiments not reported here, we could not model this behavior in a single network. However, our model contained no mechanism for simply learning good (invertible) representations of experienced stimuli. The network actually saw basic level stimuli as "all looking alike:" In order to prevent this, we added a process, implemented here as an autoencoder, to learn faithful perceptual representations of frequently encountered stimuli. The autoencoder helps spread out the representation of the basic stimuli as well as the subordinate stimuli, and it can be considered as a model for mere exposure. In our final multi-module networks, we found it necessary to further assume that the brain can *choose* the representations it will use to compare stimuli. This is a novel prediction, as far as we know, and one that could be verified via fMRI.

Acknowledgments

We would like to thank Gary's Unbelievable Research Unit, Carrie Joyce for a preliminary experiment. This was supported by NIMH grant MH57075 to GWC.

References

Bishop, C. M. *Neural networks for pattern recognition.* Oxford University Press, 1995.

Cottrell, G. W., Brandon, K. M., & Calder, A. J. Do expression and identity need separate representations? In *Proceedings of the 24th Annual Conference of the Cognitive Science Society* (Mahwah, New Jersey, 2002), The Cognitive Science Society.

Dailey, M. N., & Cottrell, G. W. Organization of face and object recognition in modular neural network models. *Neural Networks* 12 (1999), 1053-1073.

Dailey, M. N., Cottrell, G. W., Padgett, C., & Adolphs, R. Empath: A neural network that categorizes facial expressions. *Journal of Cognitive Neuroscience* 14, 8 (2002), 1158-1173.

Daugman, J. G. Uncertainty relation for resolution in space, special frequency, and orientation optimized by two dimensional visual cortical filters. *Journal of the Optical Society of American* A 2 (1985), 1160-1169.

Devalois, R. L., & Devalois, K. K. *Spatial Vision*. Oxford University Press, 1988.

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

Gauthier, I., Curran, T., Curby, K. M., & Collins, D. (2003). Perceptual interference supports a non-modular account of face processing. *Nature Neuroscience*, 6, 428-432.

Gauthier, I., & Tarr, M. J. (1997). Becoming a 'Greeble' expert: exploring the face recognition mechanism. *Vision Research*, 37. 1673-1682.

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

Gauthier, I., Williams, P., Tarr, M. J., & Tanaka, J. W. (1998). Training "Greeble" experts: a framework for studying expert object recognition processes. *Vision Research*, 38, 2401-2428.

Harnad, Stevan (1987) Uncomplemented Categories, or, What Is It Like To Be a Bachelor. Presidential address to the 1987 meeting of *Society for Philosophy and Psychology*. http://cogprints.org/2134/index.html

Johnson, K., & Mervis, C. (1997). Effects of varying levels of expertise on the basic level of categorization. *Journal of Experimental Psychology: General*, 126, 248-277.

Jacobs, R. A., Jordan, M. I., Nowlan, S. J., & Hinton, G. E. (1991). Adaptive mixtures of local experts. *Neural Computation*, 3, 79-87.

Jones, J. P., & Palmer, L. A. An evaluation of the two dimensional gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology* 58, 6 (1987), 1233-1258.

Joyce, C., & Cottrell, G. W. Solving the visual expertise mystery. In *Proceedings of the Neural Computation and Psychology Workshop* 8, Progress in Neural Processing. World Scientific, London, UK, 2004.

Lecun, Y., Bottou, L., Orr, G. B., & Muller, K. R. Efficient backprop. In *Neural Networks-Tricks of the Trade, Springer Lecture Notes in Computer Sciences* (1998), vol. 1524, pp. 5-50.

Sugimoto, M., & Cottrell, G. W., (2001) Visual Expertise is a General Skill. *Proceedings of the 23rd Annual Cognitive Science Conference*, pp. 994-999.

Tanaka, J. W., Curran, T., Sheinberg, D.L. (in press). The training and transfer of real-world, perceptual expertise. *Psychological Science*.

Appendix A: Mixed hidden layer network

In the feed-forward phase, we first compute the input weighted sum of the hidden unit u_{ij} (i is the module number and j is the unit number in the hidden layer):

$$net_{ij} = \sum_k w_{ijk} x_k$$

Then we apply the sigmoid function to the weighted sum:

$$z_{ij} = 1.7159 \tanh(\frac{2}{3} net_{ij})$$

We also compute the input weighted sum of the output unit in the gating network and then apply the softmax function to that weighted sum:

$$net_i = \sum_k \theta_{ik} x_k \;, \qquad g_i = \frac{\exp(net_i)}{\sum_j \exp(net_j)}$$

Afterward, we compute the activations of the output units:

$$o_i = \sum_m (g_m \sum_j w_{mij} net_{mj})$$

The network is trained by online back-propagation of error with the generalized delta rule. First, we compute the update weights for the output layer:

$$\delta_{o_i} = 2(t_i - o_i)$$

$$\Delta w_{mij}(t) = \eta * z_{mj} * \delta_{o_i} + \alpha * \Delta w_{mij}(t+1)$$

Second, we compute the update weights for the hidden layer:

$$\delta_{u_{ij}} = g_i \frac{dz_{ij}}{dnet_{ij}} \sum_p \delta_{o_p} w_{pij}$$

$$\Delta w_{mij}(t) = \eta * x_j * \delta_{u_{ij}} + \alpha * \Delta w_{mij}(t+1)$$

Let g_1 and g_2 wire to the basic module and subordinate module respectively. Finally, we compute the teaching signals for the output units of the gating network:

$$t_g = \begin{cases} [0,1], \max(subordinate(o_i)) \geq 0.5 \\ [1,0], \max(subordinate(o_i)) < 0.5 \end{cases}$$

Then we compute the update weights for the gating network:

$$\delta_{g_i} = 2(t_{g_i} - o_i)$$

$$\Delta\theta_{ij}(t) = \eta * x_j * \delta_{g_i} + \alpha * \Delta\theta_{ij}(t+1)$$