

Experience Matters: Modeling the Relationship Between Face and Object Recognition

Panqu Wang (pawang@ucsd.edu)

Department of Electrical and Computer Engineering, University of California San Diego
9500 Gilman Dr 0407, La Jolla, CA 92093 USA

Isabel Gauthier (isabel.gauthier@vanderbilt.edu)

Department of Psychology, Vanderbilt University
301 Wilson Hall, Nashville, TN 37240 USA

Garrison Cottrell (gary@ucsd.edu)

Department of Computer Science and Engineering, University of California San Diego
9500 Gilman Dr 0404, La Jolla, CA 92093 USA

Abstract

Some research has suggested that face and object recognition are independent abilities. Recently, however, it has been shown that they are not, and that the relationship is moderated by experience with the object categories (Gauthier et al., in press). Gauthier et al. suggest that a domain general ability underlies face and object recognition that is expressed when people have sufficient experience in that category. Using the Cambridge Face Memory Test (CFMT) and Vanderbilt Expertise Test (VET), they showed that as experience with non-face object categories grows (averaged over all eight categories of the VET), the shared variance between the CFMT and VET performance increased monotonically. This theory fits with our neurocomputational model (“The Model”, TM, Cottrell and Hsiao (2011)), since in TM, categories differentiated at the subordinate level are recruited by the face network (Tong, Joyce, & Cottrell, 2008). We model “domain general ability” as the resources available for the mapping from images to labels (the number of hidden units), and “experience” as the number of training epochs with non-face objects. We show that, as in the data, the shared variance between the performance on faces and the performance on subordinate-level object categorization increases as experience grows. Our results thus suggest that a potential source for the variance in the “domain general ability” between individuals is the amount of representational resources available for fine-level discrimination. One might have expected that faces and objects *compete* for this shared resource, leading to a negative correlation between them. Our analysis of the hidden unit representations shows that they share a “spreading” transform, that moves similar objects apart in representational space, consistent with our previous analyses suggesting that this is why the Fusiform Face Area is recruited by new categories of expertise (Tong et al., 2008).

Keywords: face recognition; neural network; individual differences; Fusiform Face Area (FFA); object recognition.

Introduction

Extensive progress has been made in discovering how complex objects, particularly human faces, are processed by the visual cortex. At the same time, there is no consensus on whether face and non-face object recognition are performed independently. Some fMRI studies of the Fusiform Face Area (FFA) suggest there is a domain-specific response to faces compared to other object categories in that region (Kanwisher, McDermott, & Chun, 1997), while other studies show that the FFA also responds to non-face objects of expertise, such as cars, birds (Gauthier, Skudlarski, Gore, & Anderson, 2000), chessboards (Bilalić, Langner, Ulrich, & Grodd,

2011), and even on novel objects when subjects are sufficiently trained in the lab (“Greeble” experts, (Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999)). Single-cell recordings in macaque show highly face-selective cell patches, but it is unknown whether the macaques had expertise in any other category (Tsao, Freiwald, Tootell, & Livingstone, 2006). However, McGugin, Gatenby, Gore, and Gauthier (2012) reported a linear correlation between behavioral car expertise and a reliable response to cars in the FFA using HR-fMRI, and suggest experience with a category may be sufficient to create this activation.

Recently there has been more focus on individual differences in face and object recognition and the relationship between them. The development of the Cambridge Face Memory Test (CFMT) (Duchaine & Nakayama, 2006) has provided a valid and reliable measure of these differences in the normal population. In a study on the heritability of face recognition, Wilmer et al. (2010) assessed the independence of face recognition by measuring the correlation between the CFMT and a similar test about abstract art, and found the correlation is low (less than 0.26). Similarly, Dennett et al. (2011) designed the Cambridge Car Memory Test (CCMT), and found it only accounted for 13.6% of the shared variance in CFMT.

The results above suggest that face recognition is independent from non-face object recognition. However, Gauthier et al. (in press) challenged this idea, arguing that there is a domain-general visual ability, v , for discriminating visually similar objects that underlies face and non-face object recognition, and that this ability is only expressed in full within a category when people have sufficient experience, E , in that category. I.e., $Ability_{cat} = v * E_{cat}$. In order to test this hypothesis, they performed the following experiment: From 256 subjects, they collected three measures. First, the subjects took the CFMT to obtain a measure of their ability with faces. The CFMT involves studying 6 target faces, and then discriminating them from other faces. The catch is that at test time, the target faces and the distractor faces are in different lighting, pose, or both, from the study faces. This is

followed by a second study phase, and then a discrimination test where the targets and distractors are embedded in Gaussian noise. Second, the subjects took the Vanderbilt Expertise Test (VET; McGugin, Richler, Herzmann, Speegle, and Gauthier (2012)), a test structured to be similar in format to the CFMT, but using 8 non-face object categories. This test gives a measure of their abilities with objects (O-PERF). Finally, they collected self-ratings from the subjects of their experience with the 8 categories, on a scale from 1 to 9 (O-EXP).

According to their hypothesis, if there is a common ability v that is expressed through experience, then their performance on the VET should be more correlated with their performance on the CFMT as their experience with the object categories grows. Hence they divided their subjects into six levels of experience, based on their standard deviation from the mean experience (see Figure 2, bottom row). Then for each experience group, they regressed the CFMT score against the VET score. They showed that, as predicted, as experience grows, the shared variance between the CFMT and VET increased monotonically. When subjects have considerable experience with objects, if they perform well (poorly) with faces, they will also perform well (poorly) on non-face objects.

Our model of face processing (“The Model” (TM); Dailey and Cottrell (1999); Cottrell and Hsiao (2011)) fits well with this hypothesis, because in TM, as more subordinate-level experience is gained with a category, the face processing network is recruited for the category. Hence any resources in the face processing network are shared with expert object processing. TM has been successfully used to simulate perceptual phenomena such as facial expression perception (Dailey, Cottrell, Padgett, & Adolphs, 2002), the recruitment of FFA for other categories of expertise (Tong et al., 2008), and the development of hemispheric lateralization of face processing (P. Wang & Cottrell, 2013).

The basic structure is similar to the expert network described in Tong et al. (2008), where a two-layer error-driven artificial neural network is trained after preprocessing the images with Gabor filters and PCA. We map the domain general ability, (v), to the number of hidden units in TM, and experience, (E), to the number of training epochs when we train on a non-face object category. We train on faces first to simulate the abilities expressed by the scores on the CFMT, and then on non-face objects to simulate the abilities tested by the VET. We apply our model to four different object categories: faces, butterflies, cars and leaves. We show that, as in Gauthier et al.’s data, the shared variance between the performance on faces and the average performance on non-face objects increases as experience with the non-face objects grows.

In addition, we make a prediction. Gauthier et al. did not observe the correlation between scores on the VET and on the CFMT at the level of experience with one particular category, but instead, of overall experience with all eight categories. Here we demonstrate that this correlation with experience on one category can be observed in our computational model, given sufficient training data.

Finally, we analyze the hidden unit activations, and show that the effect of experience is to populate a larger region of representational space - that is, to spread the representation of individual objects out. This is the same phenomenon that we have demonstrated in our model of how the FFA is recruited for other tasks Tong et al. (2008).

In the next section, we provide a detailed description of how we use TM to simulate the experiment in Gauthier et al. (in press). The result section will present our findings and analysis of the hidden unit activations. We conclude with a discussion.

Methods

Model Architecture

The version of TM we use is shown in Figure 1. TM’s structure is layered from low-level visual features to high-level object categories. The first layer is a Gabor filter layer (five scales, eight orientations) that models the response of complex cells in V1 (Daugman, 1985). In the second layer, we perform Principal Component Analysis (PCA) on the Gabor filter responses to reduce the dimensionality. PCA can be implemented using Hebbian learning (Sanger, 1989), and simulates the structural encoding level beyond primary visual cortex, up to the lateral occipital region level. The third layer is a hidden layer in the neural network that learns feature representations in the service of the task. If the task is face recognition, the hidden layer will develop representations for faces adaptively over training, and we assume that the hidden layer corresponds to the FFA. The last layer is a categorization layer, which controls the level of discrimination between different stimuli, and provides labels for them to perform the final object recognition task. The last two layers are fully connected, and are trained using online backpropagation.

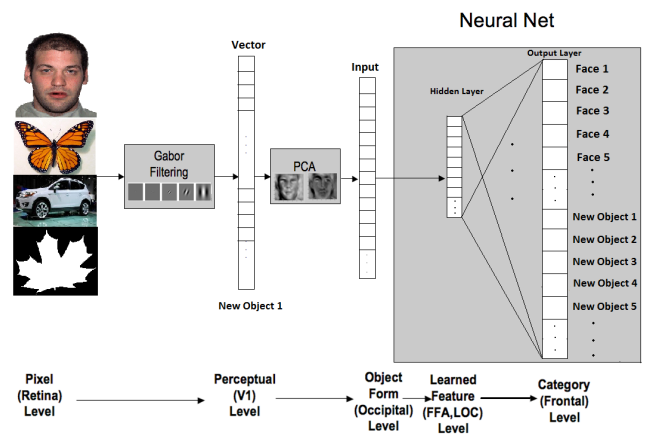


Figure 1: Model Architecture

Dataset and Preprocessing

We used four categories of objects in our studies: faces, butterflies, cars and leaves. As there is no single dataset that contains sufficient images for all these four expert categories,

we collected the images from four different datasets: 1) faces: NimStim Face Stimulus Set (Tottenham et al., 2009); 2) butterflies: Leeds Butterfly Dataset (J. Wang, Markert, & Everingham, 2009); 3) cars: Multi-View Car Dataset (Ozuysal, Lepetit, & Fua, 2009); 4) leaves: One-hundred Plant Species Leaves data Set (Mallah, Cope, & Orwell, in press). For each category, we randomly selected 16 images each from 10 individuals to form the training set (12 images per individual) and test set (4 images per individual). We transformed each image to grayscale first and cropped them to 64×64 pixels. The Gabor filtering stage consists of passing the image through the classical Gabor filter bank (Lades et al., 1993) with 5 different scales and 8 orientations ranging from 0 to $7\pi/8$. We then computed the Gabor magnitude, subsampled these vectors to an 8×8 grid, and normalized the response across orientations for each scale. The Gabor filtering process resulted in a 2560-dimensional vector to represent a single image. In order to extract a small number of features to represent the image efficiently and separate the response from each scale of the Gabor filter, PCA is performed separately on each spatial frequency component in the vectors (bandpass PCA). We kept the eight most significant eigenvectors for each scale, thus obtaining a 40-dimensional vector to represent each image prior to training the neural network.

Model Subjects Initialization and Training

There are two key variables in the psychological experiment performed by Gauthier et al. (in press): the hypothesized domain general visual ability, v , and the measured experience with a category, E . Here we assume that v corresponds to representational resources for fine-level discrimination, and hence we map visual ability (v) to the number of hidden units. With more hidden units, the network is able to generate more accurate (higher dimensional) features for a given object category, thus achieving better performance. Second, we map experience (E) to the number of training epochs on non-face object categories. To model the experiment, we make these two variables a function of the data from the 256 human subjects (see below), with one network per subject.

We first train each network on subordinate-level face recognition, as this is the first kind of expertise for most humans. We then train the network on an object class, subordinate-level classification of butterflies, cars, or leaves. Hence an extra set of output nodes are added for the second task, and error backpropagated from them will change the hidden unit representation. For Experiment 1, below, this is performed three times for each network starting with the same weights learned on faces (i.e., we “xerox” the network and run three experiments on it). The test set accuracy after training, averaged across the three tasks, is our model of their score on the VET. Note that we continue training on faces during the second task in order to avoid interference between the tasks. This is reasonable given that humans are nearly always exposed to faces every day. In the second experiment, we perform the same kind of analysis, but using only one object category, and varied training times, to simulate expertise

with the second category.

Experiments and Analysis

Experiment 1: Modeling Gauthier et al. (in press)

We first modeled the psychological experiment in Gauthier et al. (in press), described above. In the experiment, they obtained the CFMT performance, VET performance (O-PERF), and self-rating experience scores (from 1-9, O-EXP) on each category for each of the 256 human subjects. According to the average O-EXP scores across the eight categories of the VET, Gauthier et al. divided the subjects into 6 groups based on their standard deviation from the mean (see the legend to Figure 2). In each group, they performed a regression on the subjects’ CFMT scores against their object performance (O-PERF), and computed the correlation between them. They found that as O-EXP increases, the shared variance between CFMT and O-PERF increases monotonically from 6.2×10^{-6} to 0.59. This result indicated that people with considerable experience on VET object categories show a high correlation between their performance with face and non-face objects. Figure 2(a) shows their result. The bottom of the Figure shows which level of (self-reported) experience the subjects in that panel had with the VET objects.

To model these results, we make a one-to-one mapping between each subject and a network. Since, according to Gauthier et al.’s hypothesis, their score on the CFMT represents the expression of their v (because it is assumed that all subjects have high and relatively similar experience with faces), we use this score to decide each network’s representational resources. Hence, for each human subject at each experience level, we initialize a network with a number of hidden units based on their CFMT score. For each network subject s_{net} , we assign the number of hidden units from its corresponding human subject s_{hum} according to the following formula:

$$N_{hidden}(s_{net}) = \text{floor}(8 \times CFMT(s_{hum})),$$

where $CFMT(s)$ represents the fraction of correct responses for the given subject s_{hum} . The CFMT scores range from 0.4722 to 1, so the hidden unit numbers range from 3 to 8. The number “8” in the formula above was chosen arbitrarily, but 10 (for example) gives similar results.

Similarly, we mapped the self-rated experience (O-EXP) to a number of training epochs for objects as follows:

$$N_{epoch}(s_{net}) = 10 \times O-EXP(s_{hum})$$

O-EXP ranges from 1 to 9, so our training epochs range from 10 to 90. Note that while O-EXP is a noisy measure of experience (being based on self-report), here we are converting this to an exact measurement.

Stochastic gradient descent (online-backprop) is used to perform the network training. We set the learning rate to 0.015 and momentum term to 0.01 in all experiments. All weights between input layer to hidden layer, and hidden layer to output layer were set randomly between 0 and 1. As noted

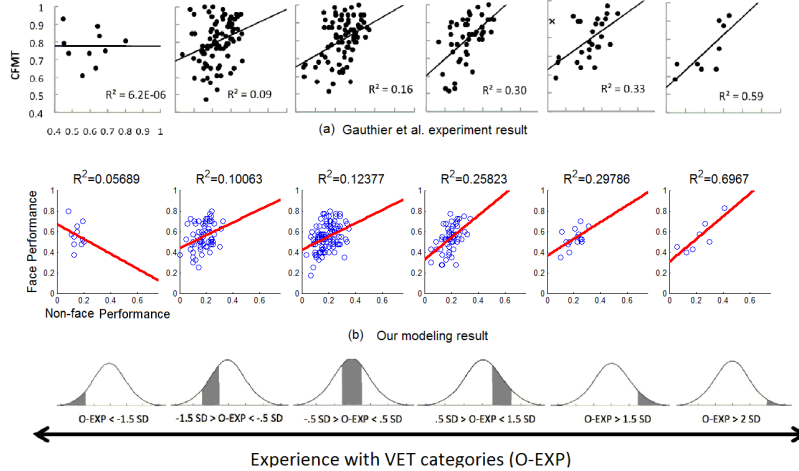


Figure 2: Gauthier et al. results and model results. Subjects are divided into groups according to their self-reported experience with the VET categories (O-EXP) (bottom row). For example, the first column (top row) shows the data from subjects whose O-EXP scores fell below 1.5 standard deviations (SD) from the mean. The top row shows the regressions for each group of their CFMT scores against their VET scores. The second row shows the results of our simulation. Each point in our graphs corresponds to a single network whose parameters (ν and E) are set based on one of the subjects in the graph above them.

above, the training begins with faces. The stopping criteria for face training is either when the training error is below 0.005 (determined by cross-validation), or the number of training epochs hits 200. Hence, either the network becomes an expert at faces and stops training, or it receives 200 passes through the training set if it can't reach that error criterion. Then training is continued on an object category for $N_{epochs}(s_{net})$. As noted above, this is repeated three times for each network, once for each category, and the results averaged across the three networks. At the end of training, we measured the recognition rate for face and non-face objects for all 256 network subjects and calculated the shared variance between the performance on faces and averaged performance on non-faces. The result is shown in Figure 2(b).

From Figure 2(b), we can clearly see that as experience (O-EXP) grows, the correlation between the recognition performance on faces and the average non-face scores increases monotonically from 0.057 to 0.698, which matches the result from Gauthier et al. (Figure 2(a)) qualitatively. This result suggests our mapping for E and ν to the network are reasonable, and the potential source for the variance in the “domain general ability” between individuals is the amount of representational resources (hidden units in the neural network).

Experiment 2: Correlation with One Category

One issue with the Gauthier et al. experiment is that the correlation was only found when experience across all categories of the VET was combined. The same results did not obtain when the analysis was restricted to a single category. This is also true of our results in Experiment 1 (data not shown for lack of space). This is a serious problem if our goal is to show that face recognition is not independent of *any* non-face object category. We believe that this lack of correlation could be the result of too small a sample size. Hence in this exper-

iment we use a much larger number of “subjects” and ability levels. Theoretically, we should see the same result as with the averaged category experience.

As there is no psychological data corresponding to this experiment, we created the initialization of the networks’ ν and E manually: 1) we map ν to a range of hidden unit numbers, $N_{hidden} \in \{1, 2, 3, 5, 6, 8, 12, 16, 20, 24, 28, 32\}$ and 2) we map E to a set of number of training epochs for non-face objects $N_{epoch} \in \{1, 5, 10, 20, 40, 60\}$. Instead of having 256 subjects in Experiment 1, we created 800 samples for each of the three non-face categories in this experiment, and then assigned the number of samples at each level of E and ν according to a Gaussian distribution, which is used to simulate the fact more people should have intermediate level of E and ν . The training procedure is exactly the same as Experiment 1, except that we computed the correlation based on the performance on faces and a *single* non-face category.

The results are shown in Figure 3. As can be seen from the figure, as experience grows, the correlation between the performance on faces and the performance on all three non-face categories increases monotonically, regardless of what the object category is. We performed the experiment multiple times and the result is stable. This result proves that our intuition is correct: sufficient data is a critical issue in observing the experience moderation effect. One phenomena worth mentioning is the moderation differs for each category at the end point (last column of Figure 3). For example, the R^2 ends at 0.683 for the car category but ends at 0.307 for the butterfly category. The end point distinction indicates the varied difficulties of different tasks.

Internal Representation Given the results above, we may wonder how the experience moderation effect occurs during training. Given that the two processes are competing for the

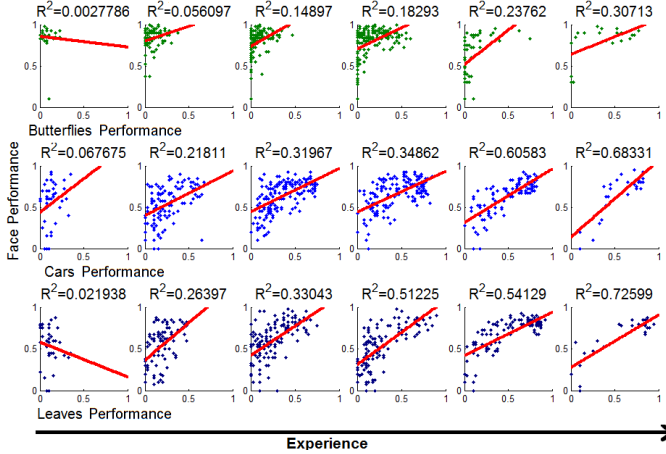


Figure 3: Experiment result (single category) of shared variance (R^2) between performance of face and non-face expert (butterflies, cars and leaves) as a function of experience. The data points are shown as colored dots in each subplot.

same resources, one might expect the correlation between them to be negative. On the other hand, if the resource was split in half, then it makes sense that more resources for one task means more resources for the other. The evidence that it is neither of these possibilities is based on the analysis used in Tong et al. (2008). They demonstrated that the hidden unit representation formed during subordinate-level training generalizes to new categories. That is, there is a “spreading transform” that separates similar-looking items on the hidden layer, and this transform generalizes to new examples. To demonstrate this, we visualize the development of internal representation by applying PCA to the hidden unit activations of the training data over learning, and plotting the second and third principal components (PCs) on a two-dimensional subspace (the first principal component just represents the magnitude of the activations, which reflects the growth in the weights).

The results of this analysis with the car category are shown in Figure 4. The plot clearly illustrates how experience contributes in the discrimination task. For faces, the data points for each individual become separated after training (left two columns). For cars, more experience also leads to more separation (fourth column). This inter-class separation is not only observed visually, but also statistically. The average between-class distance (measured using Euclidean distance between the center of each cluster in the right-hand column) is 8.448 for the network trained for 200 epochs (bottom row) and 6.405 for the network trained for five epochs (top row). This indicates that more experience leads to more separation between individuals. In addition, the within-class distance (measured using Euclidean distance between each data point belonging to a single individual to the average of that individual’s locations) in the network trained for 200 epochs is 2.849, much lower than the value 8.496 for the network trained for 5 epochs (top row). This suggests more experience also generates a more condensed and accurate data distribution, thus

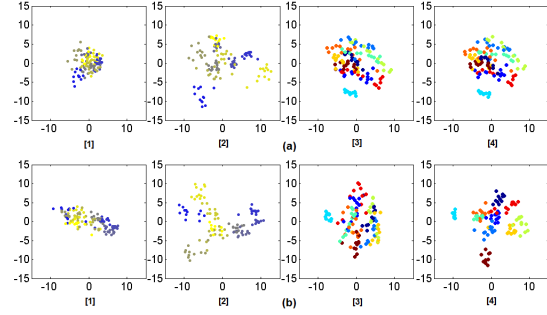


Figure 4: Visualization of hidden unit activations. The figure shows how the representational space of the hidden units (the second and third principal components of the activations) develops over network training. We take samples from epochs 1 and 200 of face training (first and second column), epoch 1 of car training (third column). The remaining column shows the difference between an experience level of 5 epochs of car training (top row) and 200 epochs of car training (bottom row). Colored dots correspond to different subordinate categories - in the left two columns, the colors correspond to individuals, in the right two columns, they correspond to individual car models.

improving the recognition rate.

For an untrained non-face object category, the performance must be low, regardless of how many hidden units there are, so the correlation is low. With more training, performance improves and becomes more dependent on the representational resources, leading to the shared variance increasing. By visualizing the development of internal representation, we can see how experience moderates face and object recognition.

At the same time, we can observe that in the third column, even without training on cars, the projections are already separated to some extent, especially compared to the first column, where the network has just started training on faces. As in our previous work Tong et al. (2008), this shows that the spreading transform learned for faces generalizes to new categories. This is an important characteristic of a subordinate-level classifier.

Conclusion

We showed that by instantiating the theoretical concept of a shared resource, v , as the number of hidden units in our network, we could replicate the moderation of experience on the relationship between face and object recognition. Experience allows the network to *express* this resource in performance on the second task. We also used TM to predict that the same phenomena will be observed at the individual category level, given sufficient training data. Finally, analysis of the developing internal representation shows how experience moderates the visual ability and recognition performance.

One potential critique of this work is that the CFMT and the VET are memory tests, and we use the recognition rate of unseen data to model this. However, we believe that if the recognition rate is high, the internal representation is well-

developed for the task. In a memory test, this would lead to better representations in memory as well. In future work, we intend to evaluate this idea more directly by using the models' internal representations in an exemplar-based memory model.

Acknowledgments

This work was supported by NSF grant SMA 1041755 to the Temporal Dynamics of Learning Center, an NSF Science of Learning Center (GWC and IG), VVRC (Grant P30-EY008126) (IG) and NEI (Grant R01 EY013441-06A2) (IG). We thank Benjamin Cipollini and Akinyinka Omigbodun for discussions on this work, and the rest of Gary's Unbelievable Research Unit (GURU) for their comments.

References

- Bilalić, M., Langner, R., Ulrich, R., & Grodd, W. (2011). Many faces of expertise: fusiform face area in chess experts and novices. *Journal of Neuroscience*, *31*(28), 10206–10214.
- Cottrell, G. W., & Hsiao, J. H. (2011). Neurocomputational models of face processing. In A. J. Calder, G. Rhodes, M. Johnson, & J. Haxby (Eds.), *The Oxford Handbook of Face Perception*. Oxford, UK: Oxford University Press.
- Dailey, M. N., & Cottrell, G. W. (1999). Organization of face and object recognition in modular neural network models. *Neural Networks*, *12*, 1053–1073.
- Dailey, M. N., Cottrell, G. W., Padgett, C., & Adolphs, R. (2002). EMPATH: A neural network that categorizes facial expressions. *Journal of Cognitive Neuroscience*, *14*(8), 1158–1173.
- Daugman, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortex filters. *Journal of the Optical Society of America*, *2*, 1160–1169.
- Dennett, H. W., McKone, E., Tavashmi, R., Hall, A., Pidcock, M., Edwards, M., et al. (2011). The cambridge car memory test: A task matched in format to the cambridge face memory test, with norms, reliability, sex differences, dissociations from face memory, and expertise effects. *Behavior Research Methods*, *44*(2), 587–605.
- Duchaine, B., & Nakayama, K. (2006). The cambridge face memory test: results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic subjects. *Neuropsychologia*, *44*(4), 576–585.
- Gauthier, I., McGugin, R. W., Richler, J. J., Herzmann, G., Speegle, M., & Gulick, A. E. V. (in press). Experience moderates overlap between object and face recognition, suggesting a common ability. *Journal of Vision*.
- Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, *3*(2), 191–197.
- Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform face area increases with expertise in recognizing novel objects. *Nature Neuroscience*, *2*, 568–573.
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*, 4302–4311.
- Lades, M., Vorbruggen, J., Buhmann, J., Lange, J., Malsburg, C. von der, Wrtz, R. P., et al. (1993). Distortion invariant object recognition in the dynamic link architecture. *Computers, IEEE Transactions on*, *42*(3), 300–311.
- Mallah, C., Cope, J., & Orwell, J. (in press). Plant leaf classification using probabilistic integration of shape, texture and margin features. *Signal Processing, Pattern Recognition and Applications*.
- McGugin, R. W., Gatenby, C. J., Gore, J. C., & Gauthier, I. (2012). High-resolution imaging of expertise reveals reliable object selectivity in the ffa related to perceptual performance. *Proceedings of the National Academy of Sciences*, *109*(42), 17063–17068.
- McGugin, R. W., Richler, J. J., Herzmann, G., Speegle, M., & Gauthier, I. (2012). The vanderbilt expertise test reveals domain-general and domain-specific sex effects in object recognition. *Vision Research*, *69*, 10–22.
- Ozuysal, M., Lepetit, V., & Fua, P. (2009, June). Pose estimation for category specific multiview object localization. In *Conference on computer vision and pattern recognition*. Miami, FL.
- Sanger, T. D. (1989). Optimal unsupervised learning in a single-layer linear feedforward neural network. *Neural Network*, *2*, 459–473.
- Tong, M. H., Joyce, C. A., & Cottrell, G. W. (2008). Why is the fusiform face area recruited for novel categories of expertise? A neurocomputational investigation. *Brain Research*, *1202*, 14–24.
- Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., et al. (2009). The nimstim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, *168*(3), 242–249.
- Tsao, D. Y., Freiwald, W. A., Tootell, R. B., & Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Science*, *311*, 670–674.
- Wang, J., Markert, K., & Everingham, M. (2009). Learning models for object recognition from natural language descriptions. In *Proceedings of the british machine vision conference*.
- Wang, P., & Cottrell, G. W. (2013). A computational model of the development of hemispheric asymmetry of face processing. In *Proceedings of the 35th annual conference of the cognitive science society*. Austin, TX: Cognitive Science Society.
- Wilmer, J. B., Germineb, L., Chabris, C. F., Chatterjeeb, G., Williamsd, M., Lokene, E., et al. (2010). Human face recognition ability is specific and highly heritable. *Proceedings of the National Academy of Sciences*, *107*(11), 5238–5241.